Université Fédérale
Toulouse Midi-Pyrénées

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse Jean Jaurès

Présentée et soutenue par

Mawa DAFREVILLE

Le 14 décembre 2022

Communication sensible à l'attention maternelle dans une perspective comparative : Approche développementale et multimodale

École doctorale et discipline ou spécialité
CLESCO – Doctorat en psychologie

Unité de recherche
Laboratoire Cognition, Langues, Langage et Ergonomie
CLLE (UMR CNRS 5263)

Directrices de Thèse
Prof. Michèle GUIDETTI
Dr. Marie BOURJADE

Jury
Prof. Maya GRATIER (Rapporteure)
Dr. Adrien MEGUERDITCHIAN (Rapporteur)
Prof. Bahia GUELLAÏ (Présidente, examinatrice)
Dr. Emilie GENTY (Examinatrice)

Université Toulouse Jean Jaurès

Laboratoire Cognition, Langues, Langage et Ergonomie (CLLE)

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ

Psychologie

Communication sensible à **l'attention maternelle dans une** perspective comparative : Approche développementale et multimodale

DAFREVILLE Mawa

Présentée et soutenue publiquement

Le 14 décembre 2022

Directrices de Recherche

Prof. Michèle GUIDETTI

Dr. Marie BOURJADE

JURY

Prof. Maya GRATIER (Rapporteure)

Dr. Adrien MEGUERDITCHIAN (Rapporteur)

Prof. Bahia GUELLAÏ (Présidente, examinatrice)

Dr. Emilie GENTY (Examinatrice)

Remerciements

Je souhaiterais remercier le professeur **Maya Gratier** et le docteur **Adrien Meguerditchian** ainsi que le professeur **Bahia Guellaï** et le docteur **Emilie Genty** pour l'honneur qu'ils me font d'avoir accepté d'évaluer ce travail et de se prêter à l'échange pluridisciplinaire qu'il suscite.

Je tiens à remercier le professeur **Michèle Guidetti** qui a su être là quand il le fallait et dont la prise de recul a grandement contribué à donner toujours plus d'horizon à cette thèse. Je remercie tout particulièrement le docteur **Marie Bourjade** avec qui nous avons partagé bien plus que cette thèse qui nous animait et qui m'a transmis bien plus que son « sens » de la Recherche tout au long de ce travail. Je vous remercie toutes deux pour votre confiance, votre patience et votre soutien.

Je remercie la Région Réunion de m'avoir financée et soutenue dans ce projet.

Je remercie le docteur **Catherine Hobaiter**, grâce à qui j'ai pu faire mes premiers pas en primatologie et qui dans une grande bienveillance, m'a donné goût à la Recherche et aux chimpanzés. Je remercie également le docteur **David Sillam-Dussès** qui a été présent avant même le commencement et qui m'a accompagnée et encouragée ces dernières années. Je tiens à remercier **Butu, Ngoya, Patrick, James et Kesier** du **projet UNBP** (*The Uaso Ngiro Baboon Project*) qui à leur manière ont eux aussi contribué à la beauté de cette thèse. Aussi, un grand merci à mes encadrantes de l'avant-thèse pour leur rôle déterminant dans mon parcours, à savoir le docteur **Laurianne Rat-Fisher** et le docteur **Sandra Molesti.**

Je remercie mes « petites sœurs de la multimodalité » à savoir **Noémie et Claire** qui ont su mettre de la couleur dans cette investigation avec une mention particulière pour toi Claire, qui n'as jamais véritablement terminé ton stage tant tu vis cette thèse avec moi. Merci pour ta pertinence, tes relectures et tes éclats de rire. Je remercie mes amies de thèse **Malvina, Lucille et Pauline** pour nos échanges sans fin sur la thèse, sur nos sujets si différents et surtout : sur tout le reste… Egalement un grand merci à mon ami **Damien** qui s'est attelé à de nombreuses heures de codage pour m'aider à construire la base de données de mes rêves. Sans oublier **Kathleen** qui aura soutenu un an tout rond juste avant moi et dont les nombreux conseils m'ont habitée tout au long de la rédaction.

Merci à mes amies de Bordeaux pour leur soutien sans faille depuis mon arrivée « loin du soleil » et avec qui j'ai tant échangé sur ma thèse. Je cite **Netichou, Benenut, Ritek, Coco, Maé, Mathildou et Paulinette**. Sans oublier mes amis-de-l'ergonomie à savoir **Camille, Marion Albert et Anthony** ainsi que toute l'équipe. Je finirai par mes amies de toujours **Coco(nut) et Marion Ducos** ainsi que **Mimi** qui ont partagé les plus grands moments de mon développement jusqu'ici. Absolument tous sans exception ont donné *un* sens à ce travail de thèse.

Je remercie **ma mamie** et **tatie Marie-Thé** d'avoir été là malgré la distance et de vivre mes aventures de thèse avec tant de bienveillance. Merci pour ces petites attentions que je recevais et qui ont égayé les différents hivers de ma thèse.

Je remercie mes grandes sœurs, **Walé et Maïré** qui ont chacune pleinement participé à l'aboutissement de cette thèse. L'une m'a poussée à l'amorcer tandis que l'autre m'a plus qu'aidée à la finaliser par ses nombreuses relectures et son sens du détail. Elles ont toutes deux donné naissance à cinq primates (humains) fascinants qui auront littéralement suscité mon intérêt pour le développement de l'enfant. Je cite **Frida, Amalia, Mariszka, Marjane et Sveinn**. Sans oublier mes beaux-frères **Loïc et Brian**. Presqu'enfin, je remercie **Robin** qui aimait bien me rappeler qu'étudier les vocalisations d'un bébé pour investiguer les origines du langage humain c'est bien mais « qu'en est-il des vocalisations du velociraptor dans tout ça ? ». Je lui suis d'une pleine reconnaissance pour son soutien ces dernières années et sa douceur dans la gestion de crises auxquelles toute thèse est confrontée. Je remercie aussi sa famille pour nous avoir accueillies maintes fois, ma thèse et moi.

Enfin, je remercie **mes parents** qui ont littéralement habité cette thèse avant même mon existence. Je remercie mon papa pour ses nombreux appels faisant suite à une émission qu'il avait entendue sur les singes ou les bébés sur France Inter et ma maman pour ses appels sans fin dans lesquels l'attention maternelle était au cœur de tous les sujets.

Je terminerai par remercier chaleureusement toutes les dyades mère-enfants de primates humains et non-humains qui m'ont permis de mener à bien cette thèse.

Sommaire

PRELUDE.....13

1. INTRODUCTION 17

1.1. ORIGINES DU LANGAGE HUMAIN ET APPROCHE *EVO-DEVO* 19

1.1.1. Origines du langage humain et « problème de réagencement modal »..... 19

1.1.2. Théories du conflit attentionnel 20

1.1.3. De l'approche *Evo-Devo* à la pragmatique développementale 23

1.1.3.1. Entre l'approche Evo-Devo et le socioconstructivisme 25

1.1.3.2. Tous ces chemins mènent à la pragmatique développementale.....26

1.2. DE LA PRAGMATIQUE DEVELOPPEMENTALE A L'APPROCHE COMPARATIVE 29

1.2.1. Pragmatique développementale : du perlocutoire au locutoire (intentionnalité) 29

1.2.2. Vers une approche comparative : entre l'illocutoire et l'intentionnalité 30

1.2.2.1. Degrés d'intentionnalité 31

1.2.2.2. Opérationnalisation 32

1.2.2.2.1. Persistance et élaboration..... 32

1.2.2.2.2. Coordination du regard 33

1.2.2.2.3. Communication sensible à l'attention visuelle du destinataire 35

1.3. ETUDE DE LA COMMUNICATION SENSIBLE A L'ETAT ATTENTIONNEL DU DESTINATAIRE 35

1.3.1. Chez le chimpanzé (*Pan troglodyte*)..... 36

1.3.1.1. Etudes expérimentales 39

1.3.1.2. *Etudes observationnelles.....41*

1.3.1.3. *Etudes développementales chez le chimpanzé immature : peu nombreuses 42*

1.3.2. Chez l'enfant préverbal (*Homo sapiens*) 43

1.3.2.1. *Etudes expérimentales 46*

1.3.2.2. *Etudes observationnelles.....49*

1.3.2.3. Etudes développementales chez l'enfant préverbal : divergences méthodologiques et de résultats.....50

1.4. VERS *UNE* APPROCHE MULTIMODALE : PERTINENCE DU CODAGE ETHOLOGIQUE 52
1.4.1. Principe du codage éthologique 52
1.4.2. Mise en application sur les travaux existants chez l'enfant préverbal 53
1.4.2.1. Développement du répertoire : émergence, combinaisons multimodales et regards 54
1.4.2.1.1. Emergence des signaux 54
Les signaux audibles 54
Les signaux silencieux-visuels 55
Les signaux tactiles 56
1.4.2.1.2. Combinaisons multimodales : audio-visuelles 57
1.4.2.1.3. Coordination des signaux de communication avec le regard 58
1.4.2.2. Développement de l'utilisation sociale de ces signaux 62
1.4.2.2.1. La communication dyadique 62
1.4.2.2.2. La communication triadique 63
1.4.3. Intégration dans une perspective comparative 64
1.5. PROBLEMATIQUE 67
1.5.1. Questions de recherche 67
1.5.2. Opérationnalisation des concepts 70
2. METHODE GENERALE 73
2.1. CHOIX METHODOLOGIQUES ET MACROSCOPIE DE L'INVESTIGATION 75
2.1.1. Choix méthodologiques généraux 75
2.1.1.1. *Espèces étudiées 75*
2.1.1.2. *Périodes développementales investiguées 77*
2.1.1.3. *Observations en contexte écologique 78*
2.1.2. Macroscopie de l'investigation générale 79
2.2. ETUDE OBSERVATIONNELLE SUR LES JEUNES CHIMPANZES (*PAN TROGLODYTES*) 81
2.2.1. Sujets-chimpanzés 81
2.2.2. Collecte de données 82
2.2.3. Echantillonnage des comportements 84
2.2.3.1. Codages des signaux produits par le jeune chimpanzé 84
2.2.3.1.1. Distinction entre gestes simples et produits en séquence .. 84
2.2.3.1.2. Association multimodale de signaux en séquence 84
2.2.3.2. Codage de l'attention visuelle maternelle 85
2.2.3.3. Concordance inter-observateur 85

2.3. ETUDES OBSERVATIONNELLES SUR LES ENFANTS PREVERBAUX (*HOMO SAPIENS*) 86

2.3.1. Sujets-humains.....86

2.3.2. Procédure et Design 87

2.3.2.1. *Déroulement de la pré-visite* *87*

2.3.2.2. *Déroulement des visites 1 et 2 à domicile* *88*

2.3.2.3. *Questionnaire du « Développement du Langage de Production en Français » (DLPF ; Bassano et al., 2005, 2020) 89*

2.3.3. Collecte de données 90

2.3.4. Echantillonnage des comportements et Procédure de codage 91

2.3.4.1. Codage des signaux produits par l'enfant *92*

2.3.4.1.1. Distinction de signaux : cas particuliers..... 92

2.3.4.1.2. Combinaison de signaux..... 97

2.3.4.1.3. Coordination des signaux de communication avec le regard 97

2.3.4.2. *Codage de l'attention maternelle*.....*98*

2.3.4.3. Concordance inter-observateur *100*

2.4. ANALYSE : OPERATIONNALISATION DES AJUSTEMENTS ET STATISTIQUES INFERENTIELLES 101

3. COMMUNICATION SENSIBLE A L'ATTENTION MATERNELLE CHEZ LE CHIMPANZE IMMATURE (*PAN TROGLODYTES*) 105

3.1. ARTICLE 1: **SENSITIVITY TO THE COMMUNICATIVE PARTNER'S ATTENTIONAL STATE: A DEVELOPMENTAL STUDY IN THE WILD CHIMPANZEE (*PAN TROGLODYTES SCHWEINFURTHII*)** 107

4. COMMUNICATION SENSIBLE A L'ATTENTION MATERNELLE CHEZ L'ENFANT PREVERBAL (*HOMO SAPIENS*).....163

4.1. ARTICLE 2: **ATTENTION-SENSITIVE SIGNALLING BY 7-TO-20-MONTH-OLD INFANTS IN A COMPARATIVE PERSPECTIVE** 165

4.2. ARTICLE 3: **BEYOND POINTING GESTURES: ATTENTION-SENSITIVE COMMUNICATION IN 7-TO-20-MONTH-OLD INFANTS DEVELOPS PRAGMATIC SKILLS** 219

5. COMMUNICATION SENSIBLE A L'ATTENTION MATERNELLE ET PREDICTION LANGAGIERE.....261

5.1. ARTICLE 4 : **L'AJUSTEMENT DE LA COMMUNICATION A L'ATTENTION VISUELLE DE LA MERE PREDIT-IL LES COMPETENCES LANGAGIERES ULTERIEURES ?**..... 263

6. DISCUSSION GENERALE 293

6.1. RAPPEL DES PRINCIPAUX RESULTATS 295

6.2. HISTOIRES DEVELOPPEMENTALES DE LA COMMUNICATION SENSIBLE A L'ATTENTION DU DESTINATAIRE CHEZ LE PRIMATE IMMATURE . 297

6.2.1. Chez le chimpanzé : une histoire de négociation et de manipulation d'objet 297

6.2.2. Chez l'enfant préverbal : une histoire d'ajustement transitoire et de signaux-leviers 303

6.2.2.1. *L'histoire des ajustements à l'attention visuelle* *303*

6.2.2.2. *Effet du Langage Adressé à l'Enfant (LAE) sur les ajustements* *309*

6.3. **L'**AJUSTEMENT DE LA COMMUNICATION A L'ATTENTION VISUELLE DE LA MERE PREDIRAIT DES COMPETENCES LANGAGIERES ULTERIEURES 313

6.3.1. Apports méthodologiques 314

6.3.2. Apports et questionnements théoriques 315

6.4. MISE EN PERSPECTIVE COMPARATIVE ET CONFRONTATION DES RESULTATS AUX THEORIES DU CONFLIT ATTENTIONNEL 317

6.4.1. Confrontation des résultats en inter-espèce 317

6.4.1.1. *Ordre différentiel d'émergence des ajustements d'une espèce à l'autre* *317*

6.4.1.2. *Intégration de l'objet dans le geste : moteur de l'ajustement intermodal* *319*

6.4.2. Théories du conflit : effet du LAE et ajustement du babillage 321

6.4.2.1. *Effet du LAE et hypothèse de Falk (2004)* *321*

6.4.2.2. *Ajustement du babillage et hypothèse de Locke (2006)* ... *323*

6.5. LIMITES ET PERSPECTIVES 324

6.5.1. Dans une perspective développementale : des histoires à compléter 324

6.5.1.1. *Dans l'étude des ajustements chez le jeune chimpanzé : vocalisations et profil du destinataire* *324*

6.5.1.2. *Dans l'étude des ajustements chez l'enfant préverbal : responsivité maternelle et élaboration de la signalisation* *326*

6.5.2. Dans une perspective comparative : un manque de représentativité à considérer et une rythmicité à explorer 327

6.5.2.1. *Représentativité de l'échantillon humain : des variations culturelles à considérer* *328*

6.5.2.2. *Rythmicité des échanges : du geste audible à la vocalisation ?* *329*

6.6. CONCLUSION 331

7. BIBLIOGRAPHIE 333

8. ANNEXES 355
Annexe 1 : Formulaire de consentement 357
Annexe 2 : Autorisation de droit à l'image 359
Annexe 3 : Amendement au consentement 360
Annexe 4 : Questionnaire de pré-visite 361
Annexe 5 : Questionnaire de visite 1 363
Annexe 6 : Questionnaire de visite 2 365
Annexe 7 : Matériel additionnel de la discussion générale 367
Annexe 8 : Liste des publications et des communications réalisées pendant la thèse 374
Annexe 9 : Enseignements réalisés pendant la thèse 375

Liste des tableaux*

1. INTRODUCTION.....17

Table 1. Résultats des études ayant testé la communication sensible à l'attention du destinataire chez le chimpanzé selon le design d'étude, l'âge de la population et les signaux testés.....38

Table 2. Résultats des études ayant testé la *communication sensible à l'attention du destinataire* chez l'enfant préverbal selon le design d'étude, l'âge des enfants et les signaux testés. 45

Table 3. Décomposition des conditions expérimentales de l'expérience de Liszkowski et al. (2008) selon l'état d'attention visuelle de l'adulte, le Langage Adressé à l'Enfant (LAE) et la responsivité de l'adulte.....48

2. METHODE GENERALE.....73

Table 1. Tranche d'âge d'appartenance, nombre de gestes produits, sexe et âges en mois à la première et dernière observation par individu et moyenne d'âge pondérée (+ écart-type) en mois par tranche d'âge.....82

Table 2. Tranche d'âge d'appartenance, sexe et âges en mois à la première et seconde visite par individu et moyenne d'âges (+ écart-type) en mois par sexe et par tranche d'âge.....87

Table 3. Synthèse des coefficients Kappa de Cohen calculés sur les signaux produits par l'enfant et l'association du regard aux signaux (sur 20 % du jeu de données total)..101

Table 4. Synthèse des coefficients Kappa de Cohen calculés sur les variables attentionnelles maternelles. 101

4. COMMUNICATION SENSIBLE A L'ATTENTION MATERNELLE CHEZ L'ENFANT PREVERBAL (HOMO SAPIENS) 133

Table S1. Repertoire of human 7-to-20-month-old infants' multimodal communication. 201

Table S2. Maternal attention variables used in the study.....207

6. DISCUSSION GENERALE.....263

Table 1. Mise en correspondance des gestes ajustés (ou proto-ajustés) avec le sens le plus probable associé à partir de Hobaiter et Byrne (2014) par tranche d'âge. 301

Table 2. Gestes audibles-ou-tactiles produits avec objet observés pour chaque espèce320

** Ici sont répertoriées toutes les Tables des chapitres hors-articles, exceptées les tables des répertoires comportementaux humains qui sont présents dans le matériel additionnel des articles 2 et 3, partie 4 et qui sont discutés tout au long de ce manuscrit.*

PRELUDE

Cette thèse sur articles a pour but de renseigner les modèles théoriques quant à l'étude des origines du langage humain (partie 1.1) et l'étude de la communication de l'enfant préverbal (partie 1.2). Toutes deux sont soumises à un questionnement relatif au *réagencement modal* des signaux menant à l'émergence de la parole (principalement vocale et audible). Ce *réagencement modal* relève à la fois du canal de production (i.e. gestuel, vocal) et de la modalité sensorielle de perception (i.e. signaux silencieux-visuels, audibles et tactiles). Pour cette double-investigation, il s'agissait d'étudier un des principaux marqueurs d'intentionnalité à savoir la *communication sensible à l'attention du destinataire* (partie 1.3) à travers l'approche développementale et comparative entre primates humains et non-humains. La *communication sensible à l'attention du destinataire* relève de la contrainte perceptive selon laquelle les signaux silencieux-visuels (e.g. gestes de quémande ou de pointage) ne peuvent pas être perçus par un destinataire visuellement inattentif (i.e. qui ne regarde pas la signaleur). Ainsi, cette contrainte perceptive allant de fait impliquer une restructuration de la signalisation (i.e. utilisation de signaux des autres modalités sensorielles), constitue un objet d'étude pertinent pour l'étude du *réagencement modal*. Faisant état des divergences méthodologiques des études existantes, l'approche multimodale (partie 1.4) semblerait être un premier outil d'investigation optimal pour alimenter la perspective comparative et développementale que nous souhaitions donner à ce travail. Sur la base de l'approche multimodale donc, nous nous proposons ici d'étudier la *communication sensible à l'attention du destinataire* dans une perspective développementale et comparative (partie 1.5).

1.
INTRODUCTION

1.1. ORIGINES DU LANGAGE HUMAIN ET APPROCHE *EVO-DEVO*

1.1.1. Origines du langage humain et « problème de réagencement modal »

L'étude des origines du langage humain est un challenge pluridisciplinaire (Wacewicz & Zywiczynski, 2016; Zywiczynski et al., 2017). Certains auteurs suggèrent l'existence d'un protolangage ne contenant pas encore la complexité syntaxique qui caractérise le langage humain moderne actuel. En revanche, la nature de ce protolangage est source de nombreux débats en termes de structure (Arbib, 2012; Bickerton, 1990), fonction (Bickerton, 1990; Fitch, 2010; Mithen, 2005) voire d'existence (Bickerton, 1990) et de modalités de production (Corballis, 2002; Dunbar, 2004; Fitch, 2010; Hewes, 1992).

En reconnaissant la fonction communicative du protolangage, certains auteurs (e.g. Arbib et al., 2008; Dunbar & Turkel, 2006; Fitch, 2000; Hewes, 1992) se sont intéressés à la nature et donc à la/les modalité(s) de production précurseur(s) du (proto)langage humain. Plusieurs théories s'affrontent à ce sujet. Certaines défendent une origine vocale (*vocal-first theories ;* e.g. Dunbar, 2004; Fitch, 2010) quand d'autres une origine gestuelle (*gesture-first theories ;* e.g. Armstrong & Wilcox, 2007; Corballis, 2002) du langage humain (Wacewicz & Zywiczynski, 2016; Zywiczynski et al., 2017). Si le langage était vocal dès son début, il semble toutefois difficile d'expliquer l'émergence de gestes au sein de la lignée humaine. Si le langage est d'abord apparu comme un système gestuel/manuel, et les systèmes non-parlés tels que ceux de la langue des signes sont tout aussi expressifs que la parole (Emmorey et al., 2007; Meguerditchian et al., 2011), pourquoi le langage aurait-il dû prendre une forme essentiellement vocale aujourd'hui (Hewes, 1992; Kendon, 2011; Wacewicz & Zywiczynski, 2016) ? En somme, quel que soit le point de départ (gestuel ou vocal), se pose un *problème de réagencement modal* (Kendon,

2011). C'est pourquoi partant du fait que le langage est multimodal et que les systèmes de communication des primates non-humains le sont aussi, on s'oriente depuis peu vers une origine multimodale du langage (e.g. Fröhlich et al., 2019; Wacewicz & Zywiczynski, 2016). Cette hypothèse multimodale suggère que le langage résulterait au moins partiellement d'une coévolution de ces différents systèmes de communication (Arbib et al., 2008; Fröhlich et al., 2019; Meguerditchian & Vauclair, 2014; Taglialatela et al., 2011). Cependant, bien que cette démarche multimodale offre un premier éclairage sur le *problème de réagencement modal* (Kendon, 2011), celui-ci n'en reste pas moins irrésolu. C'est pourquoi de nombreux questionnements persistent quant à l'articulation des différents canaux de production (vocaux et gestuels) d'une part mais aussi celle des différentes modalités sensorielles de perception mises en jeu dans la signalisation (voir partie 1.4), d'autre part.

1.1.2. Théories du conflit attentionnel

Certains auteurs suggèrent que *l'investissement parental* (Trivers, 1974) déterminant pour la survie de la progéniture, puisse avoir constitué une force évolutive et sélective plausiblement responsable de l'émergence des substrats vocaux du protolangage (Falk, 2004; Locke, 2006; Mehr & Krasnow, 2017). Cette notion *d'investissement parental* peut s'apparenter à l'effort que le parent fournit pour assurer la survie et le bon développement de sa progéniture au détriment d'autres activités qui augmenteraient sa *valeur adaptative individuelle* (survie x succès reproducteur) et surtout *inclusive,* comme celle de s'investir au profit d'autres progénitures déjà nées ou à venir (Hamilton, 1964; Trivers, 1974). De ce fait, cet investissement parental génère ce que Trivers a dénommé un *conflit parent-enfant (parent-offspring conflict)* (Trivers, 1974). En effet, il semble qu'à un certain stade de développement, parent et enfant ne s'accordent plus sur la quantité d'investissement parental à fournir : l'enfant réclame plus d'investissement parental

que le parent n'est en mesure de fournir s'il veut maximiser sa *valeur adaptative* inclusive. L'exemple de conflit parental le plus étudié est le sevrage pour lequel l'investissement est donc *matériellement* quantifiable (allaitement maternel). Toutefois, certains auteurs sous-entendent que cet investissement puisse être appréhendé de façon immatérielle via les ressources attentionnelles (Mehr & Krasnow, 2017). Ainsi, appliquées à la notion de conflit parental, ces ressources attentionnelles font l'objet d'un *conflit attentionnel* sous-jacent à certains scenarios évolutifs sur les origines du langage humain qu'on qualifiera ici de *théories du conflit attentionnel* (Falk, 2004; Locke, 2006; Mehr & Krasnow, 2017). En effet, ces scénarios relèvent de plusieurs théories telles que celles de Falk (2004), Mehr & Krasnow (2017) et Locke (2006).

Falk (2004) propose l'hypothèse du *Putting-the-baby-down* selon laquelle la parole aurait émergé chez les premiers hominidés pour compenser la diminution de contacts physiques entre la mère et l'enfant due à la généralisation de la bipédie. Falk (2004) met en avant une différence essentielle dans l'écologie des dyades mère-enfant entre humains et primates non-humains. Chez le chimpanzé par exemple, le bébé reste bien plus longtemps en contact physique avec sa mère puisqu'il est capable, dès ses premiers jours, de s'agripper à sa mère qui le transporte partout avec elle. Ce qui n'est pas le cas des bébés humains qui sont incapables, par eux-mêmes, de s'accrocher à leur mère avant plusieurs mois. Cela proviendrait de l'apparition de la bipédie exclusive dans notre lignée qui aurait engendré une réduction de la largeur du bassin et une augmentation de la taille de la boîte crânienne ; ces deux éléments entraînant le conflit obstétrical. La réponse adaptative à ce problème résulte dans un accouchement plus précoce, engendrant l'immaturité des bébés humains et leur long et intense développement post-natal. Il en découle que la mère humaine porte son bébé de manière active, ce qui, si elle n'a pas d'outil (tel qu'une écharpe ou un porte-bébé) pour le faire, ne lui permet pas d'utiliser ses mains. De ce fait, Falk (2004) propose que pour pouvoir se nourrir et libérer ses mains, les mères bipèdes des premiers représentants de la lignée humaine aient dû poser leur bébé de façon périodique. L'éloignement physique de

plus en plus fréquent entre la mère et son bébé aurait ainsi incité cette dernière à vocaliser pour compenser la distance physique et apaiser sa progéniture, aidant ainsi à l'émergence des substrats vocaux du protolangage chez l'humain (Falk, 2004).

Mehr & Krasnow (2017), à l'instar de Falk (2004), proposent une théorie selon laquelle les vocalisations maternelles auraient servi à compenser l'inattention maternelle mais plus particulièrement les vocalisations du *chant* adressé à l'enfant. Selon eux, c'est une forme d'investissement parental relativement peu coûteux mais suffisamment sophistiqué (en termes de rythmique et prosodie) pour répondre aux demandes de soin de l'enfant pendant la période de conflit parental (Mehr & Krasnow, 2017). Cette *attention mélodique* maternelle aurait ainsi permis d'assurer la survie de l'enfant tout en l'incitant à complexifier ses vocalisations (Mehr & Krasnow, 2017).

Locke (2006) propose l'hypothèse de *la sélection parentale* selon laquelle l'utilisation du babillage par l'enfant aurait constitué un indicateur de bonne santé pour le parent, qui aurait incité le parent à continuer son investissement. De ce fait, ces enfants bénéficiant de formes de soin parental plus sophistiquées que les autres, se verront augmenter leur probabilité de survie à partir de l'âge de 7 mois (cette théorie sera davantage étayée partie 1.1.3.2).

Ces différentes théories offrent des passerelles entre le développement de l'enfant humain et l'évolution de l'espèce pour suggérer une explication à l'émergence évolutive de substrats vocaux du protolangage à l'origine du langage humain. Elles s'inscrivent ainsi dans une approche *evo-devo* (Bjorklund & Pellegrini, 2000; Geary & Bjorklund, 2000; Legare et al., 2018) qui est développée ci-après.

1.1.3. **De l'approche** *Evo-Devo* à la pragmatique développementale

Ces différentes *théories du conflit attentionnel* relèvent ainsi de la psychologie *développementale évolutionnaire* qui, sur le principe de l'hérédité, s'intéresse aux *avantages adaptatifs* (en termes de survie et de reproduction) propres aux caractéristiques développementales de différentes espèces (Bjorklund & Pellegrini, 2000; Geary & Bjorklund, 2000).

Sur la Figure 1 ci-dessous, est représentée l'articulation des différentes approches, théories et différents concepts gravitant autour de la *psychologie développementale évolutionnaire* menant à la *pragmatique développementale*. Cette figure peut être parcourue en parallèle du texte ci-après à l'aide des numéros mis en exposant. Ceux-ci font référence aux notions articulées et développées dans les parties 1.1.3.1 et 1.1.3.2

Figure 1. Articulation des différentes approches, théories et différents concepts gravitant autour de la *psychologie développementale évolutionnaire* menant à la pragmatique développementale (Austin, 1962; Bates et al., 1975; Bruner, 1975).
Le numéros mis en exposant font référence aux notions articulées et développées dans le texte ci-dessous.

1.1.3.1. ***Entre l'approche Evo***-Devo et le socioconstructivisme

La *psychologie évolutionnaire*(1) s'est initialement construite sur le principe de *sélection naturelle* (initié par Darwin, 1958) qui dépendrait de l'adaptabilité d'un individu à son environnement. Les caractéristiques de l'individu facilitant cette adaptabilité (i.e. les avantages adaptatifs) permettront à celui-ci de survivre aux contraintes de son milieu et seront héritées par sa descendance. Ainsi, le fait que cette caractéristique persiste à travers plusieurs générations suggère que celle-ci ait pu constituer, au cours de l'évolution, un avantage adaptatif dans un environnement donné. Partant de là, la psychologie évolutionnaire applique ce même principe aux mécanismes psychologiques de l'humain contemporain, qui aurait ainsi résulté d'un processus d'adaptation. De plus, reconnaissant une influence de l'environnement *physique* et *social* dans l'évolution de ces mécanismes psychologiques et une expression différentielle des gènes selon diverses expériences, la psychologie évolutionnaire répond davantage au principe de l'*épigénétique* (Waddington, 1942, 2012, 2014). Depuis, certaines études ont montré qu'une grande partie des variations entre les individus proviendrait du développement de ceux-ci (Gottlieb, 2000). C'est pourquoi, certains auteurs suggèrent de rapporter l'approche développementale à la psychologie *évolutionnaire* (Bjorklund & Pellegrini, 2000; Geary & Bjorklund, 2000).

Soutenant ainsi l'idée que l'individu serait soumis à différentes pressions adaptatives propres à chaque stade de son développement, cela donnera naissance à la ***psychologie développementale évolutionnaire*(2)** (psychologie *evo-devo;* Bjorklund & Pellegrini, 2000; Geary & Bjorklund, 2000). Celle-ci permet ainsi d'articuler les notions d'héritage propre à l'espèce et de plasticité développementale propre à l'individu en les « situant » dans un environnement écologique et socio-culturel (Bogin, 1997; Bronfenbrenner, 1992, 2005; Geary & Bjorklund, 2000). Les stratégies adaptatives observables chez l'individu immature résulteraient donc de sa confrontation à des *pressions* environnementales *physiques* et *sociales* dont celles liées à l'**investissement parental (3)** (e.g. liées au conflit telles que la distanciation

physique par exemple). Reconnaître ainsi une compatibilité entre ce qui relève de l'environnement physique et social pour appréhender le développement de l'enfant permet d'articuler cette approche avec une approche **socioconstructiviste**[4] (pour revue voir Martinez, 1989). Cette dernière reconnait l'importance de la *médiation sociale* entre l'enfant et son environnement notamment assurée par le **parent médiateur** [5] (Bruner, 1981; Vygotsky, 1978).

L'influence parentale[6] considérée à l'échelle évolutive et développementale constitue ainsi un véritable point de *jonction* entre les théories *evo-devo* et socioconstructivistes. Que ce soit à travers l'approche *evo-devo* ou socioconstructiviste, **l'acquisition du langage**[7] impliquerait la part active (voire créative) de l'enfant dans son développement et la reconnaît comme étant entre autres, fonction de la *responsivité parentale*. Il nous a donc semblé pertinent de juxtaposer les théories de Locke (2006, 2017) et Bruner (1975), s'inscrivant respectivement dans une approche *evo-devo* et socioconstructiviste. Selon la théorie de *sélection parentale* de **Locke**[8] (2006, 2017) le babillage de l'enfant constituerait un indicateur de valeur adaptative (fitness), incitant le parent à poursuivre ses soins parentaux et donc plus responsivité parentale. Sur la base des **actes de langage**[9] (Austin, 1962; Searle, 1965), **Bruner**[10](1975) propose de l'appliquer au développement communicationnel de l'enfant humain. Ainsi, il propose que le développement des fonctions prélinguistiques soit initialement motivé très tôt par l'effet perlocutoire (i.e. l'effet produit sur le destinataire) et donc de responsivité parentale. Ce qui donnera naissance à la **pragmatique développementale**[11] (Austin, 1962; Bates et al., 1975; Bruner, 1975; Searle, 1965).

1.1.3.2. Tous ces chemins mènent à la pragmatique développementale

Locke (2006, 2017)[8] propose l'hypothèse de *la sélection parentale* qui aurait été fonction des capacités de flexibilité vocale de l'enfant (voir Figure 2). Il compare ainsi la courbe d'incidence des pleurs d'enfants âgés de 0 à 1 ans à la courbe

d'incidence d'infanticides recensés au Etats-Unis dans les années 80 selon l'âge de l'enfant (Locke, 2006). Il part du principe que bien que les pleurs de l'enfant constituent un *signal d'alarme* incitant le receveur adulte à prodiguer des soins à celui-ci, ces pleurs peuvent aussi constituer un réel facteur de stress pour l'adulte s'ils sont anormalement fréquents (Locke, 2006). Locke (2006) montre que la fréquence des pleurs diminue à deux reprises au cours de la première année de vie : à 3 et 7 mois, de manière concomitante à la diminution d'infanticides (Shapiro-Mendoza et al., 2006) et de décès naturels (Murphy et al., 2010). Ces deux âges coïncident également avec deux étapes du développement vocal *fondamental* (voir partie 1.4.2.1.1). En effet, c'est à partir de l'âge de 3 mois que l'enfant entre dans une phase d'extension du répertoire vocal (voir Figure 3, partie 1.4.2.1) marquée à la fois par une diversification des vocalisations prélinguistiques mais aussi par une flexibilité d'usage de celles-ci pendant l'exploration vocale (Oller, 2000; Oller et al., 2013). De plus, l'âge de 6-7 mois est associé à l'émergence du babillage (voir Figure 3, partie 1.4.2.1) constituant la vocalisation prélinguistique la plus sophistiquée en comparaison aux autres et la plus ressemblante à la parole (Esteve-Gibert & Guellaï, 2018a; Oller et al., 1999; Stoel-Gammon, 2011). Locke (2006) propose que ces deux compétences vocales procurent des avantages sélectifs aux enfants en signalant aux parents qu'ils sont en bonne santé et qu'il y a un intérêt à poursuivre leur investissement parental. Cet investissement soutiendrait leur développement vocal pour *in fine* acquérir une maîtrise suffisante de celui-ci (en termes de flexibilité) avant l'accès à d'autres dimensions plus complexes de l'échange (*pragmatique*, *sémantique* et *syntaxique*; Locke & Bogin, 2006).

Figure 2. Illustration de la théorie évolutive de la *sélection parentale* de Locke (2006) construite à partir de travaux empiriques actuels (Murphy et al., 2010; D. K. Oller, 2000; D. K. Oller et al., 2013; Shapiro-Mendoza et al., 2006). Cette théorie constitue ainsi une bonne illustration de l'approche *evo-devo* défendue par la psychologie développementale évolutionnaire.

Sur la base de la théorie des **actes de langage**[9] (Austin, 1962; Searle, 1965), **Bruner (1975)**[10] propose une théorie de l'*ontogenèse* des actes de langage. Il articulera ainsi l'approche socioconstructiviste avec la pragmatique développementale (Martinez, 1989). Selon Bruner (1975), le développement langagier dépend d'une connaissance *pratique* du monde à travers les interactions avec l'adulte médiateur, qu'il rapportera au concept de *pragmatique*. Certes l'enfant devra s'armer de précurseurs syntaxiques et sémantiques mais selon lui, c'est avant tout la pragmatique qui sera primordiale dans cet apprentissage (Bruner, 1975, 1981). Il part de l'observation que très tôt dans le développement, l'adulte reprend assez peu l'enfant pour ses erreurs syntaxiques ou sémantiques puis plus tard, il le fera davantage (Bruner, 1975, 1981). Le langage s'acquiert au service de la fonction communicative en tant que telle (son efficacité) et non directement à celui de la conformité grammaticale (Bruner, 1975, 1981). L'enfant apprendrait donc en premier lieu à se *positionner* face à son interlocuteur selon les usages culturels du groupe auquel il appartient (Martinez, 1989). Ce qui nous amène donc à considérer

la **pragmatique développementale**[11] (Austin, 1962; Bates et al., 1975; Bruner, 1975; Searle, 1965) qui sera étayée ci-dessous.

1.2. DE LA PRAGMATIQUE DEVELOPPEMENTALE A L'APPROCHE COMPARATIVE

1.2.1. Pragmatique développementale : du perlocutoire au locutoire (intentionnalité)

Dans les années 60, plusieurs courants tels que celui de la *pragmatique* puis des *actes de langage* prennent en compte le langage en termes d'échange social et non seulement de façon abstraite en s'intéressant au *sujet interactif* (Austin, 1962; Bates et al., 1975; Martinez, 1989). Austin (1962) propose un modèle innovant pour son époque, il suggère que certaines phrases ne soient pas des *descriptions* d'évènements mais des *évènements* en soi, c'est-à-dire des actes qui se manifestent à l'instant même où la phrase est produite. Il affirme que chaque phrase produite est porteuse de trois actes des langages : *locutoire*, *illocutoire* et *perlocutoire* (Austin, 1962). L'acte *locutoire* implique tous les actes nécessaires pour produire un énoncé verbal (construction syntaxique et articulation des mots). L'acte *illocutoire* est un acte socialement conventionnel, reconnu comme tel par celui qui parle et par son interlocuteur (actes d'ordonner, de conseiller, de proposer…). L'acte *perlocutoire* relève plus de la conséquence produite sur l'interlocuteur à l'énonciation de la phrase.

A cela, Searle (1965) précise la distinction entre illocution et locution en soulignant que chaque phrase peut être divisée en deux parties : un contenu propositionnel (*locutoire*) et performatif (*illocutoire)*. Par exemple, quand je dis « Amalia lance la balle », le contenu performatif pourrait être déclaratif mais aussi impératif voire interrogatif. Des études de l'époque ont alors montré qu'à l'âge de 2 ans, la fonction performative semblait être maîtrisée (Bates, 1971; Volterra & Antinucci, 1979) mais qu'elle pourrait résulter d'une histoire développementale

qu'on apparentera à l'émergence de l'*intentionnalité* à travers l'étude du geste, du regard et des vocalisations prélinguistiques (Bates et al., 1975).

Sur la base de la théorie des *actes de langage* (Austin, 1962; Searle, 1965), Bates et al. (1975) vont expérimentalement tester cette théorie et découvrent comment ces trois actes (i.e. locutoire, illocutoire et perlocutoire) s'échelonnent au cours du développement de l'enfant préverbal. Les auteurs délimitent ainsi trois stades : (1) le stade *perlocutoire* dans lequel le signal de l'enfant provoquera systématiquement un effet sur son destinataire sans avoir intentionnellement contrôlé cet effet ; (2) le stade *illocutoire* dans lequel l'enfant utilise intentionnellement des signaux non-verbaux (i.e. reconnus pour leur valeur de signal) afin de demander et de diriger l'attention de l'adulte vers des objets ou évènements ; (3) le stade *locutoire* dans lequel l'enfant construit ses premières propositions verbales et prononce des sons dans les mêmes séquences performatives qu'il exprimait précédemment de façon non-verbale (Bates et al., 1975).

Bates et al. (1975) estiment que le stade *illocutoire* associé à la manifestation d'une signalisation non-verbale intentionnelle débuterait autour de l'âge de 10 mois (Bates et al., 1975). Ce stade est marqué par d'importants changements comportementaux qui ont inspiré d'autres champs d'études investiguant notamment la communication intentionnelle chez d'autres espèces non-verbales telles que les primates non-humains (e.g., Leavens et al., 2005; Townsend et al., 2017), comme développé ci-après.

1.2.2. Vers une approche comparative : entre l'illocutoire et l'intentionnalité

En effet, les travaux réalisés sur l'enfant ont inspiré le champ de la primatologie comparée qui se confrontait à de nombreuses contraintes méthodologiques relatives à l'investigation de l'intentionnalité dans l'étude de la communication du primate non-humain (e.g., Bourjade et al., 2014; Fröhlich et al., 2018; Leavens et al., 2005; Maille et al., 2012; Townsend et al., 2017).

1.2.2.1. **Degrés d'**intentionnalité

Jusqu'alors, l'étude de l'intentionnalité de la communication humaine et non-humaine reposait sur la théorie de Grice (1969). Selon celle-ci, si un individu communique de façon intentionnelle, il agit non seulement avec l'intention de modifier le comportement ou l'état mental de son destinataire, mais il fera également en sorte que ce destinataire reconnaisse cette intention pour y répondre. Dans la continuité de cela, Dennett (1983) propose de distinguer plusieurs niveaux d'intentionnalité allant des ordres zéro à quatre que nous pouvons illustrer en donnant l'exemple d'un bébé, Olive, qui pleure en ressentant la sensation de faim.

- L'intentionnalité d'*ordre zéro* ne reconnaît aucun état mental ou d'intention de communiquer chez le signaleur. Exemple : *Olive pleure en réaction à sa sensation de faim.*
- L'intentionnalité de *premier ordre* implique que le signaleur émette son signal dans le but de provoquer une réponse comportementale chez le destinataire sans nécessairement admettre que ce dernier le reconnaisse en tant que tel. Exemple : *Olive pleure pour que sa maman lui donne le sein.*
- L'intentionnalité de *second ordre* implique celle du premier ordre à la différence que cette fois, le signaleur attribue un état mental au destinataire. Exemple : *Olive pleure pour que sa maman pense qu'il a faim.*
- L'intentionnalité de *troisième* ordre implique que le signaleur agisse avec l'intention de changer l'état mental du receveur, pour que celui-ci reconnaisse son intention de changer son comportement. Exemple : *Olive pleure pour que sa maman pense qu'Olive veut qu'elle lui donne le sein.*
- Enfin, l'intentionnalité de *quatrième ordre* implique que le signaleur agisse avec l'intention de changer l'état mental du receveur, pour que celui-ci reconnaisse son intention de changer son comportement. Exemple : *Olive pleure pour que sa maman sache qu'Olive veut qu'elle pense qu'Olive a faim.*

Chez les animaux non-humains, l'étude de la communication intentionnelle est principalement basée sur l'intentionnalité de *premier ordre*. Toutefois, des auteurs soulignent que l'intentionnalité de *second ordre* serait plus pertinente pour l'étude des précurseurs du langage (e.g. Graham et al., 2019) mais l'investigation de l' « attribution d'un état mental » semble difficilement opérationnelle d'un point de vue méthodologique chez les animaux non-humains (Premack & Woodruff, 1978; Townsend et al., 2017). De plus, l'approche comparative dans l'étude de la communication intentionnelle n'est pas de reconnaître chez les animaux non-humains des compétences linguistiques (Townsend et al., 2017). Elle constituerait plutôt des éléments de réponses quant à l'évolution de cette propriété langagière en en recherchant les précurseurs (Townsend et al., 2017). Partant du fait que l'intentionnalité de *second ordre* implique nécessairement celle de l'ordre inférieur au préalable, étudier l'intentionnalité de *premier ordre* constitue en soi, un précurseur suffisant (Graham et al., 2019; Townsend et al., 2017).

1.2.2.2. Opérationnalisation

S'inspirant des travaux réalisés sur l'enfant d'âge préverbal (Bates et al., 1975), les travaux sur la communication des primates non-humains proposent trois principaux *marqueurs comportementaux* de communication intentionnelle à savoir : la persistance et l'élaboration, la coordination du regard vers le destinataire (*visual checking)*, et la communication sensible à l'attention du destinataire (*sensitivity to the recipient's attention state* ; Fröhlich et al., 2018; Leavens et al., 2005; Townsend et al., 2017).

1.2.2.2.1. Persistance et élaboration

L'étude de la persistance et de l'élaboration relève du fait que l'on reconnaisse qu'un signal a été produit dans un but précis (*goal-directedness* ; Townsend et al., 2017). On peut tester la persistance et l'élaboration en observant si le signaleur arrête de produire ce signal une fois que le but présumé est atteint (*Pan troglodytes* : Leavens et al., 2005). Ainsi, si le destinataire ne répond pas

favorablement au signal produit, on s'attend à ce que le signaleur mette en place des stratégies comportementales pour atteindre son but, en répétant ce même signal (*persistance*) ou en modifiant la forme du signal, ou en le combinant à d'autres (*élaboration*) (*Pan troglodytes* : Hobaiter & Byrne, 2011a; Leavens et al., 2005; Liebal et al., 2004). Chez l'enfant préverbal (*Homo sapiens*), le marqueur de la persistance et de l'élaboration a été assez peu investigué en tant que tel. Toutefois, les études auxquelles il est indirectement possible de rapporter ce marqueur sont celles qui auraient testé l'effet de la responsivité *contrôlée* (Igualada et al., 2015; Liszkowski et al., 2008; Wu & Gros-Louis, 2017) ou *spontanée* (Gros-Louis & Wu, 2012; Wu & Gros-Louis, 2014) d'un adulte sur la production du geste de pointage combinée à des vocalisations chez l'enfant de 10 à 18 mois. Elles comparaient la signalisation (synchrone et diachronique) de l'enfant quand le parent répondait *correctement* au pointage (e.g. en s'engageant conjointement en direction de l'objet désigné par l'enfant) ou *incorrectement* à celui-ci (i.e. sans y répondre, ou en y répondant de façon incorrecte par exemple en se retournant en direction d'un objet différent que celui désigné par l'enfant).

1.2.2.2.2. Coordination du regard

La coordination du regard vers le destinataire (*visual checking*) relève du fait que l'on reconnaisse qu'un signal soit dirigé vers un receveur pour atteindre un but social (*social-directedness* ; Townsend et al., 2017). On peut tester la coordination du regard en observant si le signaleur s'oriente physiquement en direction du receveur *avant*, *pendant* et/ou *après* avoir produit son signal, via l'orientation du corps, du visage ou des yeux (voir revue : Emery, 2015). En effet, chez le primate plus que dans de nombreuses autres espèces, le comportement du regard aurait été soumis à de nombreux changements au cours de l'évolution (Emery, 2015). Ces changements pourraient avoir été provoqués par les changements morphologiques du visage et des yeux des primates, des limitations dans l'anatomie faciale des autres vertébrés, des changements environnementaux propres aux primates, et une nécessité de communiquer des informations sur ces environnements, leurs états

émotionnels et mentaux (Emery, 2015). Chez l'enfant préverbal (*Homo sapiens*), l'étude du regard a été largement investiguée (voir partie 1.4.2.2.1) via l'étude *de l'engagement visuel* (coordination des regards avec certains signaux ; voir partie 1.4.2.2.1; Colonnesi et al., 2012; Lin & Green, 2009; Yale et al., 2003) avant l'âge de 9 mois. A partir de 8-9 mois, les études s'intéresseront davantage à l'*alternance des regards* entre le destinataire et un objet de référence dans l'étude de l'*attention conjointe* pour appuyer la présence de la communication intentionnelle (Beuker et al., 2013; Carpenter et al., 1998; Franco & Butterworth, 1996). Soulignons par ailleurs que de nombreuses études réalisées sur des enfants âgés de 10 et 24 mois ont exclusivement restreint leurs analyses à des gestes (déictiques, conventionnels et symboliques) et vocalisations associés à des regards (Aureli et al., 2017; Gros-Louis & Wu, 2012; Guidetti, 2005; Igualada et al., 2015; Iverson & Goldin-Meadow, 2005; Rodrigues et al., 2021; Wu & Gros-Louis, 2014, 2017). Ces études considéraient que la coordination du regard vers le destinataire des signaux produits constituait un marqueur d'intentionnalité *suffisant* dans l'étude de la communication de l'enfant (Csibra, 2010; Csibra & Volein, 2008). Comme souligné par Donnellan et al.(2019), c'est *a minima* un marqueur d'intentionnalité de *premier ordre* (c'est-à-dire que le signal est produit dans le but de provoquer un changement comportemental chez le destinataire mais sans nécessairement que le signaleur ne lui infère des états mentaux ; voir partie 1.2.2.1 ; Dennett, 1983; Scott-Phillips, 2015; Sperber & Wilson, 1995; Townsend et al., 2017). C'est d'ailleurs probablement dans une perspective d'appréhender de plus près l'intentionnalité de *deuxième ordre* (i.e. qui, en plus de celle de premier ordre, implique que le signaleur infère nécessairement des états mentaux à son destinataire ; voir partie 1.2.2.1), que l'alternance de regards suscite un intérêt tout particulier dans l'étude de *l'attention conjointe* (voir partie 1.4.2.2.2). Notons par ailleurs, que la coordination du regard vers le destinataire d'un signal produit n'est absolument pas considérée comme étant suffisante pour justifier d'une quelconque intentionnalité dans la communication du primate non-humain (e.g., *Pan troglodytes* : Byrne et al., 2017; Leavens et al., 2004; Leavens et al., 2005; Townsend et al., 2017).

1.2.2.2.3. ***Communication sensible à l'attention visuelle du destinataire***

Tout comme pour la coordination du regard vers le destinataire, la *communication sensible à l'attention du destinataire* (*sensitivity to the recipient's attention state*) relève elle aussi du fait que l'on reconnaisse qu'un signal soit dirigé vers un receveur pour atteindre un but social (*social-directedness* ; Townsend et al., 2017). Ce marqueur relève d'une *contrainte perceptive* selon laquelle un signal silencieux (e.g. geste de pointage, geste de quémande) ne sera perceptible par son destinataire uniquement si ce dernier est visuellement attentif (i.e. regarde dans sa direction). Ce marqueur sous-tendrait une capacité du signaleur à d'une part, discriminer la situation où le destinataire est visuellement *attentif* (i.e. regarde dans la direction du signaleur) de celle où il est visuellement *inattentif* (i.e. ne regarde pas dans la direction du signaleur) et d'autre part, ajuster ses signaux en fonction (e.g., *Papio anubis* : Bourjade et al., 2014; *Pan troglodytes* : Leavens et al., 2004; *Cercocebus torquatus torquatus* : Maille et al., 2012). Les études ayant investigué ce dernier marqueur chez le chimpanzé (*Pan troglodytes*) et l'enfant préverbal (*Homo sapiens*) sont étayées dans la partie 1.3 ci-après.

1.3. ETUDE DE LA COMMUNICATION SENSIBLE A L'ETAT ATTENTIONNEL DU DESTINATAIRE

Comme mentionné plus haut, la *communication sensible à l'attention du destinataire* constitue un des principaux marqueurs d'intentionnalité qui relève de la contrainte perceptive selon laquelle des signaux silencieux-visuels tels qu'un pointage produit isolément ne sont pas perceptibles par un destinataire visuellement inattentif. Dans la littérature éthologique, ce marqueur d'intentionnalité a été grandement appréhendé dans l'étude comportementale de nombreuses espèces de primates (*Pongo pygmaeus*, *gorilla gorilla*: Botting & Bastian,

2019; *Papio anubis*: Bourjade et al., 2014; *Macaca tonkeana*: Canteloup & Bovet, 2014; *Macaca radiata*: Deshpande et al., 2018; *Gorilla gorilla*: Genty et al., 2009; *Cercocebus torquatus*: Maille et al., 2012) et non-primates (*Canis familiaris*: Horowitz, 2009; *Gymnorhina tibicen*: Kaplan, 2011; *Corvus corax*: Pika & Bugnyar, 2011; *Equus caballus*: Proops & McComb, 2010). Etant donné les espèces étudiées dans ce travail de thèse, nous ne développerons ici que les travaux relatifs à ce sujet réalisés sur les populations de chimpanzés (*Pan troglodyte*) et d'humains (*Homo sapiens*).

1.3.1. Chez le chimpanzé (*Pan troglodyte*)

Les études qui appréhendent la *communication sensible à l'attention du destinataire* (*sensitivity to the recipient's attention state* ; Fröhlich et al., 2018; Leavens et al., 2005; Townsend et al., 2017) chez le chimpanzé s'apparentent aux études s'intéressant à la communication intentionnelle. Elles sont en majorité expérimentales et réalisées sur des populations d'âge très variable (voir Table 1).

Sur la Table 1 ci-dessous, sont répertoriés les résultats des études ayant testé la *communication sensible à l'attention du destinataire* chez le chimpanzé selon le design d'étude, i.e. expérimental (Exp.) ou observationnel (Obs.) sur des individus captifs (Lab.) ou sauvages (Eco.) mais aussi selon l'âge des individus, les signaux testés et la mesure de comparaison. En ce qui concerne la mesure de comparaison, on distinguait les études dans lesquelles les auteurs comparaient le nombre de signaux d'une même modalité d'une condition à l'autre (i.e. *distribution*) ou la fréquence relative de cette modalité de signal sur l'ensemble des signaux produits dans une condition (i.e. *proportion*) d'une condition à l'autre. Les études mentionnées sont développées dans les parties 1.3.1.1 et 1.3.1.2 et discutées au regard de l'approche développementale dans la partie 1.3.1.3.

Table 1. Résultats des études ayant testé la communication sensible à l'attention du destinataire chez le chimpanzé selon le design d'étude, l'âge de la population et les signaux testés.

				Gestes Silencieux-visuels		***Gestes Audibles***		***Gestes Tactiles***		***Vocalisations***	
Design	*Etudes*	*Age (mois)*	*Mesure*	*Att.*	*Inatt.*	*Att.*	*Inatt.*	*Att.*	*Inatt.*	*Att.*	*Inatt.*
Exp.	Povinelli et al., 2003	Juvéniles (N=7)	Distribution	>							
	Hostetter et al., 2001	≥Juvéniles (N=49)	Distribution (latence)	>		ns				<	
	Kaminski et al., 2004	≥Juvéniles (N=7/16)*	Distribution	>							
	Leavens et al., 2004a	≥Juvéniles (N=101)	Distribution	>							
	Leavens et al., 2004b	Adultes (N=57)	Distribution	>		<					
	Hostetter et al., 2007	≥Juvéniles (N=116)	Distribution	>		ns					
	Leavens et al., 2010	≥Juvéniles (N=110)	Distribution	>		<				<	
	Tempelmann et al., 2011	≥Juvéniles (N=18)	Distribution	>		ns				ns	
	Liebal et al., 2004	≥Juvéniles (N=9)	Proportion	>		ns		ns			
Obs.-Lab.	Tomasello et al., 1994	≥Juvéniles (N=17) *EA = ns*	Distribution	>		>		<			
	Tomasello et al., 1997	Enfants-Juvéniles (N=20)	Distribution	>		>		ns			
	Pika & Liebal, 2006	Juvéniles-Adolescents (N=2/8)*	Proportion	>		ns		ns			
Obs.-Eco.	Hobaiter & Byrne, 2011	≥Enfants (N=81)	Proportion	>		ns		<			
	Fröhlich et al., 2018	Enfants-Juvéniles (N=12) *EA = sig*	Proportion	>		<					

*Note : EA = Effet de l'âge. Les catégories d'âges sont définies ainsi : enfants (0-4 ans), juvéniles (5-9 ans), adolescents (10-15 ans), adultes (>16 ans ; Hobaiter & Byrne, 2011; Reynolds, 2006). *Quand l'espèce de chimpanzé n'était pas la seule testée, la fraction A/B veut donc dire qu'il y avait A chimpanzés sur B sujets testés.*

1.3.1.1. Etudes expérimentales

Les études expérimentales étudiant la *communication sensible à l'attention du destinataire* s'inscrivent dans le paradigme de requête alimentaire (*food-requesting paradigm* ; Hostetter et al., 2001, 2007; Kaminski et al., 2004; Leavens, Hopkins, et al., 2004; Leavens, Hostetter, et al., 2004; Leavens et al., 2010; Liebal et al., 2004; Povinelli et al., 2003; Tempelmann et al., 2011) dans lequel on contraint le primate à communiquer avec un expérimentateur humain pour obtenir de la nourriture. Cela consistait à positionner de la nourriture hors de portée du sujet et à l'inciter à quémander de la nourriture de façon audible et/ou visuelle pour que l'expérimentateur humain la lui donne. Partant du principe que la stratégie de quémande diffère selon l'état attentionnel de l'expérimentateur (par exemple qu'il soit de face ou de dos), ces études faisaient varier son attention visuelle, délimitant ainsi les conditions d'attention et d'inattention visuelle (Hostetter et al., 2001, 2007; Kaminski et al., 2004; Leavens, Hopkins, et al., 2004; Leavens, Hostetter, et al., 2004; Leavens et al., 2010; Liebal et al., 2004; Povinelli et al., 2003; Tempelmann et al., 2011).

Pour cela, certains auteurs ont cherché à savoir quels indices pertinents les chimpanzés utilisaient pour détecter l'attention visuelle : était-ce l'orientation du *corps* (face/profil/dos), l'orientation du *visage* (face/profil/dos) ou des *yeux* (ouvert / fermé et direction du regard) ? En comparaison aux autres indices d'attention visuelle (i.e. orientation du corps et du regard, Hostetter et al., 2007), l'indice d'attention visuelle le plus expérimentalement utilisé est l'orientation du visage (Hostetter et al., 2007; Kaminski et al., 2004; Tempelmann et al., 2011). Les chimpanzés captifs produisaient significativement plus de gestes silencieux-visuels (en l'occurrence des gestes de quémande ou de pointage) face à un expérimentateur humain dont le visage était orienté vers eux que dans une autre direction (Tempelmann et al., 2011).

Ces études expérimentales s'attelaient à comparer pour la plupart la distribution des gestes silencieux-visuels, audibles (i.e. gestes produisant un son, par exemple taper sur le sol ou sur les barreaux de l'enclos) et tactiles (i.e. gestes impliquant un contact physique avec le destinataire) ainsi que des vocalisations pour certaines d'entre elles (voir Table 1). Bien que les âges observés soient variables, ces études sont pour la plupart menées sur des individus dont les plus jeunes sont totalement sevrés de leur mère (i.e. *juvéniles*, âgés de plus de 4 ans ; Reynolds, 2006), et ne sont plus portés par leur mère lors des déplacements (Bard, 2019; Bard et al., 2019). Elles s'accordent toutes sur le fait que les chimpanzés produisaient significativement moins de *gestes silencieux* quand l'expérimentateur n'était pas visuellement attentif que quand il l'était (Hostetter et al., 2001, 2007; Kaminski et al., 2004; Leavens, Hopkins, et al., 2004; Leavens, Hostetter, et al., 2004; Leavens et al., 2010; Liebal et al., 2004; Povinelli et al., 2003; Tempelmann et al., 2011). Concernant la distribution des *gestes audibles* d'une condition à l'autre, les résultats sont moins consensuels. Certaines études montrent que les chimpanzés en produisaient plus quand l'expérimentateur n'était pas visuellement attentif que quand il l'était (Leavens et al., 2004; Leavens et al., 2010) mais la plupart d'entre elles ne trouvait pas d'utilisation différentielle significative de ces gestes d'une condition d'attention visuelle à l'autre (Hostetter et al., 2001, 2007; Liebal et al., 2004; Tempelmann et al., 2011). Concernant la distribution des *gestes tactiles* d'une condition à l'autre, seuls Liebal et al. (2004) ont relevé des signaux tactiles dans ce contexte expérimental (*pull*, *touch with/without object* ; Liebal et al., 2004) mais ne montrent aucune utilisation différentielle significative de ces signaux d'une condition à l'autre. Par ailleurs, une autre spécificité de cette étude est que le design expérimental utilisé donnait également l'opportunité aux chimpanzés de se déplacer pour se mettre face à un expérimentateur initialement de dos à eux, et ainsi se replacer dans son champ visuel (Liebal et al., 2004). Les résultats montrent que les chimpanzés préféraient davantage se replacer dans le champ de vision du destinataire que produire des gestes audibles ou tactiles lorsque celui-ci était initialement visuellement inattentif (Liebal et al., 2004). Concernant la distribution

des *vocalisations* d'une condition à l'autre, la majorité des études ayant testé cette variable montre que les chimpanzés produiront davantage de vocalisations face à un expérimentateur visuellement inattentif qu'attentif (Hostetter et al., 2001, 2007; Leavens et al., 2010; à l'exception de Tempelmann et al., 2011).

1.3.1.2. *Etudes observationnelles*

A la grande différence des études expérimentales, les études observationnelles menées sur les chimpanzés investiguaient la *communication sensible à l'attention du destinataire* dans les interactions entre les individus d'une même espèce (i.e. en intra-espèce) et non plus avec un expérimentateur humain (i.e. en inter-espèce). Parmi ces études observationnelles, certaines ont été réalisées sur des chimpanzés captifs (Pika & Liebal, 2006; Tomasello et al., 1994, 1997) et d'autres sur des chimpanzés sauvages (Fröhlich & Hobaiter, 2018; Hobaiter & Byrne, 2011). Comme pour les études expérimentales, ces études ont été réalisées sur des âges assez variables mais à la différence que certaines d'entre elles incluaient également des enfants (âgés de moins de 4 ans; Reynolds, 2006) aux analyses (Fröhlich et al., 2018; Hobaiter & Byrne, 2011; Tomasello et al., 1997) et parfois même n'y incluaient pas les adultes (Fröhlich et al., 2018; Pika & Mitani, 2006; Tomasello et al., 1997).

Toutes ont montré que les chimpanzés produisaient moins de gestes silencieux-visuels quand leur destinataire conspécifique (i.e. de même espèce) n'était pas visuellement attentif que quand il l'était (Fröhlich et al., 2018; Hobaiter & Byrne, 2011; Pika & Liebal, 2006; Tomasello et al., 1994, 1997). En ce qui concerne les gestes audibles, les études ayant testé leur distribution d'une condition à l'autre montrent que les chimpanzés captifs en produisaient significativement plus quand leur destinataire était visuellement attentif que quand il ne l'était pas (Tomasello et al., 1994, 1997). D'autres études s'attelaient à comparer l'évolution en proportion des gestes audibles et/ou tactiles (par rapport à l'ensemble des signaux produits) et ne montrent pas d'utilisation différentielle de ces signaux d'une condition à l'autre chez les chimpanzés captifs (Pika & Liebal, 2006). Alors que les

chimpanzés sauvages produiraient significativement plus de gestes tactiles (et pas audibles) quand leur destinataire n'est pas visuellement attentif que quand il l'est (Hobaiter & Byrne, 2011).

1.3.1.3. *Etudes développementales chez le chimpanzé immature : peu nombreuses*

A notre connaissance, peu d'études se sont spécifiquement intéressées au développement des capacités d'ajustement des signaux associées à la *communication sensible à l'attention du destinataire* sur le chimpanzé immature (Fröhlich et al., 2018; Tomasello et al., 1997). Cependant, une étude récente réalisée sur des chimpanzés enfants et juvéniles montre un effet significatif de l'âge sur l'ajustement de la modalité des signaux de communication du jeune chimpanzé (Fröhlich et al., 2018) notamment avec des destinataires autres que leur mère. Toutefois, ayant été réalisée sur une population de chimpanzés enfant (0-4 ans) et juvéniles (4-9 ans), cette étude ne nous permet pas concrètement de situer quand est-ce que ces capacités d'ajustement des signaux émergent précisément au cours du développement (Fröhlich et al., 2018). Toutefois, notons que la majorité des études réalisées incluait des chimpanzés appartenant *a minima* à la catégorie des juvéniles (Hostetter et al., 2001, 2007; Kaminski et al., 2004; Leavens, Hopkins, et al., 2004; Leavens, Hostetter, et al., 2004; Leavens et al., 2010; Liebal, Call, Tomasello, et al., 2004; Pika & Liebal, 2006; Povinelli, Eddy, Hobson, & Tomasello, 1996; Tempelmann et al., 2011; Tomasello et al., 1994). Ainsi, on peut dire que la distribution des gestes silencieux-visuels (moins produits quand le destinataire n'est pas visuellement attentif que quand il l'est) semble être *a priori* présente dès le stade juvénile. Toutefois, cette contrainte perceptive va de fait impliquer une restructuration de la signalisation (i.e. utilisation de signaux des autres modalités sensorielles à savoir audibles et tactiles) et cette réorganisation modale n'est pas encore tout à fait renseignée dans la littérature et son développement l'est encore moins.

En s'intéressant à l'acquisition du répertoire multimodal des grands singes immatures, Schneider et al. (2012) ont montré que les signaux de différentes

modalités n'apparaissaient pas en même temps et que la part d'utilisation de ceux-ci variait au cours du développement (voir partie 1.4.3). Ainsi, il semblerait pertinent de davantage considérer la dynamique développementale sous-jacente aux capacités d'ajustement relatives à la *communication sensible à l'attention du destinataire*. Ceci nous permettrait d'alimenter nos connaissances quant à la réorganisation modale conséquente à l'ajustement des signaux à l'attention visuelle du destinataire, affinant ainsi notre « carte macroscopique » d'utilisation de la communication multimodale.

1.3.2. Chez l'enfant préverbal (*Homo sapiens*)

Les études qui appréhendent la *communication sensible à l'attention du destinataire* chez l'enfant sont moins nombreuses que celles réalisées sur d'autres espèces notamment celles sur les chimpanzés présentées ci-dessus. Elles se concentraient essentiellement sur des enfants ayant autour de l'âge d'un an et s'attelaient à quantifier les gestes de pointage, les vocalisations ainsi que les combinaisons pointage-vocalisation pour comparer leur utilisation d'une condition à l'autre. On distingue les études expérimentales (Igualada et al., 2015; Liszkowski et al., 2008; Wu & Gros-Louis, 2017), des études observationnelles réalisées en laboratoire (Gros-Louis & Wu, 2012; Wu & Gros-Louis, 2014) ou en crèche (Rodrigues et al., 2021).

Sur la Table 2 ci-dessous, sont répertoriés les résultats des études ayant testé la *communication sensible à l'attention du destinataire* chez l'enfant préverbal selon le design d'étude, i.e. expérimental (Exp.) ou observationnel (Obs.), mais aussi selon l'âge des individus, les signaux testés et la mesure de comparaison. Comme pour la Table 1, la mesure de comparaison indiquait si les auteurs testaient leur hypothèse avec des *distributions* ou des *proportions*, i.e. fréquences relatives (voir explications partie 1.3.1). Les études mentionnées sont développées dans les parties 1.3.2.1 et 1.3.2.2 et discutées au regard de l'approche développementale dans la partie 1.3.3.3.

Table 2. Résultats des études ayant testé la ***communication sensible à l'attention du destinataire*** **chez l'enfant préverbal selon le design d'étude, l'âge des enfants et les signaux testés.**

				Pointages		***Vocalisations***		***Combinaisons pointage-vocalisation***	
Design	*Etudes*	*Age (mois)*	*Mesures*	*Attentive*	*Inattentive*	*Attentive*	*Inattentive*	*Attentive*	*Inattentive*
Exp.	Liszkowski et al., 2008	12-18 (N=64)	Distributions	>				12 : ns 18 : <	
	Igualada et al., 2015	12 (N=19)	Distributions	ns		ns		>	
	Wu & Gros-Louis, 2017	10 (N=39)	Distributions	ns		>		ns	
Obs.	Gros-Louis & Wu, 2012	12 (N=30)	Proportions *Combi/Point.**					<	
	Wu & Gros-Louis, 2014	10-13 (N=51)	Proportions *Combi/Point.**					<	
(Autre répertoire)				***Gestes silencieux-visuels***		***Gestes audibles et tactiles***			
Design	*Etudes*	*Age*	*Mesures*	*Attentive*	*Inattentive*	*Attentive*	*Inattentive*		
Obs.	Rodrigues, 2021	7-12 (N=10)	Proportions	>		<			

*Note : les « * » indiquent que ces études comparaient la proportion de combinaisons pointage-vocalisations sur l'ensemble des gestes de pointage produits entre les conditions d'attention et d'inattention visuelle.*

1.3.2.1. *Etudes expérimentales*

Les études expérimentales qu'on pourrait rapporter à l'étude de la *communication sensible à l'attention du destinataire* (Igualada et al., 2015; Liszkowski et al., 2008; Wu & Gros-Louis, 2017) ont toutes été réalisées selon un même design expérimental à savoir une tâche de pointage déclaratif initialement proposée par Liszkowski et al. (2008). Durant l'expérience, l'enfant et l'adulte étaient placés l'un en face de l'autre et un objet attractif (censé attirer l'attention de l'enfant) passait à travers la pièce dans le dos de l'adulte. La dyade pouvait être soumise à trois conditions expérimentales définies selon le comportement de l'adulte complice, faisant suite au pointage de l'enfant produit lorsque l'objet traversait la pièce. Dans la (1) *condition de référence* (*baseline condition*; Liszkowski et al., 2008), l'adulte était visuellement attentif à l'enfant (il le regardait) et quand l'enfant pointait l'objet traversant dans le dos de l'adulte, ce dernier se retournait et orientait son attention visuelle vers l'évènement ou l'objet de référence, s'engageant ainsi conjointement avec l'enfant en ne s'adressant pas nécessairement à l'enfant. Dans la (2) *condition disponible* (*available condition* ; Liszkowski et al., 2008), l'adulte était visuellement attentif à l'enfant et quand celui-ci se mettait à pointer l'objet, l'adulte questionnait verbalement l'enfant sans jamais se retourner vers l'évènement ou l'objet de référence. Dans la (3) *condition indisponible* (*unavailable condition* ; Liszkowski et al., 2008), l'adulte était visuellement inattentif à l'enfant en prétextant de lire un livre et ne réagissait aucunement au pointage de l'enfant (ni par la parole, ni par le comportement).

Sur la base de ces mêmes conditions, les autres études expérimentales citées Table 2 avaient construit leur propre design expérimental (Igualada et al., 2015; Wu & Gros-Louis, 2017). Malgré certaines divergences méthodologiques (e.g. soumettre une même dyade aux trois conditions ou seulement à deux d'entre elles), ces conditions expérimentales avaient été testées sur des enfants âgés de 10 mois (Wu & Gros-Louis, 2017), de 12 mois (Igualada et al., 2015) et de 12 à 18 mois (Liszkowski et al., 2008). Toutes ces études cherchaient à comparer l'utilisation des

gestes de pointage seuls, de vocalisations seules et de combinaisons de pointage-vocalisation. Pour cela, il s'agissait d'observer la distribution de ces signaux d'une condition expérimentale à l'autre.

Wu & Gros-Louis (2015) ont montré que chez les enfants âgés de 10 mois, il n'y avait pas d'utilisation différentielle significative de gestes de pointage seuls et de combinaisons pointage-vocalisations d'une condition à l'autre. De plus, ils produisaient globalement plus de vocalisations seules quand la mère était visuellement attentive (*condition indisponible*) que quand elle ne l'était pas (*condition de référence* et *disponible*). Toutefois, le quart des enfants de ce même âge qui produisaient plus de vocalisations, co-occurrentes à un geste de pointage ou non, quand la mère n'était pas visuellement attentive (*condition indisponible*) que quand elle l'était (*condition de référence* et *disponible*), présentaient de meilleures performances de production langagière à l'âge de 15 mois. **Liszkowski et al. (2008)** ont réalisé cette même expérience en ne soumettant les enfants qu'à deux conditions : la *condition de référence* et une des deux *conditions disponible* ou *indisponible*. Les enfants âgés de 12 mois et 18 mois produisaient moins de gestes de pointages seuls quand les expérimentateurs étaient visuellement attentifs et répondaient (*condition de référence*) que non (*condition indisponible*). Aussi, seuls les enfants âgés de 18 mois produisaient plus de combinaisons pointage-vocalisations quand l'expérimentateur n'était pas visuellement attentif (*condition indisponible*) que quand il l'était (*condition de référence*). **Igualada et al. (2015)** ont montré que chez les enfants âgés de 12 mois, il n'y avait pas d'utilisation différentielle significative de gestes de pointage seuls et de vocalisations seules d'une condition à l'autre.

Toutefois, dans ce design expérimental, la variable d'attention visuelle de l'adulte n'avait pas été isolée puisqu'elle pouvait être confondue à celle du Langage Adressé à l'Enfant (LAE). En effet comme on peut le voir sur la Table 3, ces trois conditions expérimentales semblaient confondre trois variables : deux variables attentionnelles, i.e. d'*attention visuelle* et de *LAE* d'une part, et la variable de *responsivité* (i.e. engagement conjoint du parent suite au pointage).

Table 3. Décomposition des conditions expérimentales de l'expérience de Liszkowski et al. (2008) **selon l'état d'attention visuelle de l'adulte, le Langage Adressé à l'Enfant (LAE) et la responsivité de l'adulte.**

Condition	Attention visuelle	LAE	Responsivité
condition de référence	Oui	Oui/Non	Oui
condition disponible	Oui	Oui	Non
condition indisponible	Non	Non	Non

Cette confusion des variables attentionnelles et de responsivité limitent véritablement nos interprétations quant à l'étude de la communication sensible à l'attention (visuelle) du destinataire. Par exemple, ces études ont montré que les enfants âgés de 12 mois produisaient plus de combinaisons pointage-vocalisations en *condition disponible* (i.e. l'adulte était visuellement attentif, s'adressait verbalement à l'enfant et mais ne se retournait pas, voir Table 3) qu'en *condition de référence* (i.e. l'adulte était visuellement attentif, *pouvait* s'adresser à l'enfant et se retourne vers l'objet, voir Table 3 ; Igualada et al., 2015; Liszkowski et al., 2008). On pourrait alors se dire que c'est l'absence de responsivité qui pousse l'enfant à produire plus de combinaisons pointage-vocalisations. Pourtant, quand on compare la condition *indisponible* (i.e. l'adulte était visuellement inattentif, ne s'adressait pas à l'enfant et ne se retournait pas non plus, voir Table 3) à la *condition de référence* (i.e. l'adulte était visuellement attentif, *pouvait* s'adresser à l'enfant et se retourne vers l'objet, voir Table 3), on ne trouve aucune utilisation différentielle de signalisation (Igualada et al., 2015; Liszkowski et al., 2008). De plus, les résultats entre les conditions *disponible* et *indisponible* diffèrent d'une étude à l'autre tant pour les gestes de pointages seuls que combinés (Igualada et al., 2015; Liszkowski et al., 2008) . Ces derniers résultats soulignent donc l'importance de considérer de manière distincte ces trois variables (voir Table 3) lors de l'investigation de la *communication sensible à l'attention du destinataire* chez l'enfant.

1.3.2.2. *Etudes observationnelles*

Des études observationnelles réalisées sur les enfants au stade préverbal peuvent également être rapportées à l'étude de la *communication sensible à l'attention du destinataire*. Elles ont été réalisées en laboratoire (Gros-Louis & Wu, 2012; Wu & Gros-Louis, 2014) ou en crèche (Rodrigues et al., 2021) dans un contexte de jeu libre. Les études réalisées en laboratoire (Gros-Louis & Wu, 2012; Wu & Gros-Louis, 2014) se sont attelées à comparer la proportion de gestes de pointage combinés à des vocalisations (par rapport au nombre de pointages produits) d'une condition d'attention visuelle du parent à l'autre sur des enfants âgés de 10 à 13 mois. Toutes deux ont montré que les enfants combinaient plus leurs gestes de pointage à des vocalisations quand le parent n'était pas visuellement attentif que quand il l'était (Gros-Louis & Wu, 2012; Wu & Gros-Louis, 2014). De plus, il semble que ce pattern comportemental prédise de meilleures compétences de compréhension langagière à l'âge de 15 mois (Wu & Gros-Louis, 2014). A l'instar des études réalisées sur les primates non-humains (Byrne et al., 2017; Genty et al., 2009; Hobaiter & Byrne, 2011; Kersken et al., 2019), une étude récente a été réalisée sur une population d'enfants plus jeunes, âgés de 7 à 12 mois en crèche (Rodrigues et al., 2021). A la différence des autres études évoquées ci-dessus, ces travaux s'appuyaient sur un répertoire de gestes différents et ne considéraient pas les vocalisations. En effet, les gestes de communication relevés étaient catégorisés selon la modalité sensorielle associée (voir partie 1.4) et comparaient donc les proportions de chaque modalité de gestes (audibles, tactiles et silencieux-visuels) d'une condition visuelle à l'autre (Rodrigues et al., 2021). Cette étude rapporte que les enfants âgés de 7 à 12 mois produisaient significativement plus de gestes audibles mais aussi essentiellement tactiles quand le destinataire (d'âge variable, non nécessairement adulte) n'était pas visuellement attentif que quand il l'était (Rodrigues et al., 2021). Par ailleurs, qu'elles que soient les études développées ici, aucune d'entre elles ne montrent un effet significatif de l'âge sur la capacité à ajuster ses signaux à l'attention visuelle du destinataire (Rodrigues et al., 2021; Wu & Gros-Louis, 2014). Notons que la divergence méthodologique de ces études

observationnelles (i.e. répertoire comportemental considéré tenant compte de vocalisations ou non, le nombre d'enfants observés, le contexte d'observation et le profil du destinataire) limite leur comparabilité au regard de l'investigation de la *communication sensible à l'attention du destinataire* chez l'enfant.

1.3.2.3. ***Etudes développementales chez l'enfant*** *préverbal : divergences méthodologiques et de résultats*

La majorité des études, expérimentales et observationnelles, mentionnées ci-dessus s'intéresse aux mêmes signaux de communication (Gros-Louis & Wu, 2012; Igualada et al., 2015; Liszkowski et al., 2008; Wu & Gros-Louis, 2014, 2017) exceptée une étude réalisée selon un codage éthologique mais qui ne considérait que les gestes (Rodrigues et al., 2021). Les études expérimentales se sont globalement appuyées sur le même design expérimental de tâche de pointage déclaratif (Liszkowski et al., 2008) et comparent toutes la distribution des pointages, des vocalisations et la combinaison des deux, d'une condition à l'autre (Igualada et al., 2015; Liszkowski et al., 2008; Wu & Gros-Louis, 2017). Toutefois, certaines caractéristiques méthodologiques similaires (e.g. la non-distinction de différentes variables attentionnelles telles que l'attention visuelle et le LAE) ou divergentes (e.g. le fait de soumettre les enfants à deux ou trois conditions) rendent les résultats difficilement comparables et « compilables » pour retracer l'histoire développementale des capacités d'ajustement des signaux à l'attention du destinataire. En effet, aucun des résultats rapportés ne concordent (voire Table 2). Pour ce qui est des études observationnelles, elles divergent également d'un point de vue méthodologique mais leurs résultats sont davantage concordants en comparaison aux études expérimentales (Gros-Louis & Wu, 2012; Rodrigues et al., 2021; Wu & Gros-Louis, 2014). Malgré cela, ces études ne semblent pas trouver d'effet de l'âge sur ces capacités d'ajustement entre l'âge de 7 et 13 mois alors que c'est une période où la communication de l'enfant devrait connaître des

changements *modaux* majeurs (voir Figure 3, partie 1.4.2). De plus certaines différences méthodologiques (e.g. différences de répertoire, considération des vocalisations ou non, contexte d'observation) limitent également la construction d'un modèle développemental quant au développement des capacités d'ajustement des signaux relatives à la *communication sensible à l'attention du destinataire*.

En résumé, la *communication sensible à l'attention du destinataire* qui semblerait pertinente pour l'étude du *réagencement modal* relative à l'acquisition du langage chez l'enfant, questionne également le champ d'étude relatif aux origines du langage humain. Ce marqueur a bien été investigué chez de nombreuses espèces : humaines (Gros-Louis & Wu, 2012; Igualada et al., 2015; Liszkowski et al., 2008; Rodrigues et al., 2021; Wu & Gros-Louis, 2014, 2017) et non-humaines (*Papio* anubis: Bourjade et al., 2014; *Pan troglodytes*: Fröhlich et al., 2018; *Gorilla* gorilla: Genty et al., 2009; *Canis* familiaris: Horowitz, 2009; Kaplan, 2011; *Corvus corax*: Pika & Bugnyar, 2011). Toutefois, les résultats sont difficilement comparables les uns aux autres, en intra-espèce comme en inter-espèces. Chez l'enfant humain, on étudie essentiellement son développement alors que chez d'autres espèces, l'investigation se fait davantage sur des sujets plus matures et/ou d'âges très variables et l'effet développemental est très peu appréhendé. De plus, les divergences méthodologiques limitent la comparabilité des résultats et *in fine* la construction d'un modèle relatif à ce remaniement modal dont dépend ce marqueur. Il semblerait donc pertinent d'orienter davantage les études faites sur la communication de l'enfant préverbal vers *une* approche multimodale supportée par le codage éthologique.

1.4. VERS *UNE* APPROCHE MULTIMODALE : PERTINENCE DU CODAGE ETHOLOGIQUE

1.4.1. Principe du codage éthologique

Dans ce travail de thèse, *l'approche multimodale* consiste à catégoriser des signaux selon une modalité qu'on leur attribue. Selon la discipline concernée, le terme de *modalité* diffère (Fröhlich et al., 2019). Par exemple, en psychologie du développement et comparée, on considèrera la modalité d'un signal selon le canal de production. Ainsi on distinguera les gestes, les vocalisations et les expressions faciales (e.g. Murillo & Casla, 2020; Murillo et al., 2019). En éthologie, on considérera la modalité d'un signal à travers la modalité sensorielle de perception du destinataire. Dans les études sur la communication gestuelle du primate non-humain (*Gorilla gorilla*: Genty et al., 2009; *Pan troglodytes*: Hobaiter & Byrne, 2011a; *Pongo pygmaeus*: Liebal et al., 2006; *Papio anubis*: Molesti et al., 2017), les gestes sont catégorisés selon la modalité sensorielle de perception dominante. On y distinguera donc les gestes *silencieux-visuels*, *audibles* et *tactiles*. Il existe certaines divergences quant à la définition du *geste* (Bourjade et al., 2020) mais elles semblent toutefois s'accorder sur le fait que le geste constitue *a minima* un moyen de communication et possède une composante visuelle. Toutefois, si le geste produit implique un son (e.g. taper dans les mains ou sur une table) ou un contact physique (e.g. une tape dans le dos), ils seront respectivement considérés comme étant des gestes *audibles* ou *tactiles*. Dans le cadre de ce travail, on reconnaît ces deux approches de la *modalité*, c'est-à-dire que, comme la psychologie du développement, elle accordera une importance au canal de production (geste, vocalisation, expression faciale) mais aussi à la modalité sensorielle de perception pour appréhender la communication non-verbale. Dans la continuité de cette catégorisation, si l'on y intègre les signaux oraux telles que les vocalisations dans cette étude, on considèrera que ces signaux sont *principalement* audibles (voire *essentiellement* audibles dans le cas d'un chantonnement avec bouche fermée). Ces vocalisations feront donc partie des signaux audibles du répertoire comportemental. Par ailleurs, concernant les gestes

silencieux-visuels les études divergent quant aux signaux qu'on peut qualifier de gestes ou non (e.g. les expressions faciales, les mouvements corporels non-manuels ; Bourjade et al., 2020). C'est pourquoi tout au long de ce travail les gestes manuels ou non et expressions faciales seront qualifiés de *signaux* (et non de gestes) silencieux-visuels.

Selon la discipline concernée, le terme de *multimodalité* diffère (Fröhlich et al., 2019). En *psychologie comparée*, la modalité du signal faisant référence au canal de production, la multimodalité implique qu'il y ait au moins deux types de signaux produits (i.e. geste et expression faciale, vocalisation et expression faciale ou geste et vocalisation ; Fröhlich et al., 2019). En *éthologie*, la modalité du signal faisant référence à la modalité sensorielle de perception du destinataire, on distingue deux types de multimodalités : *physique* et *comportementale* (Fröhlich et al., 2019). La multimodalité *physique* fait référence à la multimodalité inhérente au signal pour lequel les modalités sont couplées (Fröhlich et al., 2019). Par exemple, lorsqu'un individu s'adresse verbalement à son destinataire en face à face, le destinataire pourra à la fois voir ses lèvres bouger et entendre son discours. La multimodalité *comportementale* fait référence à la combinaison (production synchrone) d'au moins deux signaux distincts appartenant à deux modalités différentes (Fröhlich et al., 2019). Dans ce travail le terme de *multimodalité* sera exclusivement utilisé en termes de *combinaison multimodale comportementale*, e.g. la combinaison d'un signal silencieux-visuel (e.g. geste de pointage) et d'un signal audible vocal (e.g. le babillage) ou gestuel (e.g. taper sur la table avec une cuillère), tout en tenant compte du canal de production de chaque signal combiné.

1.4.2. Mise en application sur les travaux existants chez l'enfant préverbal

De nombreuses études en psychologie du développement de l'enfant ont investigué la communication gestuelle et/ou vocale de l'enfant au stade préverbal. Elles cherchaient pour la plupart à alimenter une trajectoire développementale

conduisant à l'émergence de la parole (signal audible) via l'étude du *pointage* notamment (signal silencieux-visuel). Peu d'entre elles considéraient les signaux manipulés à travers la modalité sensorielle perçue, à l'instar du codage éthologique (i.e. catégoriser l'ensemble des signaux d'un répertoire comportemental selon la modalité sensorielle dominante perçue à savoir silencieuse-visuelle, audible ou tactile). Toutefois, même si la catégorisation des signaux diffère d'une étude à l'autre, il est possible de rapporter les résultats de nombreuses études réalisées au principe du codage éthologique. C'est pourquoi, il s'agit ici d'appliquer le principe du codage éthologique sur les travaux existants chez l'enfant préverbal afin de mettre en évidence *une* trajectoire de développement sous le prisme de la modalité de signal en procédant de façon chronologique.

1.4.2.1. Développement du répertoire : émergence, combinaisons multimodales et regards

1.4.2.1.1. Emergence des signaux

Les signaux audibles

Sur le principe du codage éthologique, les signaux audibles sont principalement appréhendés à travers l'étude des vocalisations. A cet égard, les études s'accordent à reconnaître l'importance des vocalisations prélinguistiques dans le développement communicationnel de l'enfant menant aux premiers mots Oller et al., 1999). Elles émergent selon 4 stades qui se chevauchent plus ou moins: 1) *phonatoire*, 2) *d'articulation primitive*, 3) *d'élargissement* et 4) *canonique* (Oller et al., 2016). Le stade *phonatoire* recouvre les deux premiers mois au cours duquel l'enfant produira des sons non-articulés quasi-vocaliques (i.e. de quasi-voyelles; Oller, 2000; Oller et al., 2016). Le stade *d'articulation primitive* s'étend de l'âge de 1 à 4 mois. L'enfant articule progressivement les sons quasi-vocaliques notamment par l'utilisation de roucoulement (Oller, 2000; Oller et al., 2016). Le stade *d'élargissement* s'étend de l'âge de 3 à 8 mois durant lequel l'enfant commence véritablement à produire des sons vocaliques (Oller, 2000; Oller et al., 2016) et ne réussit pas encore à transiter d'un son vocalique à un son consonnique rapidement pour produire des

syllabes. Jusque-là, ces vocalisations prélinguistiques sont considérées comme exploratoires (Oller, 2000; Oller et al., 2016). Ce troisième stade chevauche toutefois le dernier stade *canonique* qui s'étend de l'âge de 5 à 10 mois durant lequel les enfants produisent des syllabes bien formées (*canonical babbling*; Oller, 2000; Oller et al., 2016). Ces quatre stades mènent à l'utilisation prédominante du babillage. Le babillage est la vocalisation prélinguistique la plus sophistiquée en comparaison aux autres. En effet, elle se caractérise par l'utilisation de syllabes composées d'au moins un élément vocalique et un élément consonnique et d'une transition rapide de l'un vers l'autre (Oller et al., 1999). De plus, elle apparaît relativement tard par rapport aux autres vocalisations non-syllabiques autour de 6-7 mois et constitue ainsi un intérêt certain pour de nombreuses études (Esteve-Gibert & Guellaï, 2018a; Oller et al., 1999; Stoel-Gammon, 2011). En effet possédant des caractéristiques relativement proches de la parole (par l'utilisation de syllabes), le babillage sera plus facilement perçu comme étant communicatif par l'adulte. Il constitue ainsi un marqueur de maîtrise vocale de l'enfant préverbal qui serait un prédicteur langagier (Donnellan et al., 2019; Jensen et al., 1988; Oller et al., 1998; Oller et al., 1999; Stoel-Gammon, 2011; Stolt et al., 2012).

Les signaux silencieux-visuels

Sur le principe du codage éthologique, les signaux silencieux-visuels sont principalement appréhendés à travers les études s'intéressant au *sourire* (e.g. Colonnesi et al., 2012; Hsu et al., 2001; Lin & Green, 2009; Northrup & Iverson, 2020; Yale et al., 2003) aux *mouvements corporels* (Iverson & Fagan, 2004; Nakata & Trehub, 2004; Scola, Bourjade, & Jover, 2015) mais aussi et surtout aux *gestes déictiques* (e.g. Bates et al., 1975; Beuker et al., 2013; Cameron-Faulkner et al., 2015; Carpenter et al., 1998; Choi et al., 2021; Cochet & Vauclair, 2010; Franco & Butterworth, 1996; Moreno-Núñez et al., 2020). Autour de 2 mois, émergent les premiers sourires *sociaux* (autres que réflexes) dans divers contextes socio-culturels (Kilbride, 1980; Super & Harkness, 2015; Wörmann et al., 2014). Ces *sourires* constituent les premiers signes de développement socio-émotionnel marquant les premiers mois de l'enfant. Les *mouvements corporels* (mouvements répétitifs des

membres et/ou du tronc) sont relativement peu considérés dans les études relatives au développement communicationnel de l'enfant. Toutefois, certaines études montrent qu'ils sont d'ores et déjà bien présents chez l'enfant âgé de 5 mois (Nakata & Trehub, 2004; Scola et al., 2015). Autour de 9-10 mois, on voit émerger les premiers gestes *symboliques* (i.e. gestes qui représentent un objet, une action ou un évènement) et *conventionnels* (gestes résultant d'un apprentissage au travers de routines sociales tel que le « coucou de la main » ou des applaudissements ; Cohen & Billard, 2018; Guidetti, 2002, 2005; Murillo et al., 2018). A partir de cette période, commencent à émerger les *gestes déictiques*. On y distingue les gestes déictiques *proximaux*, i.e. les gestes faisant référence à un objet que le signaleur tient dans sa/ses mains (e.g. *giving, showing*) et gestes déictiques *distaux*, i.e. gestes faisant référence un objet à distance du signaleur (e.g. *whole-hand/index-finger pointing*, *reaching* ; Beuker et al., 2013; Murillo et al., 2021). Les gestes déictiques *proximaux* et le geste déictique distal de *reaching* (i.e. tentative de préhension) émergeraient à l'âge de 9-10 mois juste avant le geste de pointage à 10-11 mois (Bates et al., 1975; Cameron-Faulkner et al., 2015; Choi et al., 2021). Parmi les gestes *proximaux*, les enfants utilisent préférentiellement les *showing* jusqu'à l'âge de 13 mois (Moreno-Núñez et al., 2020). Parmi les gestes *distaux*, les enfants produisent préférentiellement les gestes de *pointage* à partir de 12 mois (Cameron-Faulkner et al., 2015; Franco & Butterworth, 1996). Par ailleurs, une étude expérimentale a montré une corrélation positive entre la fréquence d'utilisation des gestes *proximaux* chez les enfants âgés de 10-11 mois et la fréquence d'utilisation de gestes de pointage (*whole-hand/index-finger pointing*) et non de *reaching* (Cameron-Faulkner et al., 2015). L'émergence ainsi que l'utilisation en contexte social de ces gestes déictiques ont principalement été étudiées dans le développement de la communication triadique (voir partie 1.4.2.2.2).

Les signaux tactiles

Sur le principe du codage éthologique, l'étude des signaux tactiles reviendrait à considérer des études considérant des gestes impliquant un contact physique avec le destinataire. A notre connaissance, assez peu d'études ont véritablement

investigué l'émergence de ces signaux mais davantage mis l'accent sur l'utilisation de ceux-ci (Kersken et al., 2019; Rodrigues et al., 2021). Notons, qu'aucune d'entre elles ne semble trouver un effet de l'âge sur l'utilisation de ces signaux sur les enfants âgés de 7 à 24 mois

1.4.2.1.2. Combinaisons multimodales : audio-visuelles

Dans la continuité de ces travaux, l'utilisation des combinaisons multimodales (cooccurrence de signaux audibles et silencieux-visuels) a elle aussi été investiguée. Les études sur les enfants âgés de moins de 9 mois portent globalement sur la *coordination* (synchrone ou diachronique) des sourires ou mouvement corporels et des vocalisations (Hsu, Fogel, & Messinger, 2001; Iverson & Fagan, 2004; Lin & Green, 2009; Yale et al., 2003). Les études sur les enfants âgés de plus de 9 mois portent globalement sur la *combinaison* (synchrone) des gestes déictiques et des vocalisations verbales ou non (Aureli et al., 2017; Fasolo & D'Odorico, 2012; Iverson & Fagan, 2004; Iverson & Goldin-meadow, 2005; McCune et al., 2020; Murillo & Casla, 2020; Murillo et al., 2018).

Les études sur la *coordination* conçoivent celle-ci comme le fait de produire séquentiellement des signaux se chevauchant temporellement ou non dans le temps. La coordination entre sourires et vocalisations apparaîtrait relativement tôt autour de l'âge de 2 mois (Hsu et al., 2001; Yale et al., 2003). Ce n'est qu'à partir de l'âge de 7 mois que l'enfant combinera significativement ses sourires à des vocalisations (Lin & Green, 2009). Par ailleurs, une étude montre qu'entre 6 et 9 mois, les enfants coordonneront de plus en plus leurs mouvements de bras (plutôt que de tronc ou de jambes) avec préférentiellement des babillages (Iverson & Fagan, 2004). De plus, la précision de la synchronisation de la combinaison vocalisations-mouvement est bien meilleure après émergence du babillage qu'avant (Iverson & Fagan, 2004).

Les études sur la *combinaison* des gestes déictiques et des vocalisations sont bien plus nombreuses (Aureli et al., 2017; Fasolo & D'Odorico, 2012; Iverson & Fagan,

2004; Iverson & Goldin-meadow, 2005; McCune et al., 2020; Murillo & Casla, 2020; Murillo et al., 2018). Parmi ces études, Murillo et al. (2018) montrent que dès l'âge de 9 mois, les enfants produiront significativement plus de *reaching* (i.e. tentative de préhension) avec des vocalisations que sans (Murillo et al., 2018). Ils commenceront également à combiner les gestes de *showing*, conventionnels et symboliques avec des vocalisations mais préfèreront les utiliser sans vocalisations (Murillo et al., 2018). A l'âge de 12 mois, l'enfant commence à combiner ses gestes de *giving* et de *pointage* avec des vocalisations mais préfèreront les utiliser sans vocalisations (Murillo et al., 2018). A partir de 15 mois, les gestes de *pointage* seront les seuls gestes déictiques à être préférentiellement produits avec des vocalisations jusqu'à ce que les gestes de *showing* suivent cette même tendance à l'âge de 18 mois (Aureli et al., 2017; Murillo et al., 2018). Parallèlement à cela, le babillage, le *mot unique* et les *successions de mots* commenceront à être combinés aux gestes déictiques ou non (i.e. symbolique, conventionnels et autres) aux âges respectifs de 9, 12 et 18 mois. Toutefois, contrairement aux gestes déictiques mentionnés plus haut, ces vocalisations seront préférentiellement utilisées sans gestes jusqu'à 18 mois (Murillo et al., 2018).

1.4.2.1.3. Coordination des signaux de communication avec le regard

Dans cette littérature, les coordination (synchrone ou diachronique) des signaux avec le regard occupe une place centrale. Avant l'âge de 8-9 mois, on parlera essentiellement *d'engagement visuel* (regard du signaleur en direction du destinataire), la littérature recouvrant cette période s'intéressera notamment aux temps de fixation et à la coordination de ceux-ci avec d'autres signaux tels que les sourires par exemple à l'âge de 4 mois (Colonnesi et al., 2012; Lin & Green, 2009; Yale et al., 2003) dans les interactions avec la mère. A partir de 9 mois, la coordination du regard semble se faire en plusieurs temps que ce soit pour les vocalisations (D'Odorico et al., 1997) ou les gestes déictiques (Franco & Butterworth, 1996). Pour ce qui est des signaux *audibles*, les enfants tendraient à

regarder leur destinataire plutôt avant puis après avoir produit leurs vocalisations aux âges respectifs de 12 mois et 20 mois (D'Odorico et al., 1997). Pour ce qui est des signaux *silencieux-visuels*, les enfants tendraient à regarder leur destinataire plutôt après, pendant et avant avoir produit leurs gestes déictiques aux âges respectifs de 12, 14 et 16 mois (Franco & Butterworth, 1996). Toutefois, lorsqu'on s'attèle à distinguer les signaux à l'intérieur de chaque modalité, on se rend compte que la coordination se fait également en plusieurs temps. Donnellan et al. (2019) ont montré qu'à l'âge de 11 mois, les enfants coordonnaient leur regard aux vocalisations sauf au babillage. Ils montrent également qu'à cet âge-là, les enfants coordonnaient leur regard aux gestes déictiques *proximaux* (*showing, giving*) mais pas aux gestes *distaux* de pointages (*whole-hand/index-finger pointings*; Donnellan et al., 2019). Aussi, ils montrent que globalement les combinaisons multimodales de gestes déictiques sont elles aussi associées à des regards dirigés vers le destinataire mais pas les combinaisons impliquant le pointage (Donnellan et al., 2019). Ces travaux suggèrent ainsi que la coordination du regard serait fonction de la modalité sensorielle de perception du signal, de l'âge et de la nature du signal.

Sur la Figure 3 ci-dessous, sont représentées les frises chronologiques d'émergence des signaux audibles et silencieux-visuels et d'apparition des différentes combinaisons audio-visuelles. Comme on peut le voir, le développement des signaux audibles et celui des silencieux-visuels se fait conjointement dès la naissance.

Durant le stade *perlocutoire* (avant 10 mois ; Bates et al., 1975), il y a une notable évolution des vocalisations avec l'émergence du babillage autour de 6-7 mois ; ce qui représente une *première* étape « phare » de la signalisation (1). Conjointement à cette évolution du répertoire vocal, certains signaux silencieux-visuels tels que des expressions faciales à valence émotionnelle fixe (e.g. le sourire) ou les mouvements corporels émergeront. A la fin de ce stade perlocutoire, émergent les gestes déictiques avec une transition rapide vers l'utilisation de combinaisons multimodales avec les vocalisations.

Une fois passé au stade *illocutoire* (à partir de 10 mois ; Bates et al., 1975), le geste déictique de pointage sera préférentiellement utilisé (en comparaison aux autres gestes déictiques) à l'âge de 13 mois (Moreno-Núñez et al., 2020); ce qui représente une *deuxième* étape « phare » de la signalisation (2). Puis, ce geste de pointage sera préférentiellement combiné à des vocalisations à l'âge de 15 mois, marquant ainsi un *pic de multimodalité* (Aureli et al., 2017; Murillo et al., 2018) ; ce qui représente une *troisième* étape « phare » de la signalisation (3). La précision de la synchronisation entre pointage et vocalisation s'affinera jusqu'à l'âge de 18 mois (Murillo et al., 2018) âge à partir duquel les 'propositions verbales' (successions de mots) seront de plus en plus combinées aux gestes déictiques jusqu'à 30 mois (Murillo & Casla, 2020).

Figure 3. Frises chronologiques d'émergence des signaux audibles et silencieux-visuels et d'apparition de combinaisons audio-visuelles. Etant donné le peu d'études recensées existantes quant à l'émergence des signaux tactiles, il semblait difficile de les positionner sur une frise chronologique. Les numéros entre parenthèses représentent les étapes « phares » explicitées dans le texte présentant le graphique.

1.4.2.2. ***Développement de l'utilisation sociale de ces signaux***

Tous ces signaux combinés ou non présentent un grand intérêt dans l'étude du développement communicationnel de l'enfant au stade préverbal puisqu'ils sont associés à des capacités prélinguistiques nécessaires à l'émergence du langage. Ainsi, leur utilisation en contexte est également étudiée. Avec l'émergence des gestes déictiques à l'âge de 9 mois (juste avant le stade *illocutoire*, voir Figure 3) , les études qui s'intéressaient jusque-là à la communication dyadique, s'intéresseront davantage à la *communication triadique* notamment, avec l'étude de l'*attention conjointe* (e.g. Beuker et al., 2013; Carpenter et al., 1998) et de la *communication sensible à l'attention du destinataire* (Gros-Louis & Wu, 2012; Igualada et al., 2015; Liszkowski et al., 2008; Wu & Gros-Louis, 2014, 2017).

1.4.2.2.1. La communication dyadique

La communication dyadique entre l'enfant et son parent a principalement été étudiée par des études se rattachant à *la théorie de l'intersubjectivité* (Trevarthen & Aitken, 2003) . Celle-ci défend l'idée que dès la naissance, il existe une coordination rythmique yeux-mains-bouche (*eye-hand-mouth coordination*) en contexte de *face-to-face*. Cette coordination serait ainsi l'expression d'une conscience active et d'une intention communicative précoce allant de pair avec le développement socio-émotionnel. A l'instar de cette théorie, de nombreux travaux ont montré la coordination (synchrone ou diachronique) des regards et de signaux communicatifs tels que les expressions faciales et des vocalisations dans les *face-to-face* avec la mère (e.g. Beebe et al., 2016; Lavelli & Fogel, 2005). Par ailleurs, quand la mère s'adresse (en parlant ou chantant) à l'enfant âgé de 5 à 9 mois, ceux-ci tendent à réduire la quantité de mouvements corporels, suggérant ainsi que l'enfant commence à délimiter des épisodes interactifs (Nakata & Trehub, 2004; Scola et al., 2015) autrement qu'avec des vocalisations. De plus, dans cette même tranche d'âge, des expériences s'inscrivant dans le paradigme du *still-face* (Tronick, 2003;

Tronick et al., 1978) ont montré que les enfants étaient sensibles à la rupture interactionnelle (et attentionnelle). En effet, quand l'adulte (mère ou expérimentateur) interrompait les épisodes de *face-to-face* en se montrant impassible (i.e. visage neutre, sans s'adressait à l'enfant), l'enfant produisait plus de signaux audibles (vocalisations hors pleurs) et moins de signaux silencieux-visuels (sourires ; Delgado et al., 2002; Franklin et al., 2014; Goldstein et al., 2009). Ce qui suggère que le choix de la *modalité* du signal produit par l'enfant puisse être fonction de l'attention du destinataire de façon relativement précoce.

1.4.2.2.2. La communication triadique

Comme mentionné ci-dessus, à partir de l'âge de 8-9 mois, les études s'intéressent notamment à l'émergence de l'attention conjointe avec l'utilisation des gestes déictiques en contexte d'interaction. L'*attention conjointe* se définit par le fait de coordonner sa propre attention avec l'attention d'autrui vers une entité externe (Bourjade, 2017; Carpenter et al., 1998; Klein et al., 2009; Scaife & Bruner, 1975) : (1) *partager* son attention (e.i. alterner son regard entre l'objet et le partenaire d'interaction) 2) *suivre* l'attention d'autrui (e.i. suivre son regard ou son geste de pointage) et 3) *orienter* l'attention d'autrui (*sharing, following and directing attention*; Beuker et al., 2013). Cet engagement triadique est considéré comme un marqueur de la communication référentielle (Cochet & Vauclair, 2010; Colonnesi, Stams, Koster, & Noom, 2010; Liszkowski, Carpenter, & Tomasello, 2007) même si la capacité à abstraire ou à se représenter un objet non visible se développe ultérieurement (McCune & Zlatev, 2015; Murillo & Casla, 2020). Les gestes de pointages ont également été étudiés dans le cadre de l'émergence de l'intentionnalité (voir partie 1.2.2.2) et la *communication sensible à l'attention du destinataire* dont l'investigation est l'objet central de cette thèse (voir partie 1.3.2). Notons toutefois que ces études se sont essentiellement intéressées à l'effet de l'attention du destinataire sur les gestes déictiques (combinés ou non) de l'enfant alors que ceux-ci possèdent les mêmes caractéristiques physiques de production

(i.e. gestuelle) et de perception (i.e. signaux silencieux-visuels) que d'autres signaux ayant émergé plus précocement tels que le sourire ou les mouvements corporels.

1.4.3. Intégration dans une perspective comparative

Comme il a été défendu précédemment l'approche multimodale constitue un véritable « outil d'investigation » quant à l'étude de la *communication sensible à l'attention du destinataire* (voir partie 1.3). Initiée par les études sur des espèces non-humaines, elle n'est jusqu'alors pas suffisamment considérée dans les études sur le développement de la communication de l'enfant préverbal. Pourtant cette approche pourrait véritablement servir la perspective comparative et développementale.

Dans la perspective comparative, il subsiste un véritable besoin méthodologique d'optimiser la comparabilité de différentes études. En comparaison aux autres espèces de primates, la période de maturation humaine est plus longue (Bründl et al., 2021; Hawkes et al., 2017). Il subsiste de nombreux questionnements sur les traits comportementaux dont le développement est significativement plus long. Dans cette perspective, une étude récente sur des chimpanzés sauvages âgés de 1 à 36 mois a donc cherché à investiguer *quand* et *comment* se développaient quatre traits comportementaux : la motricité globale (*gross motor*), la motricité fine (*fine motor*), la socialisation (*social interaction*) et la communication audible (*communication traits*; Bründl et al., 2021). Les traits qu'il a été le moins facile à comparer (entre chimpanzés et humains) étaient les traits relatifs à la communication vocale et la motricité fine (dans lequel on retrouve l'utilisation de signaux silencieux-visuels). Ceux-ci se développent plus tardivement que les autres et montrent une plus grande variabilité interindividuelle quelle que soit l'espèce considérée (Bründl et al., 2021). Les auteurs suggèrent ainsi que ces traits soient les plus dépendants de l'expérience notamment liée aux modalités d'interaction avec la mère (Bard, 1994). De façon intéressante, ces résultats

soulignent à la fois la difficulté et la nécessité de comparer ces deux espèces. Aussi, elle donne des arguments qui appuient le fait que l'étude du développement communicationnel, de par l'importance de l'expérience notamment avec la mère, peut être pertinemment observable au niveau comportemental et interprétable dans une perspective comparative. D'où la nécessité d'utiliser un outil d'investigation optimisant la démarche.

L'approche multimodale constitue un outil d'investigation optimal puisqu'elle peut être retranscrite sur des travaux existants chez l'enfant préverbal. Toutefois, notons que deux études récentes (Kersken et al., 2018; Rodrigues et al., 2021) avaient cherché à faciliter la comparabilité de chimpanzés et humains en étudiant la communication de l'enfant préverbal de 7 à 24 mois sur la base d'un *répertoire* de gestes de grands singes (Byrne et al., 2017; Genty et al., 2009; Hobaiter & Byrne, 2011; Kersken et al., 2019). Kersken et al. (2018) ont ainsi pu comparer le répertoire de chimpanzés et d'enfants humains de 12 à 24 mois à partir des mêmes signaux. Bien que les résultats rapportés facilitent la comparabilité des espèces, ces études semblent difficilement « compilables » à tout un champ de la littérature de la psychologie du développement (voir partie 1.4.2.1) de par la difficulté à faire correspondre les termes et les méthodes d'observation. Pourtant, comme il a été montré plus haut (partie 1.4.2), les signaux étudiés dans la littérature existante peuvent assez facilement être rapportés à l'approche multimodale (i.e. catégoriser les signaux selon leur modalité sensorielle de perception). L'étude comparative ne dépendrait donc pas nécessairement d'un répertoire de signaux similaire pour y voir des différences développementales intéressantes.

Une première confrontation des travaux existants entre chimpanzés et humains souligne d'ores et déjà des pistes d'investigations pertinentes au regard de l'approche *evo-devo*. Les études menées sur le développement communicationnel du chimpanzé ont montré que l'émergence des différents signaux communicatifs selon la modalité sensorielle à laquelle ils sont associés présentait une certaine

histoire développementale. Schneider et al. (2012) ont montré que les premiers gestes communicatifs émergent aux alentours de 10 mois chez les chimpanzés. Ils seront d'abord principalement *tactiles* et *silencieux-visuels* puis les signaux *audibles* émergent bien plus tard aux alentours de 18 mois (Schneider et al., 2012). Aussi, tandis que l'utilisation des gestes tactiles s'amoindrit, l'utilisation des gestes silencieux-visuels et audibles s'intensifie avec l'âge (Fröhlich, Wittig, & Pika, 2016; Schneider et al., 2012). Les premières vocalisations émergeraient bien avant des premiers gestes puisque dès la naissance (i.e. les *whimpers,* les *screams*, qu'on peut assimiler à des pleurs chez l'enfant humain et les *grunts* qui ne seront socialement dirigés qu'à partir de 2 mois ; Bründl et al., 2021; Laporte & Zuberbühler, 2011). Par ailleurs l'utilisation du *grunt* serait utilisée de façon de plus en plus flexible à travers les différents contextes dès la première année (Dezecache et al., 2020) tout comme les protophones chez l'enfant préverbal (Oller et al., 2013) et suivront une véritable histoire développementale jusqu'à l'âge adulte (Laporte & Zuberbühler, 2011). Globalement, les chimpanzés produisent plus de gestes et vocalisations de façon isolée que de combinaisons des deux (Hobaiter et al., 2017). Toutefois, il semble y avoir un véritable effet de l'âge sur l'importance d'utilisation de gestes et de vocalisations mais pas de combinaisons geste-vocalisations (Hobaiter et al., 2017). A partir du stade juvénile, les individus produiraient de moins en moins de gestes et de plus en plus de vocalisations jusqu'à l'âge adulte où ils augmenteraient leur utilisation de gestes (Hobaiter et al., 2017). Chez l'enfant préverbal, les vocalisations émergent elles aussi bien avant les gestes manuels (voir partie 1.4.2.1.1). Aucun geste déictique (donc manuel) ne sera préférentiellement combiné à des vocalisations avant l'âge de 12 voire 15 mois alors que les vocalisations seront préférentiellement utilisées sans gestes jusqu'à 18 mois (Murillo et al., 2018). Ce qui souligne d'ores et déjà une différence en termes d'articulation des gestes et des vocalisations et l'importance de la combinaison geste-vocalisation dans la transition vers la communication *mature* de chaque espèce. Ainsi, retranscrire les résultats des études existantes à travers l'approche

multimodale permet d'ores et déjà de mettre en évidence des différences développementales qui appellent à être investiguées.

En résumé, comme il avait été discuté précédemment (voir partie 1.3) l'étude de la *communication sensible à l'attention du destinataire* semble pertinente pour étudier le réagencement modal menant à l'émergence et l'acquisition du langage au regard de l'approche *evo-devo*. Partant de la contrainte perceptive que la *communication sensible à l'attention du destinataire* implique, son investigation appellerait à l'approche multimodale (i.e. considérer les signaux à travers sa modalité sensorielle de perception). L'application de cette approche multimodale aux travaux existants sur l'enfant préverbal nous aura permis d'avoir une vision plus macroscopique du développement multimodal de l'enfant humain et de mettre en lumière les « pièces manquantes » de la fresque de connaissances que nous avons du développement communicationnel de l'enfant (e.g. déséquilibre entre l'ensemble des signaux émergeants au stade préverbal et les signaux dont l'utilisation sociale est étudiée, des questionnements quant aux mécanismes sous-jacents à l'acquisition de la parole chez l'enfant ; discutés partie 1.5). Par ailleurs, une première confrontation des travaux existants sur les primates immatures humains et non-humains souligne d'ores et déjà la pertinence de l'approche développementale dans l'investigation du *réagencement modal* à l'échelle de l'espèce.

1.5. PROBLEMATIQUE

1.5.1. Questions de recherche

A travers ce travail de thèse, il s'agit d'utiliser une approche multimodale pour étudier **le développement de la *communication sensible à l'attention du destinataire* dans une perspective comparative** afin de renseigner davantage les modèles théoriques sur (i) l'émergence du langage à travers l'approche comparative et (ii) l'acquisition du langage chez l'enfant préverbal âgé de 7 à 20 mois. En effet,

l'étude de la *communication sensible à l'attention du destinataire* semble pertinente pour investiguer le réagencement modal qui questionne à la fois l'étude sur les origines du langage humain et l'étude sur le développement communicationnel de l'enfant. L'approche *evo-devo* (voir partie 1.1.3) pourrait servir cette double-investigation. Il s'agira donc ici de répondre aux questions ci-dessous :

- **Le primate immature ajuste-il ses signaux à l'attention visuelle maternelle et comment développe-t-il cette capacité d'ajustement ?**

La *communication sensible à l'attention du destinataire* a été investiguée chez le chimpanzés (*Pan troglodytes*) et l'enfant préverbal (*Homo sapiens*) comme étayée partie 1.3. Toutefois, les résultats de l'ensemble de ces études sont d'ores et déjà difficilement comparables les uns aux autres à l'intérieur d'une même espèce. En effet, qu'elle que soit l'espèce, il semble y avoir certaines incohérences entre les résultats rapportés (voir partie 1.3, tables 1 et 2). Cette discordance de résultats pourrait avoir été causée par certaines divergences méthodologiques (e.g. design expérimental ou non, profil du destinataire, mesure de *distribution* ou de *proportion* comparée d'une condition à l'autre…). Chez le chimpanzé, l'investigation se fait davantage sur des sujets adultes ou d'âges très variables et l'effet développemental n'est que trop peu appréhendé. Chez l'humain, on l'étudie essentiellement dans une perspective développementale chez les enfants préverbaux. Toutefois, les études sont relativement peu nombreuses et restreignent leurs analyses au geste de pointage et aux vocalisations qui ne représentent qu'une partie de l'ensemble des signaux émergeants pendant la période non-verbale (voir partie 1.4.2.1.1). De plus, les âges testés recouvrent une période relativement courte entre 10 et 13 mois (voir partie 1.3.2, tables 1) pour la plupart et l'effet de l'âge n'a encore été que trop peu renseigné alors que ce sont des périodes marquées par d'importants changements communicationnels (voir partie 1.4.2.1). Par ailleurs, comme souligné partie 1.3.2, certaines variables attentionnelles (e.g. le Langage Adressé à l'Enfant) pourraient avoir influencé l'ajustement à l'attention visuelle retrouvées dans certaines études.

Ce premier point sera investigué dans les études réalisées sur de jeunes chimpanzés (Article 1 ; partie 3) et des enfants préverbaux (Article 2 et 3 ; partie 4).

- **Dans une perspective développementale, ces capacités d'ajustement constituent-elles des prédicteurs langagiers chez l'enfant préverbal ?**

On sait que l'enfant passe par différentes étapes clés de développement avant d'acquérir le langage (voir partie 1.4.2). Si ces capacités d'ajustement participent au réagencement modal nécessaire à l'acquisition du langage chez l'enfant, il est possible que ces ajustements prédisent des compétences langagières. Par ailleurs, on sait d'ores et déjà que parmi les signaux à ajuster, qu'ils soient *audibles* (e.g., Oller et al., 1999; Stoel-Gammon, 2011), *silencieux-visuels* avec le geste de pointage notamment (e.g., Beuker et al., 2013; Cochet & Byrne, 2016) ou *combinés* (Fasolo & D'Odorico, 2012; Iverson & Goldin-Meadow, 2005), certains constituent des prédicteurs langagiers. De ce fait, leur ajustement pourrait plausiblement constituer un prédicteur potentiel et si c'était le cas, l'ajustement d'autres signaux possédant des caractéristiques perceptives similaires pourraient l'être également. A notre connaissance, seules deux éudes ont testé et montré que l'ajustement des vocalisations à 10-13 mois prédisaient de meilleures performances langagières à l'âge de 15 mois (Wu & Gros-Louis, 2014, 2017). Toutefois, les auteurs comme bien d'autres (voir partie 1.3.2) limitent leurs analyses à un répertoire restreint aux gestes de pointages seuls ou combinés à des vocalisations. De plus, ces études portaient sur une tranche d'âge relativement limitée autour de l'âge d'un an (Wu & Gros-Louis, 2017), ce qui réduit l'investigation de l'effet de l'âge sur l'effet de prédiction.

Ce second point sera investigué dans l'étude sur questionnaire réalisée sur les parents des enfants préverbaux (Article 4 ; partie 5).

- **Dans une perspective comparative, quelles différences développementales inter-espèces retrouve-t-on ?**

Les répertoires de signaux communicatifs de chacune des espèces étudiées ici (humaine et de chimpanzés) contiennent à la fois des gestes et des vocalisations (*Pan trodlogytes*: Crockford & Boesch, 2005; Hobaiter et al., 2017; *Homo sapiens*: Liszkowski et al., 2008; Oller, 2000). Toutefois, la forme *mature* de communication de chacune de ces espèces a ses propres caractéristiques, e.g. le répertoire de gestes est plus varié que celui des vocalisations chez le chimpanzé alors que ce serait l'inverse chez l'humain. De plus, quelle que soit l'espèce, les gestes (*Pan trodlogytes*: Schneider et al., 2012 ; *Homo sapiens*: Beuker et al., 2013; Murillo et al., 2018) et les vocalisations (*Pan trodlogytes*: Dezecache et al., 2020; *Homo sapiens*: Laporte & Zuberbühler, 2011; Oller et al., 2021) montreraient une histoire développementale. Si chacune de ces espèces est soumise à un réagencement modal potentiellement supporté par de mêmes capacités d'ajustement et qui mène au langage chez l'humain et non chez le chimpanzé, il semble pertinent de discuter les différences de trajectoires développementales de la *communication sensible à l'attention du destinataire* entre ces deux espèces.

Ce dernier point sera discuté dans la discussion (partie 6.4) à partir des études réalisées sur chimpanzés et enfants préverbaux (article 1, 2 et 3 ; parties 3 et 4).

1.5.2. Opérationnalisation des concepts

Comme évoqué ci-dessus, bien que la *communication sensible à l'attention du destinataire* ait fait l'objet de nombreuses études chez différentes espèces, il semble y avoir une divergence méthodologique quant aux descripteurs utilisés pour l'appréhender. Pour opérationnaliser la *communication sensible à l'attention du destinataire* chez le jeune primate ici, on utilise deux descripteurs d'ajustement *unimodal* et *intermodal*, respectivement définis par la capacité de l'enfant à traiter les mères visuellement inattentives (i) en évitant de produire des signaux silencieux-visuels et/ou (ii) en favorisant la production de signaux perceptibles par l'utilisation de signaux audibles ou tactiles. Pour appréhender l'attention visuelle maternelle, l'attention maternelle a été décomposée selon trois variable distinctes : l'attention visuelle, le contact physique et le Langage Adressé à l'Enfant (LAE). La distinction

de ces trois variables nous permet ainsi d'étudier isolément l'ajustement des signaux à l'attention visuelle du destinataire, d'une part. Cela nous permet également de tester l'effet potentiel des autres variables attentionnelles (i.e. le contact physique et le LAE) sur ces capacités d'ajustement *unimodal* et *intermodal*., centrales dans notre investigation.

2.
METHODE GENERALE

Ce travail de thèse s'inscrivant dans une perspective comparative, il était question d'étudier plusieurs espèces de primates. Nous avons donc mené nos différentes études sur les primates immatures chimpanzés (*Pan troglodytes schweinfurthii*) et humains (*Homo sapiens*).

2.1. CHOIX METHODOLOGIQUES ET MACROSCOPIE DE L'INVESTIGATION

2.1.1. Choix méthodologiques généraux

2.1.1.1. *Espèces étudiées*

Comme évoquée précédemment (voir partie 1.1.3), l'approche *evo-devo* permet d'articuler les notions d'*héritage* propre à l'espèce et de *plasticité* développementale propre à l'individu en les « situant » dans un environnement écologique et socio-culturel (Bogin, 1997; Geary & Bjorklund, 2000). Sur la base de ce principe, l'étude sur les espèces humaine et chimpanzé semble pertinente du fait que (i) celles-ci possèdent des caractéristiques morphologiques et génétiques relativement proches et que (ii) leurs signaux de communications suivent des trajectoires développementales relativement longues et dépendantes des influences environnementales.

Chimpanzés et humains ont une importante proximité phylogénétique en comparaison à celle que peuvent avoir d'autres espèces de primates non-humains avec l'espèce humaine (voir Figure 1). Elles font toutes deux partie de la famille des hominidés qui forme avec la famille des hylobatidés (i.e. les gibbons) la super-famille des hominoïdes (communément appelés *Grands Singes*). En effet, au sein de cette même famille, par rapport à l'espèce humaine, le dernier ancêtre commun au chimpanzé ne date que de 6,4 millions d'années (MA) contre 8,6 MA pour le gorille (*Gorilla*), 15,2 MA pour l'orang-outan (*Pongo*) voire 20,4 MA pour les gibbons (*Hylobatidae*) (Hedges et al., 2015; Kumar et al., 2017, 2022). Il est important de souligner que les hominoïdes se distinguent des autres espèces de primates de par

leur taille plus importante, l'absence de queue (qui aurait des conséquences morphologiques au niveau des membres et du torse), la vision binoculaire (et non sur les côtés), entre autres.

Figure 1. Arbre phylogénétique des singes hominidés. Seules les dates de divergence de la famille et sous-famille communes aux humains et chimpanzés sont renseignées en millions d'années, avec l'intervalle de confiance à 95% entre crochets (source : TimeTree, Kumar et al., 2017, 2022)

D'un point de vue développemental, les traits émergeants tardivement (e.g. motricité fine dont la communication gestuelle dépend) seraient plus sujets que d'autres traits (e.g. motricité globale) à des influences environnementales chez les deux espèces (Bründl et al., 2021). Ces dernières vivent dans un environnement physique et social complexe dans lequel la survie dépendrait de leur capacité *motrice* et *sociale* (e.g. manipulation d'objet, nécessité de coopération avec des membres qui ne sont pas nécessairement familiers ; Boesch et al., 2019; Wittig & Boesch, 2003). Ces capacités présentent une variation interindividuelle croissante quelle que soit l'espèce (Bründl et al., 2021) et l'investissement maternel influencerait le développement de celles-ci.

2.1.1.2. *Périodes développementales investiguées*

Pour chaque espèce, trois périodes développementales ont été investiguées avec une période centrale marquée par des étapes importantes du développement de la communication du jeune primate. Que ce soit pour l'enfant préverbal ou les jeunes chimpanzés, les périodes intermédiaires étudiées ([11-14] mois chez l'enfant et [5-9] ans chez le chimpanzé) sont marquées par une croissance notable du répertoire de gestes manuels et une utilisation croissante de signaux silencieux-visuels (*Pan troglodytes*: Fröhlich et al., 2016; Hobaiter & Byrne, 2011; *Homo sapiens* :Beuker et al., 2013; Choi et al., 2021; Liszkowski & Rüther, 2021; voir partie 1.4.2.1, Figure 3) . Chacune d'elle est associée à des évènements développementaux déterminants.

Chez le jeune chimpanzé, la tranche d'âge [5-9] ans (i.e. le stade *juvénile*) est marquée par le sevrage avec la mère à partir duquel le chimpanzé immature n'est plus porté par elle durant les déplacements et donc par plus d'indépendance (Clark, 1977; Nishida, 2012). Partant de là, pour investiguer l'effet développemental, il s'agissait de délimiter deux tranches d'âges la précédant puis la succédant à savoir celle de [1-4] ans (i.e. le stade *enfance*) et [10-15] ans (i.e. le stade *adolescent*). Le passage du stade juvénile à l'adolescence est marqué par la puberté à savoir le développement des caractères sexuels secondaires mais surtout une indépendance totale vis-à-vis de la mère (entre l'âge de 8 et 9 ans chez les chimpanzés sauvages; Pusey, 1990).

Chez l'enfant préverbal, la tranche d'âge [11-14] mois marque l'entrée dans le stade *illocutoire* (à partir de 10 mois ; Bates et al., 1975) et l'utilisation croissante de gestes déictiques (e.g. Murillo et al., 2018). De plus, c'est durant cette période qu'émerge l'utilisation intentionnelle du geste de pointage (Liszkowski et al., 2004, 2012) et que l'enfant fait par ailleurs ses premiers pas seul (De Onis, 2006) à 12 mois. Tout comme pour les chimpanzés, pour investiguer l'effet développemental, il s'agissait de délimiter deux tranches d'âges la précédant puis la succédant à savoir celle de [7-10] mois et [15-20] mois. L'entrée dans chacune d'elles est marquée par un évènement « phare » de la signalisation de l'enfant préverbal à savoir

l'émergence du babillage à 7 mois (Esteve-Gibert & Guellaï, 2018a; Oller et al., 1999; Stoel-Gammon, 2011) et l'utilisation préférentielle du pointage avec des vocalisations à 15 mois (« pic de multimodalité » ; voir partie 1.4.2.1.1, Figure 3 ; Aureli et al., 2017; Murillo et al., 2018). L'âge de 20 mois qui est l'âge limite de l'échantillon se fait juste après l'étape où l'enfant commence à combiner des propositions verbales (de plusieurs mots) avec ses gestes (Fasolo & D'Odorico, 2012; Murillo et al., 2018).

2.1.1.3. *Observations en contexte écologique*

Les études réalisées dans ce travail de thèse ont toutes été réalisées en condition écologique. Ce choix émane du constat que, au regard des études réalisées sur la *communication sensible à l'attention du destinataire* (voir partie 1.3), la plupart des études sur le chimpanzé ont été réalisées sur des populations captives (12 sur 14 études réalisées, soit 85%) et que celles menées sur les enfants préverbaux ne sont pas comparables entre elles (comme souligné partie 1.3.2.2).

Par ailleurs, d'un point de vue théorique, comme il a été discuté précédemment (voir parties 1.4.3 et 2.1.1), que ce soit pour les chimpanzés ou les humains, le développement communicationnel est dépendant de l'expérience puisque qu'il serait plus sujet aux influences environnementales physiques et sociales (Bründl et al., 2021). Il semble donc pertinent, à travers l'approche *evo-devo*, d'observer ce développement en condition spontanée et non contrôlée. D'autant plus que les caractéristiques environnementales physiques de chacune des populations observées ici, i.e. chimpanzés de la forêt de Budongo (Reynolds, 2006) et des humains urbains français, sont de fait très différentes et par conséquent influentes quant au développement de ces capacités (e.g. manque de visibilité sur les grandes distances chez les chimpanzés de par la végétation abondante, espaces restreints de déplacement à l'intérieur des maisons chez les enfants humains). De plus, de par l'approche comparative défendue ici, il semblait nécessaire d'observer des chimpanzés non-captifs afin que l'espèce non-humaine étudiée n'ait relativement pas été influencée par l'espèce (humaine) à laquelle on la compare. Comme

souligné par Leavens (Leavens, 2004), le fait que les primates non-humains en captivité soient maintenus en cage influencerait leur communication. Ils sont davantage sujets à des objets inaccessibles dont l'accessibilité implique l'aide du soigneur (humain). Leur signalisation se verrait, de fait, influencée par la responsivité et donc la signalisation humaine.

2.1.2. **Macroscopie de l'investigation générale**

Sur la Figure 2 ci-dessous, est représentée la trame méthodologique appliquée à chacune des espèces étudiées qui sera développée dans les parties ci-dessous (voir parties 2.2, 2.3 et 2.4).

Comme on peut le voir, contrairement aux études réalisées sur les sujets humains, la collecte de données sur les chimpanzés a été réalisée avant de composer l'échantillon sur lequel mener l'étude. Le nombre d'observations par individu chimpanzé de l'échantillon était variable alors que pour l'espèce humaine, les individus étaient observés deux fois (même si trois d'entre eux n'ont pas pu être observés une deuxième fois). Toutefois, qu'elle que soit l'espèce, toutes les observations d'un même individu étaient réalisées sur une seule et même tranche d'âge à savoir [0-4], [5-9] ou [10-15] ans pour les chimpanzés et [7-10], [11-14] ou [15-20] mois pour les enfants préverbaux (comme justifié précédemment, partie 2.1.1.2).

Suite aux observations réalisées sur les enfants préverbaux, un questionnaire de développement langagier (Développement du Langage de Production en Français, DLPF ; Bassano et al., 2005, 2020) évaluant les compétences *lexicales*, *syntaxiques* et *pragmatiques* avait été diffusé aux parents en moyenne 15,5 mois après la dernière observation. Excepté pour le DLPF, les analyses ont été réalisées sur la totalité de l'échantillon sélectionné pour les chimpanzés (N=12) et les enfants préverbaux (N=30).

Figure 2. Trame méthodologique appliquée à chacune des espèces étudiées.

Note :

(1) *La collecte de données avait été réalisée avant que la sélection des dyades soit faite pour réaliser l'étude sur les jeunes chimpanzés (voir Article 1, partie 3).*

(2) *Le questionnaire de Développement du Langage de Production en Français (DLPF) est un questionnaire d'évaluation de performance lexicale, syntaxique et pragmatique (Bassano et al., 2005,2020).*

(3) *Excepté pour le DLPF, les analyses ont été réalisées sur la totalité de l'échantillon initial.*

2.2. ETUDE OBSERVATIONNELLE SUR LES JEUNES CHIMPANZES (*PAN TROGLODYTES*)

2.2.1. Sujets-chimpanzés

Les données vidéos pour l'étude observationnelle sur chimpanzés (article 1, partie 3) sont extraites d'une base de données de vidéos de grands singes *Great Ape Dictionary video database* (www.greatapedictionary.com) gérée par le *Centre for Social Learning & Cognitive Evolution* du département de Psychologie et de Neuroscience de l'Université de Saint-Andrews (Royaume-Uni). Les données codées pour cette étude ont été collectées dans la réserve forestière de Budongo (Eggeling, 1947) sur la communauté de chimpanzés *Sonso* d'Afrique de l'est (*P. troglodytes schweinfurthii* ; Hobaiter & Byrne, 2011). La collecte originale a commencé en 2007 sur une communauté composée de 81 individus dont 18 enfants (3 mâles, 15 femelles) âgés de 0 à 4 ans, 15 juvéniles (6 mâles et 9 femelles) âgés de 5 à 9 ans et 16 adolescents (10 mâles et 6 femelles) âgés de 10 à 15 ans (pour plus de détails, voir Hobaiter & Byrne, 2011; Reynolds, 2006).

Dans cette étude transversale (article 1, voir partie 3), on s'intéresse à trois tranches d'âges d'intérêt : *enfants* (0 à 4 ans), *juvéniles* (5 à 9 ans) et *adolescents* (10-15 ans) comme mentionné et justifié partie 2.1.1.2. Comme des données longitudinales détaillées sur les changements comportementaux ou physiologiques individuels n'étaient pas disponibles, nous avons utilisé des catégories d'âge fixes entre les individus. Nous avons ainsi sélectionné les jeunes chimpanzés, pour chacune des trois catégories d'âge, ayant produit au moins 10 gestes envers leur mère dans le jeu de données total. Le jeu de données final comportait 264 gestes produits par 11 individus (voir Table 1).

Table 1. Tranche d'âge d'appartenance, nombre de gestes produits, sexe et âges en mois à la première et dernière observation par individu et moyenne d'âge pondérée (+ écart-**type) en mois par tranche d'âge**

Tranche d'âge	Sujet	Sexe	Âge à l'observation(1)	Âge à l'observation(x)	Moyenne d'âge pondérée (+ET)
Enfants 0-54 (N = 3)	Karibu	F	18	49	*Tous* : 23,25 (± 5.41)
	Klauce	M	15	53	
	Kox	F	15	25	
Juvéniles 55-102 (N = 4)	Karo	F	83	102	*Tous* : 74.15 (± 14.47)
	Kasigwa	M	60	64	
	Night	F	57	77	
	Zak	M	68	80	
Adolescents 103-180 (N = 4)	Kumi	F	107	127	*Tous* : 139,86 (± 21,79)
	Kwezi	M	129	132	
	Nora	F	141	152	
	Rose	F	154	175	

2.2.2. Collecte de données

Figure 3. Enregistrement vidéo d'un contexte d'interaction sociale entre chimpanzés en condition écologique réalisé par C. Hobaiter (CH) (réserve forestière de Budongo, Ouganda).

Les données vidéo originales avaient été recueillies par CH (voir Figure 3) pendant trois périodes de terrain entre octobre 2007 et août 2009 (pour plus de

détails sur la méthode de collecte, voir Hobaiter & Byrne, 2011a) selon une méthode d'échantillonnage du comportement (Altmann, 1974). Toute interaction sociale relevant d'une communication potentielle avait été filmée. En pratique, toutes les occasions où deux individus ou plus étaient présents et non occupés par des activités solitaires telles que se reposer ou se toiletter étaient jugées comme relevant d'une communication potentielle. Cependant, comme les chimpanzés d'Afrique de l'Est sont de type *fission–fusion*, certains individus observés avaient été présents presque quotidiennement, et d'autres absents pendant plusieurs semaines ou mois (Aureli et al., 2008). Le système social de type fission-fusion implique qu'un groupe d'individus se subdivise en sous-groupes de taille variable qui se dissocient et se recomposent en de nouveaux sous-groupes en fonction de leurs activités ainsi que de la disponibilité des ressources (Aureli et al., 2008). Il restait donc des différences dans la fréquence avec laquelle il avait été possible d'observer des individus particuliers. Par conséquent, certains individus étaient mieux représentés ou représentés sur une période plus longue dans l'ensemble de données finales. Les vidéos étaient relativement courtes, duraient généralement entre 2 et 3 minutes, mais un même événement communicatif comprenait souvent plusieurs gestes envers la mère. Après sélection de tous les individus immatures ayant produit plus de 10 gestes dirigés vers leur mère sur une même tranche d'âge d'intérêt, 141 clips vidéo ont été sélectionnés (12 enregistrements par individu en moyenne) pour l'étude de l'article 1, partie 3.

La collecte de données originale avait été approuvée par le comité d'éthique et de bien-être animal de l'Université de St Andrews (RU). Le travail de terrain de CH avait été soutenu par des subventions de la Fondation Wenner-Gren (http://wennergren.org) et le Russell Trust.

2.2.3. Echantillonnage des comportements

Comme présenté précédemment (partie 1.4.1), l'échantillonnage des comportements avait été réalisé selon le codage éthologique.

2.2.3.1. Codages des signaux produits par le jeune chimpanzé

2.2.3.1.1. Distinction entre gestes simples et produits en séquence

Pour les études réalisées sur chimpanzés (article 1, partie 3), le répertoire considéré comprenait plus de 60 types de gestes (Byrne et al., 2017; Hobaiter & Byrne, 2011a). Par ailleurs, notons que la quasi-intégralité des gestes de ce répertoire sont illustrés par des vidéos sur Dictionnaire du Grand Singe. Comme dans d'autres études (e.g., Hobaiter & Byrne, 2011b), on distingue les *gestes simples*, définis par le fait qu'un geste est suivi d'une pause d'au moins 1 seconde, des *gestes produits en séquence* produits et séparés par moins de 1 seconde de pause du geste le précédant et/ou le succédant. L'intervalle d'une seconde était considéré comme le temps minimum requis pour être interprété comme une attente de réponse (Genty & Byrne, 2010; Hobaiter & Byrne, 2011b).

2.2.3.1.2. Association multimodale de signaux en séquence

Par ailleurs, pour chaque *séquence* de gestes produite, les informations suivantes étaient relevées : la *modalité intra-séquence* et la *dimension modale* (unimodale, bimodale ou trimodale) de la séquence. La *modalité intra-séquence* indiquait la succession de modalités de gestes composant la séquence [par exemple, une séquence de trois gestes composée de *hit other - object move - stomp object* était codée comme Tactile-Audible-Audible (*contact-audible-audible ;* CAA) ; une séquence à deux gestes composée de *dangle* (i.e. suspension à une branche en pivotant sur soi-même, ce qui produit un son) – *swing* (i.e. balancement d'un des membres) a été codée

Audible-Silencieux-visuel (AS)]. La *dimension modale*, indiquait combien des trois modalités de signaux (audible, silencieuse-visuelle ou tactile) étaient incorporées dans la séquence (i.e. *unimodale*, *bimodale* ou *trimodale*).

2.2.3.2. ***Codage de l'attention visuelle maternelle***

La mère était considérée comme *visuellement attentive* uniquement lorsqu'elle regardait directement son enfant (en orientant la tête ou le regard dans sa direction) juste avant que celui-ci ne produise le geste en question. Ainsi, notre variable d'*attention visuelle maternelle* avait *deux conditions* : (1) *attentive* : le visage de la mère est entièrement orienté vers l'enfant ; en considérant la ligne de vue directe du destinataire comme étant à 0°, le signaleur ne peut être situé que dans un arc de 45° de part et d'autre de celui-ci ; (2) *inattentive* : le visage de la mère n'est pas complètement orienté vers l'enfant ; le signaleur est situé à un angle de 45° ou plus par rapport à la ligne de vue directe du destinataire. Des exemples vidéos mis en ligne pour illustrer chaque condition à chaque âge sont disponibles sur Supports vidéos Dafreville et al. 2021.

2.2.3.3. Concordance inter-observateur

Une étude de concordance inter-observateurs avait été réalisée sur 11 % du jeu de données initial par l'équipe du *Wild Minds Lab* de l'université de Saint-Andrews (pour le détail, voir Hobaiter & Byrne, 2011a). La concordance entre observateurs a été calculée selon le Kappa de Cohen. Le pourcentage de concordance des différentes variables manipulées dans l'étude réalisée sur les chimpanzés (voir partie 3).

2.3. ETUDES OBSERVATIONNELLES SUR LES ENFANTS PREVERBAUX (*HOMO SAPIENS*)

2.3.1. Sujets-humains

Trente dyades mère-enfants ont été recrutées par démarchage auprès des professionnels de crèches et des parents au nom du *BabyLab* de la plateforme CCU (UT2J, Toulouse, France). La campagne de recrutement était faite par voie d'affichage, de distribution de flyers et en allant directement à la rencontre des parents en fin de journée dans une vingtaine de crèches gérées par le Centre Communal d'Action Sociale (CCAS) de Toulouse via lequel l'annonce avait pu être diffusée par e-mail. Après recrutement, une pré-visite était programmée avec les parents (déroulement de la pré-visite détaillé partie 2.3.2.1). Il s'agit majoritairement de mères primipares (N = 18 sur 30 dyades mère-enfants, 60 %) et le nombre moyen de frères et sœurs par enfant observé était de 1,87. L'âge de la mère variait entre 27 et 40 ans. La mère avait un haut niveau d'éducation formelle, toutes étaient diplômées d'études supérieures (i.e. au moins au niveau du baccalauréat). Les enfants ont été allaités pour la plupart (N = 24 sur 30 dyades mère-nourrisson, 80%). Neuf enfants avaient rencontré des difficultés de santé à leur naissance (i.e. prématurité, accouchement difficile, asthme).

Ces études observationnelles (articles 2,3 et 4, parties 4 et 5) étant transversales, les dyades mère-enfant étaient recrutées selon l'âge de l'enfant dont les âges aux observations devaient se situer entre 7 et 20 mois. Les dyades ont été regroupées selon trois tranches d'âges d'intérêt qui se situaient *avant, pendant* et *après* émergence d'une forme visible de communication intentionnelle du geste de pointage à l'âge de 12 mois (Camaioni et al., 2004; Carpenter et al., 1998; Colonnesi et al., 2010). Les trois tranches d'âges d'intérêt de l'enfant étaient donc : [7-10], [11-14] et [15-20] mois (dont le choix est discuté partie 2.1.1.2).

Table 2. Tranche d'âge d'appartenance, sexe et âges en mois à la première et **seconde visite par individu et moyenne d'âges (+ écart**-type) en mois par sexe **et par tranche d'âge**

Tranche d'âge	Sujet	Sexe	Age à la visite(1)	Age à la visite(2)	Moyenne d'âge (+ET)
7-10 (N = 11)	TH23	M	7	8	*Filles* : 8,83 (± 1,03) *Garçons* : 8,80 (± 1,09) *Tous*: 8,81 (± 1,00)
	SA8	M	7,3	8,2	
	CH19	F	7,5	8,3	
	AD3	F	7,7	9,1	
	ZO13	F	8,1		
	AP26	F	8,5	9,5	
	CL7	M	8,8	9,8	
	EL15	F	9,1		
	AL30	M	9,2	10,1	
	PA4	F	9,2	10,2	
	TH16	M	9,8		
11-14 (N = 12)	CL17	M	11,3	12,5	*Filles* : 12,78 (± 0,99) *Garçons* : 13,11 (± 0,91) *Tous* : 13,03 (± 0,90)
	AR11	M	11,4	12,4	
	LO6	F	11,4	12,6	
	AL24	M	11,8	13,6	
	CO15	M	11,8	12,9	
	MA26	F	12	12,9	
	AL20	M	12,9	14,2	
	MA12	M	13	13,9	
	LA16	F	13,4	14,4	
	AI23	M	13,5	14,5	
	GA3	M	13,5	14,4	
	AX15	M	13,6	14,8	
15-20 (N = 7)	AU30	M	14,9	16,1	*Filles* :18,53(± 1,59) *Garçons* : 16,41 (± 0,93) *Tous*: 17,01 (± 1,44)
	AR13	M	15	16,1	
	LO3	M	15,9	17,4	
	AN13	M	16	17,2	
	VI13	F	16,9	17,9	
	NI7	M	17	18,5	
	MA25	F	19	20,3	

2.3.2. Procédure et Design

2.3.2.1. *Déroulement de la pré-visite*

Cette pré-visite à domicile, avait un but principal de familiarisation mutuelle entre l'observateur et la maman et un repérage les lieux en vue du positionnement

d'une des caméras lors de la phase concrète d'observation à venir. Cette pré-visite dure environ 30 minutes durant laquelle après avoir présenté son projet, l'observateur présentait succinctement le contenu des deux documents qu'il lui fallait faire signer aux parents. Ces deux documents étaient un formulaire de *consentement* et d'*autorisation d'utilisation du droit à l'image* (voir Annexes 1 et 2). Aussi, l'observateur précisait qu'à la fin des observations, il leur serait demandé de signer un *amendement au consentement* (Annexe 3) qu'ils seront libres de signer ou non. Enfin, si la mère était d'accord, après avoir signé les documents nécessaires, l'observateur lui posait quelques questions sur son enfant (configuration familiale, mode de garde, …) au regard d'un *questionnaire de pré-visite* (Annexe 4).

2.3.2.2. *Déroulement des visites 1 et 2 à domicile*

Suite à la pré-visite, une première observation est programmée. Deux visites durant entre 1h30 et 2h ont été réalisées au domicile de chaque dyade mère-enfant. Chacune d'entre elles incluait 60 minutes (± 15 minutes) d'enregistrement vidéo (exploitable pour analyse). Les visites 1 et 2 étaient espacées de 4 semaines (± 1 semaine). Chaque visite se faisait pendant le moment du repas. L'enregistrement devait (idéalement) se décomposer en trois parties : *avant*, *pendant* et *après* le repas dont les cinq dernières minutes étaient dites « expérimentales » puisque l'expérimentateur cherchait à détourner l'attention de la mère en lui demandant de remplir un questionnaire tout en interagissant oralement avec elle. Cette phase « semi-expérimentale » n'a pas été traité dans nos analyse mais les réponses aux questionnaires l'ont, elles, étaient. Le but ayant été d'avoir accès à un repas des plus habituels possibles, il pouvait y avoir d'autres personnes que la dyade mère-enfant (i.e. père, frères et sœurs…) durant l'observation et aucune consigne n'était donnée quant à leurs interactions avec l'enfant. Les plages horaires des visites étaient : entre 11h et 13h (pour le déjeuner), entre 15h et 17h (pour le goûter) ou entre 18h et 20h30 (pour le dîner). L'expérimentateur ne devait s'engager dans aucune interaction verbale ou visuelle avec les protagonistes jusqu'à 5 minutes avant la fin de l'enregistrement. Les visites se déroulaient ainsi :

1) L'observateur installait une première caméra dite *fixe* (i.e. installée sur trépied) dans un coin de la pièce en s'assurant qu'au moment où la mère donnerait à manger à son enfant, la caméra puisse filmer le visage de la mère.
2) L'observateur prévenait la mère que juste avant la fin de l'enregistrement, il lui ferait remplir un questionnaire.
3) Au moment où la mère se disait prête, l'observateur activait la caméra *fixe*. Il rappelait qu'à partir du moment où il reprendrait une autre caméra (*mobile*) entre les mains, il n'interviendrait à aucun moment et qu'il était important que la mère n'interagisse pas avec lui.
4) Puis, l'observateur activait cette seconde caméra dite *mobile* (i.e. qu'il tenait entre ses mains) et se positionnait de telle façon à pouvoir suivre l'enfant et garder son visage dans son champ de vision.
5) Cinq minutes avant la fin de l'enregistrement, l'observateur, tout en continuant à tenir la caméra, tendait à la mère le *questionnaire visite 1* de distraction (voir Annexe 5) ayant eu pour but de détourner son attention visuelle de l'enfant. Pendant que la mère remplissait le questionnaire, l'observateur s'adressait volontairement à elle pour renforcer ce phénomène de diversion.
6) L'observateur stoppait l'enregistrement au bout de 60 minutes.

Lors de la deuxième visite, l'expérimentateur faisait signer un amendement au consentement (voir Annexe 3).

2.3.2.3. *Questionnaire du « Développement du Langage de Production en Français » (DLPF ; Bassano et al., 2005, 2020)*

Suite aux visites, un questionnaire sur le développement langagier (*DLPF* ; Bassano et al., 2005, 2020) à remplir en ligne avait été diffusé par e-mail aux parents de novembre à décembre 2020. Le DLPF est un outil d'évaluation et d'analyse de production langagière pour les enfants âgés de 18 à 42 mois (Bassano et al., 2020, 2005). Etant tridimensionnel, il évalue à la fois les compétences *lexicales*, *syntaxiques*

et *pragmatiques* (Bassano et al., 2005, 2020). Les parents pouvaient le remplir en plusieurs fois sur une période de 4 semaines (voir détails dans l'article 4, partie 5).

2.3.3. Collecte de données

Figure 4. Enregistrement vidéo d'une observation en condition écologique de dyade mère-enfant réalisé par Mawa Dafreville (MD) (Toulouse, France).

Les données vidéo étaient recueillies par MD (voir Figure 4) entre mars 2019 et mars 2020 en utilisant une approche d'échantillonnage focal sur l'enfant (Altmann, 1974). Les enfants étaient observés un par un et les signaux de communication étaient enregistrés en continu sur 3 focus de 5 minutes où chaque focus était aléatoirement sélectionné *avant*, *pendant* et *après* le repas pour chacune des deux visites. Lorsqu'il était impossible d'avoir un focus avant ou après le repas parce que l'observation ne couvrait pas ces trois temps, on remplaçait le moment de repas manquant (e.g. un focus *après-repas*) par un deuxième moment de repas non-manquant (e.g. un second focus *avant-repas*). Les visites étaient filmées à l'aide de deux caméras au total : une caméra *fixe* et une caméra *mobile*. La caméra *fixe* positionnée sur trépied permettait d'avoir une vision d'ensemble de la pièce et donc des déplacements de chacun notamment de la mère. La caméra *mobile* était tenue par l'observateur qui se focalisait uniquement sur l'enfant et le suivait pendant tous ses déplacements à une distance de plus d'un mètre (voir Figure 4). Les vidéos

étaient relativement longues, elles duraient généralement entre 45 et 60 minutes. Approximativement, un enfant de [7-10], [11-14] et [15-20] mois pouvait respectivement produire en moyenne 125, 146 et 166 signaux envers la mère sur 15 minutes d'un enregistrement. Après la collecte de données nous disposions de 57 enregistrements au total pour analyse (deux vidéos par enfant pour 27 dyades et qu'une seule pour les 3 restantes pour cause d'abandon). Le questionnaire ayant été diffusé aux parents de 25 enfants (l'adresse e-mail de ceux des 5 enfants restant n'étaient plus valides), fut complété par 21 des parents contactés. La période séparant la dernière visite au remplissage du questionnaire était en moyenne de 15,56 mois (± 2,97 mois).

Les données vidéo et de codage étaient anonymisées et stockées sur des disques externes stockés dans des endroits sûrs verrouillés jusqu'à 15 ans après la fin de l'étude, date à laquelle elles seront détruites. Les données vidéo seront exclusivement utilisées à des fins d'analyse. Les parents pouvaient également signer un formulaire de droit à l'image nous autorisant à utiliser les données vidéo à des fins pédagogiques ou de communication scientifique (voir Annexe 2). La collecte de données avait été approuvée par le comité d'éthique de l'Université de Toulouse (i.e., CERNI ; n° IRB00011835-2019-03-19-141).

2.3.4. Echantillonnage des comportements et Procédure de codage

Comme développé et argumenté précédemment, l'analyse comportementale a été réalisée par le codage éthologique, dans une approche multimodale (voir partie 1.4). Ainsi, tous les signaux produits par l'enfant étaient relevés et catégorisés selon leur modalité sensorielle perceptive dominante (Byrne et al., 2017; Genty et al., 2009; Hobaiter & Byrne, 2011a). On y distinguait donc les signaux *audibles*, *tactiles* et *silencieux-visuels* (voir partie 1.4) à l'intérieur desquels certains critères de distinctions entre signaux seront présentés ci-dessous (partie 2.3.4.1). Chaque signal produit était assigné à un état d'attention maternelle défini

selon plusieurs variables attentionnelles qui seront présentées ci-dessous (partie 2.3.4.2).

2.3.4.1. ***Codage des signaux produits par l'enfant***

Les signaux produits par l'enfant préverbal étaient codés avec le logiciel ELAN (version 5.9; Lausberg & Sloetjes, 2009) selon un répertoire de 23 signaux différents (voir répertoire dans le *matériel additionnel* de l'article 2, partie 4.1). Pour chaque signal de l'enfant relevé, les informations suivantes étaient enregistrées : (1) la *modalité sensorielle* (audible, tactile ou muette-visuelle), (2) si le signal a été *associé au moins une fois à un regard* (vers la mère et/ou vers un même objet d'attention) ou non, (3) le *canal de production* (gestuel ou oral), (4) si le signal était *combiné* à un autre signal d'une autre modalité sensorielle, (5) la présence d'un autre receveur plus proche que la mère (i.e. *receveur non maternel*), (6) l'état *d'attention de la mère* et du receveur non maternel le cas échéant.

2.3.4.1.1. Distinction de signaux : cas particuliers

A l'intérieur de chaque catégorie de signaux, la différenciation de certains signaux n'était pas évidente et cela a suscité de nombreux échanges, notamment avec les stagiaires (i.e. Claire Lerbet et Noémie Pantel) avec qui le double codage a été réalisé. En voici quelques exemples.

Les signaux audibles

Parmi les signaux audibles il nous aura fallu distinguer les signaux oraux entre eux ainsi que les signaux gestuels.

Concernant la distinction des signaux oraux, l'identification des différents sons oraux avec et sans cordes vocales (i.e. respectivement les protophones et les bruits de bouche) se faisait à l'oreille. Pour s'entraîner à les reconnaître, il aura fallu dans un premier temps s'entraîner à partir des vidéos du *matériel additionnel* librement disponible de Oller et al. (2013). Cet entraînement permettait de distinguer les

différents protophones à savoir les *raspberries, growls*, *squeals, grunts* et babillage (Oller et al., 2013). A partir de là, hors signaux oraux végétatifs (e.g. rots, éternuements, toussements), tous les sons oraux ont été relevés selon les catégories détaillées dans le *matériel additionnel* de l'article 2, partie 4.1.

Concernant la distinction des signaux gestuels audibles, il s'agissait de créer deux catégories : les gestes audibles *avec* objet et *sans* objet. Le critère de distinction de ces deux catégories était la *préhension* manuelle de l'objet par lequel le son était produit. Figure 5, est qualifié de geste audible *sans* objet, le fait que l'enfant tape sur la table avec sa main. Bien que la table soit un objet, la production de geste n'implique pas que l'enfant tienne un objet entre ses mains. Par ailleurs, comme illustré Figure 6, le geste audible *avec* objet peut prendre plusieurs formes. L'enfant peut par exemple percuter l'objet tenu entre les mains contre un autre objet mais peut aussi le secouer (voir Figure 6). Notons que lorsque l'enfant appuie sur le bouton d'un jouet induisant un son non-mécanique (e.g. une mélodie), le comportement n'était pas qualifié de geste audible.

Figure 5. Exemple de gestes audibles produits sans objet.

Figure 6. Exemples de gestes audibles produits avec objet

Les signaux silencieux-visuels

Concernant les signaux silencieux-visuels, les définitions qu'il nous aura fallu affiner pour faciliter la distinction de certains gestes dont l'interprétation visuelle n'était pas évidente, concernaient essentiellement les gestes déictiques. On peut évoquer ici deux cas particuliers de distinction : (i) entre les gestes de *showing* et de *giving* et (ii) entre les gestes de *reaching* et pointage avec la main (i.e. *whole-hand pointing gestures*).

Concernant la différence entre *showing* et *giving* (donc les définitions sont précisées dans le *matériel additionnel* de l'article 2, partie 4.1). Ces deux gestes font partie des gestes déictiques *proximaux* puisqu'ils impliquent la préhension manuelle de l'objet de référence. Le critère de distinction le plus déterminant pour ces gestes est l'extension du bras qu'il y aurait pour le *giving* et non le *showing* (comme illustré Figure 7).

Figure 7. Exemples de gestes de *showing* et *giving*

Concernant la différence entre *reaching* et pointage avec la main (i.e. *whole-hand pointing gestures* ; dont les définitions sont précisées dans le *matériel additionnel* de l'article 2, partie 4.1). Ces deux gestes font partie des gestes déictiques *distaux* puisque n'impliquent pas la préhension manuelle de l'objet de référence. Le critère de distinction le plus déterminant pour ces gestes est l'orientation du tronc de l'enfant puisque pour le geste de *reaching* (i.e. relevant d'une tentative de préhension), le tronc doit être orienté en direction de l'objet alors que pour le pointage main ce n'est pas le cas (comme illustré Figure 8). Notons toutefois que si le tronc est avancé en direction de l'objet et que la main pointe l'objet avec le doigt, on considèrera bien ce geste comme étant un geste de pointage avec index (comme illustré Figure 9).

Figure 8. Exemples de gestes de *reaching* et pointage avec la main

Figure 9. Exemples de gestes de pointage avec index

Les signaux tactiles

Une des principales difficultés de codages des gestes tactiles relevait de la temporalité du geste. La question était de savoir si un contact initié par l'enfant qui perdure dans le temps relevait ou pas d'un signal communicatif et si oui, quel était le critère de fin de ce geste. Il a donc été question de ne considérer que les gestes tactiles produit avec les mains ou la tête dont le contact avait été initié par l'enfant

et durait moins de 5 secondes. Si le contact durait moins de 1 seconde mais qu'il était répété plusieurs fois à intervalle de moins de 1 seconde alors on le considérait comme un seul geste tactile. La définition est détaillée dans le *matériel additionnel* de l'article 2, partie 4.1.

2.3.4.1.2. *Combinaison de signaux*

Les *combinaisons multimodales* (i.e. combinaison de signaux de différentes modalités sensorielles) étaient également relevées lors du codage. Elles étaient catégorisées selon les modalités sensorielles qui les composaient. On y distinguait donc les combinaisons *distales* (combinaison de signaux audibles et silencieux-visuels *sans* signaux tactiles) des combinaisons *proximales* (combinaison d'au moins un signal tactile et d'un signal d'une autre modalité). Une *combinaison* était enregistrée à chaque chevauchement entre un signal et un autre dont la durée entre le moment où se termine le premier signal et celui où débute le second était plus de 200 millisecondes.

2.3.4.1.3. *Coordination des signaux de communication avec le regard*

Comme mentionné plus haut, il était renseigné si l'enfant regardait au moins une fois son destinataire et/ou l'objet observé par son destinataire (i.e. un objet d'attention conjointe) sur une fenêtre allant de moins de 1 seconde avant le début du signal à 1 seconde après la fin du signal. Ainsi nous avions créé une variable à quatre modalités pour renseigner *la coordination du regard* à un signal :

a) L'enfant regardait au moins une fois sa mère (GTM)

b) L'enfant regardait au moins une fois le même objet que sa mère (JA)

c) L'enfant regardait au moins une fois sa mère et le même objet que sa mère (GTM+JA)

d) L'enfant ne montrait aucun des patterns ci-dessus (0)

Pour l'article 3 (partie 4.2), cette variable a été transformée en variable binaire (présence/absence) dans laquelle on considérait qu'il y avait *présence* de

coordination du regard seulement si l'enfant avait regardé au moins une fois sa mère, i.e. cas a) et c).

2.3.4.2. ***Codage de l'attention maternelle***

Pour définir l'état d'attention du destinataire, les variables attentionnelles ci-dessous étaient (pour plus de détails, voir dans le *matériel additionnel* de l'article 2, partie 4.1) : 1) L'*attention visuelle* (attention/inattention) définie selon les mêmes règles de codage que pour l'étude réalisée sur chimpanzés (voir Figures 10 et 11) ; 2) Le *Langage Adressé à l'Enfant* (LAE/silence) défini selon que la mère s'adressait verbalement à son enfant ou non ; 3) Le *contact physique* (contact/non-contact) défini selon si la mère avait au moins une partie de son corps en contact physique avec son enfant ou non. L'*attention auditive*, c'est-à-dire la possibilité que la mère puisse entendre son enfant, était considérée comme toujours présente car il n'était pas possible de déterminer son absence.

Figure 10. Exemple de cas d'attention visuelle maternelle (qui en plus était codé comme du Face-to-Face)

Figure 11. Exemples de cas d'inattention visuelle maternelle

De plus, dans le codage initial, d'autres formes d'attention spécifiques telles que le *face-to-face* (i.e. un regard mutuel maintenu pendant plus de 2 secondes ou le parent et l'enfant sont physiquement proches, c'est-à-dire à moins de 2 mètres l'un de l'autre, voir Figure 10), le *portage* (i.e. le parent porte son enfant dans ses bras) ou l'*étreinte* (i.e. le parent a son enfant contre elle ou l'enlace avec ses bras sans objet solide entre eux et leurs deux têtes sont physiquement très proches l'une de l'autre) avaient également été relevées. Cela avait aussi été le cas pour des formes spécifiques d'inattention telle que *l'attention distraite* (i.e. le parent s'adresse à quelqu'un d'autre que l'enfant *ou* quelqu'un d'autre que l'enfant s'adresse à lui en

présentiel ou en distanciel, e.g. par téléphone). Toutefois, ces « autres » variables d'attention et d'inattention spécifiques ne furent pas exploitées dans ce travail de thèse se concentrant sur l'étude de la *communication sensible à l'attention du destinataire*.

2.3.4.3. Concordance inter-observateur

Les enregistrements des enfants préverbaux ont été entièrement codés par nous-même, ce qui avait donc servi de référence pour former un deuxième codeur aveugle. Le deuxième codeur (Noémie Pantel) s'étant entraîné pendant un mois, a ensuite codé à l'aveugle 20 % de l'ensemble des données. La concordance entre observateurs était calculée selon le Kappa de Cohen. Les coefficients sont présentés ci-dessous (tableaux 4 et 5) et ceux des différentes variables manipulées dans les différentes études sont précisés dans la partie 4. On considère qu'un coefficient Kappa est bon à partir de 70%, très bon à partir de 80% et excellent à partir de 90%. Comme on peut le voir dans les tableaux ci-dessous, les coefficients les plus bas sont ceux des sons oraux et des mouvements corporels. Concernant les sons oraux, ce n'était pas surprenant du fait que la perception des sons soit soumise à une grande variabilité interindividuelle. Les discordances étaient principalement dues à la détection des sons plus qu'à l'identification de ceux-ci. Concernant les mouvements corporels, le problème était principalement dû au *découpage* de celui-ci en plusieurs mouvements corporels si l'enfant arrêtait un des mouvements pendant plus de 1 seconde. Or, pour ces comportements, notamment pur ceux impliquant des répétitions comme le balancement (*swagger*) ou les mouvements répétitifs des membres (*circular movement of the limbs*), le critère de fin de la dernière répétition pouvait être sujet à plus de variabilité entre codeurs (e.g. ralentissement du mouvement, changement vers un autre type de signal). Notons également que ces mouvements corporels répétitifs (e.g. un balancement sur une chaise) pouvaient parfois produire un son. Dans ce cas, ce signal devait être relevé comme étant un geste audible *sans* objet. Tout comme pour les vocalisations, la perceptibilité de ce son était sujette à de la variabilité interindividuelle.

Table 3. Synthèse des coefficients Kappa de Cohen calculés sur les signaux **produits par l'enfant et l'association du regard aux** signaux (sur 20 % du jeu de données total).

Types de signaux	Signaux	Kappa de Cohen		Moyenne du Kappa de Cohen
Signaux audibles	Vocalisations	0,72		0,71
	Bruits de bouche			
	Gestes audibles	0,83		
Signaux silencieux-visuels	Gestes déictiques	0,88	0,80	0,79
	Mouvements corporels	0,65		
	Gestes conventionnels/symboliques	/		
	Sourires	/		
Signaux tactiles	Gestes tactiles	/		
	Regards	*0,92*		*0,92*

Table 4. Synthèse des coefficients Kappa de Cohen calculés sur les variables attentionnelles maternelles.

Etat d'attention maternelle	Coefficient du Kappa de Cohen
Attention maternelle *visuelle* et *tactile*	0.89
IDS	0.81

2.4. ANALYSE : OPERATIONNALISATION DES AJUSTEMENTS ET STATISTIQUES INFERENTIELLES

Quel que soit le jeu de données manipulé, i.e. du jeune chimpanzé (N = 12) ou de l'enfant préverbal (N = 30), aucun des résultats ne suivait une loi normale donc toutes les analyses statistiques ont été réalisées à l'aide de tests non-paramétriques de corrélation de *Spearman*, des tests de *permutation* et des tests de *Fisher* (Siegal & Castellan, 1988) selon les hypothèses à tester. Pour certaines études (article 1 ,2 et 4 retrouvées parties 3, 4.1 et 5) ont été réalisées au préalable des sélections de modèles linéaires mixtes généralisés manipulant des occurrences et des données de proportions selon l'étude (*GLMM*; Bolker et al., 2009; Harrison,

2015). Pour chaque étude, les hypothèses testées sont renseignées dans les parties 3, 4 et 5.

Notons toutefois que « l'ossature commune » à toutes ces analyses est construite sur les principes d'ajustement *unimodal* et *intermodal* permettant l'étude de la *communication sensible à l'attention du destinataire* (parties 1.3 et 1.5.2). L'ajustement *unimodal* consiste à produire moins de signaux silencieux-visuels quand le destinataire est visuellement inattentif que quand il l'est. De ce fait, la *distribution* de signaux silencieux-visuels d'une condition visuelle à l'autre était testée à l'aide d'un test de permutation. L'ajustement *intermodal* consiste à (i) produire plus de signaux perceptibles (audibles et tactiles) que signaux non- perceptibles (silencieux-visuels) quand le destinataire est visuellement inattentif et (ii) que cette différence de *proportions* (audibles-et-tactiles versus silencieux-visuels) dans la situation où le destinataire n'est pas visuellement attentif est significativement différente (plus grande ou de signe inversé) que celle retrouvée dans la situation opposée. La comparaison de ces proportions qui relèvent à la fois de la modalité sensorielle du signal (audibles-et-tactiles *ou* silencieux-visuels) et de l'attention visuelle maternelle (attentive/inattentive) était réalisée à l'aide de tests exacts de Fisher sur échantillon apparié (Pearce, 2016). Des tests de Fisher ont également été réalisés sur des échantillons non-appariés pour tester l'effet de l'âge par exemple sur l'ajustement *unimodal* (e.g. l'effet des variables âge et attention maternelle visuelle sur l'utilisation de signaux silencieux-visuels).

Concernant la dernière étude sur questionnaire (testant les performances langagières de l'enfant préverbal après nos observations ; article 4, partie 5.1), il s'agissait de construire des scores de performance aux ajustements *unimodal* et *intermodal* (selon la méthode détaillée dans l'article 4, partie 5). Ces scores constituaient ainsi des variables indépendantes testées sur les scores de performances lexicales, syntaxiques et pragmatiques du questionnaire de *Développement du Langage de Production en Français* (DLPF ; Bassano et al., 2005, 2020). Sur cette base ont d'abord été conduites des sélections de modèles pour chacun de ces scores du DLPF puis des corrélations de Spearman (Siegal & Castellan, 1988).

Tous les tests ont été effectués à l'aide du logiciel R v3.6.1 (http://cran.r-project.org) en considérant les p-values égales ou inférieures à 0,05 comme étant significatives.

3.
COMMUNICATION SENSIBLE A L'ATTENTION MATERNELLE CHEZ LE CHIMPANZE IMMATURE (*PAN TROGLODYTES*)

3.1. ARTICLE 1: SENSITIVITY TO THE COMMUNICATIVE PARTNER'S ATTENTIONAL STATE: A DEVELOPMENTAL STUDY IN THE WILD CHIMPANZEE (*PAN TROGLODYTES SCHWEINFURTHII*)

Mawa Dafreville[1]*, Catherine Hobaiter[2], Michèle Guidetti [1], David Sillam-Dussès[3], Marie Bourjade[1,4]

[1] CLLE, Université de Toulouse, CNRS, Toulouse, France
[2] Origins of Mind, School of Psychology and Neuroscience, University of St Andrews, St Andrews, Scotland
[3] Laboratoire d'Ethologie Expérimentale et Comparée UR 4443, Université Sorbonne Paris Nord, Villetaneuse, France
[4] Institut Universitaire de France

Article publié dans American Journal of Primatology en 2021

Synthèse de l'article 1

Contexte : Nous disposons de relativement peu d'information quant aux développement des compétences sociocognitives nécessaires à l'émergence de la communication intentionnelle. C'est pourquoi, cette étude vise à investiguer la capacité du jeune chimpanzé à ajuster ses signaux à l'attention visuelle maternelle (i.e. la *communication sensible à l'attention du destinataire*), considérée comme un des principaux marqueurs de communication intentionnelle.

Méthode : En condition écologique, onze jeunes chimpanzés ont été observés en interaction avec leur mère : trois *enfants* (0-4 ans), quatre *juvéniles* (5-9 ans) et quatre *adolescents* (10-15 ans). La *communication sensible à l'attention du destinataire* a été testée à partir de descripteurs d'ajustement *unimodal* et *intermodal*, respectivement définis par la capacité de l'enfant à traiter les mères visuellement inattentives (i) en évitant de produire des signaux *silencieux-visuels* et/ou (ii) en favorisant la production de signaux perceptibles par l'utilisation de signaux *audibles* ou *tactiles*.

Résultats : Nos résultats montrent une diminution progressive de l'utilisation de gestes silencieux-visuels quand la mère n'est pas visuellement attentive avec l'âge. Les jeunes chimpanzés montraient un ajustement *intermodal* à partir du stade juvénile et un l'ajustement *intermodal* à partir du stade adolescent. De plus, durant l'ajustement *intermodal*, les jeunes chimpanzés ajustaient préférentiellement leurs gestes tactiles qu'audibles jusqu'à ce que la tendance s'inverse à l'adolescence

Conclusion : Dans l'ensemble, cette étude montre que les jeunes chimpanzés n'ajustaient pas pleinement leurs gestes à l'attention visuelle maternelle avant l'adolescence mais qu'au stade juvénile, ils présentaient un profil intermédiaire. En effet, les juvéniles montraient de l'ajustement *intermodal* mais non encore *unimodal* et une grande variabilité interindividuelle quant à l'émergence de ces ajustements.

Abstract

Gestural communication permeates all domains of chimpanzees' social life and is intentional in use. However, we still have only limited information on how young apes develop the socio-cognitive skills needed for intentional communication. In this cross-sectional study, we document the development of behavioural adjustment to the recipient's visual attention – considered a hallmark of intentional communication – in wild immature chimpanzees' gestural communication. We studied 11 immature chimpanzees (*Pan troglodytes schweinfurthii*): 3 infants, 4 juveniles, and 4 adolescents gesturing towards their mother. We quantified silent-visual, audible, and contact gestures indexed to maternal visual attention and inattention. We investigated *unimodal adjustment,* defined by the capacity of young chimpanzees to deploy *fewer* silent-visual signals when their mothers did not show full visual attention towards them as compared to when they did. We then examined *cross-modal adjustment,* defined as the capacity of chimpanzees to deploy *more* audible-or-contact gestures gestures than silent-visual gestures in the condition where their mothers did not show full visual attention as compared to when they did. Our results show a gradual decline in the use of silent-visual gestures when the mother is not visually attentive with increasing age. The absence of silent-visual gesture production toward a visually inattentive recipient (complete *unimodal* adjustment) was not fully in place until adolescence. Immature chimpanzees used more audible-or-contact gestures than silent-visual ones when their mothers did not show visual attention and vice-versa when they did. This *cross-modal* adjustment was expressed in juveniles and adolescents but not in infants. Overall, this study shows that infant chimpanzees were limited in their sensitivity to maternal attention when gesturing, whereas adolescent chimpanzees adjusted their communication appropriately. Juveniles present an intermediate pattern with cross-modal adjustment preceding unimodal adjustment and with variability in the age of onset.

Keywords: Intentional communication, visual attention, chimpanzees, mother-infant dyads, development

Introduction

Gestural communication permeates all aspects of great apes' social lives from infancy to old age, and from soliciting nursing or grooming, to soliciting sex or reconciliation. Great apes employ their gestures in communication that is directed to a particular partner (Genty et al., 2009; Leavens et al., 2004a; Liebal et al., 2006; Pika et al., 2003, 2005; Tomasello et al., 1989), and show sensitivity to their partner's visual attention by shifting the modality of their gestural signals to match the channels of information that their partner is able to perceive (Genty et al., 2009; Hobaiter & Byrne, 2011a; Liebal et al., 2004, 2005; Tomasello et al., 2007). This communicative adjustment to the recipient's state of visual attention (their ability to receive the visual information in the signal) is a key feature of intentional communication, and one of the criteria used to infer that a signal is intentionally produced (that is directed to a particular partner in order to achieve a particular goal that the signaller has in mind; Fröhlich et al., 2018; Leavens et al., 2004a).

In order to adjust communication to a partner, apes must be capable of discriminating subtle cues of attention in others (Hostetter et al. 2007; Tempelmann & Kaminski, 2011) or even, for some authors, assigning them a mental state (i.e., here attentional state; Dennett, 1983). Human children show a transition from perlocutionary (broadcast) to illocutionary (targeted and intentional) communication over the first year of life (Bates et al., 1975). A perlocutionary act creates an effect on the audience (e.g. a parent's response to the hunger cry of a newborn infant), whereas an illocutionary act is a conventional social act, recognized by both signaller and recipient that necessarily implies intentional communication on the part of the signaller (e.g. indicating an object with a pointing gesture; Bates et al., 1975). Longitudinal studies of great ape infant communication remain rare, and we still have only limited information on how young apes develop the socio-cognitive skills needed to recognise other individuals' behaviour and/or minds, in order to produce intentional gestural communication.

The shared intentional nature of both great ape gestural communication and human language has been used to argue that gestures may have represented a precursor state in human language evolution (Arbib et al., 2008; Armstrong & Wilcox, 2007; Corballis, 2002;

Hewes, 1992; Tomasello, 2008). Given the importance of intentionality for this argument, there has been substantial focus on establishing its presence, and different research groups provide particular behavioural criteria to define intentional signals (Leavens et al., 2005; Townsend et al., 2017). Three main hallmarks are regularly employed: *audience checking*, *persistence to the goal*, and *sensitivity to the recipient's attentional state* (Leavens et al., 2005; Townsend et al., 2017). While these stem from the research describing the transition to illocutionary communication in human infants (Bates et al., 1975), research on great ape gesture has largely focused on using associated criteria to detect intentional communication in juvenile and mature individuals where it is already clearly present (Cartmill & Byrne, 2007, 2010; Christiansen & Kirby, 2003; Hobaiter & Byrne, 2011a; Leavens et al., 2005; Liebal et al., 2004).

Sensitivity to the recipient's attentional state is operationalized in terms of communication match/mismatch between the sensory modality of the signal and the modalities that can be perceived by the recipient. Signals *match* recipient's attention when they are conveyed in a modality that can be perceived by the recipient. Communication *mismatch* occurs when the signaler addresses a visually inattentive recipient with a silent-visual signal (waving when your recipient has turned away from you), or an auditorily inattentive recipient with an audible signal (talking across a noisy bar). Individuals display attention-sensitive signaling when communication mismatches are absent or in a significant lower proportion than communication matches. Two descriptors of communication match/mismatch can be used to address attention-sensitive signaling. *Unimodal adjutment* refers to the capacity of the signaler to deploy *fewer* signals of a given modality '*x*' when they are unable to be perceived, as compared to when they are able to be percieved, i.e., avoiding communication mismatch. *Cross-modal adjustment* refers to the capacity of the signaler to not only inhibit the production of signals of modality '*x*' that cannot be perceived by the recipient, but also to deploy more signals that include other modalities (i.e. '*y*' and/or '*z*') that can be perceived, rather than signals of modality '*x*', as compared to the condition in which signals of modality '*x*' can be perceived, i.e., switching across modalities to favor communication match.

Many gestures incorporate multiple sensory modalities – all gestures include a visible component, but some also include physical contact or audible sound (Hobaiter & Byrne, 2011a; Pika et al., 2003; Tomasello et al., 1994, 1997). Visual attention is relatively easy to infer, for example via gaze and head direction (Hostetter et al., 2001, 2007; Kaminski et al., 2004; Leavens et al., 2004a,2004b; Leavens et al., 2010; Povinelli et al., 2003; Tempelmann et al., 2011); however, it is much more difficult for a human observer to detect another individual's attention towards audible or tactile signals, as it does not require the recipient to display an observable position. For example: many primates, including humans, tilt their head and orient an ear to a sound they are attending to, but we are not incapable of receiving signals in the auditory modality from other positions. While primate recipients have physical limitations on the auditory or contact information they can detect (for example sounds above or below a particular frequency), these are relatively stable – and do not vary substantially with the particular orientation and environment of the signaler-recipient pair. It is possible that a recipient who is paying attention to one set of sounds may be inhibited (or primed) in its ability to perceive another, or that pre-existing physical contact with a signaller alerts the recipient to the possibility of further information in that modality. For example, the physical contact between the mother and her offspring could be defined as a specific form of parental engagement involving physical attention from both mother and infant (Falk, 2004; Mehr & Krasnow, 2017). But, neither case provides observers with a consistent externally observable indication of these states of attention. As a result, studies exploring the sensitivity of a signaller's to their audiences attention have focused on visual-silent gesture use (Fröhlich et al., 2018; Hobaiter & Byrne, 2011a; Hostetter et al., 2007a; Kaminski et al., 2004; Leavens et al., 2004a, 2010; Liebal et al., 2004; Tempelmann & Kaminski, 2011; Tomasello, 2008).

One early hypothesised function for gestures that include audible and contact components was that they served only to regain the recipient's visual attention in cases of visual inattention (*manipulation* by means of *attention-getters*; Tomasello et al., 1994). If this were the case, these gestures should typically be followed by an additional visual signal, should not be used individually, and should only be used when visual attention was unavailable. However, studies of great apes gesturing across age groups found little evidence for attentional *manipulation* (as sequences of audible/contact + silent-visible gesturing;

Gorilla gorilla: Genty et al., 2009; *Pan troglodytes:* Hobaiter & Byrne, 2011b; Liebal et al., 2004). Instead, where visual attention was not present, signallers of any age either moved to place themselves in the line of sight (*unimodal adjustment*; *Pan troglodytes, Pan paniscus:* Liebal et al., 2004b), and/or selected a gesture that conveyed the information in an alternative modality (*cross-modal adjustment*), for example through sound or touch (*Pan troglodytes*, mixed ages: Hobaiter & Byrne, 2011a; *Pan paniscus*, infants and juveniles: Pika et al., 2005); suggesting that gestures of all modalities convey information, including in non-visual channels.

Further evidence for chimpanzees' ability to adjust their signal use comes from captive studies employing *food-requesting* paradigms. Here, captive chimpanzees (*Pan troglodytes*) need to communicate with a human experimenter to access food. Chimpanzees were capable of *unimodal adjustment* of silent-visual signals: producing significantly more silent or silent-gesture+vocal communication when human experimenters were visually attentive than when they were not (*Pan troglodytes*: (adults and adolescents) Hostetter et al., 2001, (mixed ages) 2007b; (mixed ages) Kaminski et al., 2004; (adults) Leavens et al., 2004b, (adults) 2010; (5-year old chimpanzees) Povinelli et al., 1996; *Pongo pygmaeus, Gorilla gorilla, Pan paniscus, Pan troglodytes* (mixed ages): Tempelmann et al., 2011). Nevertheless, like studies of their conspecific communication, the majority of these experiments focused on the behaviour of mature great apes, leaving open questions on when and how modal adjustment in chimpanzee communication emerges.

Only a few studies have examined sensitivity to the recipient's visual attention in immature apes (e.g. Fröhlich et al., 2019; Tomasello et al., 1997). In a young chimpanzee's life – the mother is typically a privileged partner, and mother-infant relationships are considered a "developmental niche" that shapes behavioural ontogeny (Stamps, 2003; West et al., 2003). The emergence of the different gestural modalities appears to parallel changes in mother-infant interactions. Contact and silent-visual gestures appear earlier in young chimpanzees (around 8 months of age) than audible gestures (around 18 months of age), and the use of silent-visual gestures increases while that of contact and audible gestures decreases with age (Schneider et al., 2012). In chimpanzees, body contact with the mother gradually decreases across infancy, and is limited by the time weaning occurs at around 5-years of age (Clark, 1977; Nishida, 2012). Over the juvenile period (from 5-years to puberty

at ~10 years), proximity to the mother further decreases, with an increasing tendency to move out-of-sight and at times to travel independently. Thus, during their development, young chimpanzees experience progressively increasing physical distances to their mother, potentially promoting the use of distal communication and hence the use of more silent-visual and/or audible gestures (Fröhlich et al., 2016; Lonsdorf et al., 2014).

Here, we document the development of young chimpanzees' gestural communication within the mother-infant dyad from infancy to adolescence. We focus on differences in the modality of immature chimpanzees' gestural signals, given their mother's visual attention. First, we describe immature chimpanzees' use of the different modalities of gesture, and how these are combined into sequences. We then investigate whether immature chimpanzees employ *unimodal adjustment;* producing *fewer* silent-visual signals when their mothers did not show full visual attention towards them as compared to when they did. Second, we investigate whether immature chimpanzees employ *cross-modal adjustment*; producing *more* audible-or-contact gestures than silent-visual gestures where their mothers did not show full visual attention as compared to when they did (Leavens et al., 2004b, 2010; Liebal et al., 2004; Tomasello, 2008). To do so, we test whether they differentially produce silent-visual and audible-or-contact gestures across conditions of maternal visual attention.

Methods

The research adhered to the legal requirements of the countries in which it was conducted and to the principles of the "Ethical Treatment of Nonhuman Primates," as outlined by the American Society of Primatologists. All original video data collection occurred with permission from the Ugandan National Council for Science and Technology, the Uganda Wildlife Authority, and the Budongo Conservation Field Station, following ethical review by the University of St Andrews Animal Welfare and Ethics Committee, and followed the International Primatogical Society's Code of Best Practices for Field Primatology.

Study sites and subjects

Original video data for this study were extracted from the Great Ape Dictionary video database and then further coded (www.greatapedictionary.com). This database contains labelled video-data and a set of coded gestural signals produced by wild apes during communication across species and sites. The data coded for this study were originally collected in the Budongo Forest Reserve in North-West Uganda (Eggeling, 1947) from the Sonso community of East African chimpanzees (*Pan troglodytes schweinfurthii*), and are fully reported in Hobaiter & Byrne (2011a). Original data collection started in October 2007. At this time, the community consisted of 81 named individuals including 18 infants (3 males and 15 females) aged between 0 and 4 years old, 15 juveniles (6 males and 9 females) aged between 5 and 9 years old, and 16 adolescents (10 males and 6 females) aged between 10 and 15 years old (more details: Hobaiter & Byrne, 2011a; Reynolds, 2006).

In this cross-sectional study, we focused on development across three developmental periods: infant (0-54 months), juvenile (55-102 months), and adolescent (103-180 months). The transition from infancy to the juvenile period is marked by the period in which the mother stops breastfeeding and the infant ceases to travel on the mother's body, and typically occurs around 4 or 5 years old in wild chimpanzees (Clark, 1977; Nishida, 2012). The transition from the juvenile period to the adolescence is defined by puberty which typically occurs around the age of 8 or 9 years old in wild chimpanzees (Pusey, 1990) and is marked by the development of secondary sexual characteristics, and full independence from the mother. The transition from the adolescent period to adulthood typically occurs around 15 years old, but is more variable (Nishida, 1988; Pusey, 1990). As detailed longitudinal data on individual behavioural or physiological changes were not available, we used fixed age categories across individuals. We extracted coded gesture data for individuals in each of the three developmental categories who had produced at least 10 gestures in communication with their mother. The final data set selected for this study included 264 gestures produced by 11 individuals (see Table 1): 3 infants (1 male and 2 females; aged between 15 and 53 months), 4 juveniles (2 males and 2 females; aged between 57 and 102 months), and 4 adolescents (1 male and 3 females; aged between 107 and 175 months, see Table 1).

TABLE 1. Observation age range, weighted mean age, sex, and relationship of the mother–offspring dyads. We show the age, sex, and mother-offspring dyads for all pairs included in the study, together with the number of gestures directed to the mother by the immature chimpanzees. Weighted mean age refers to the mean age weighted by the number of gestures at a given age (see methods).

Age category	Signaller (# gestures)	Mother	Sex	Observation age range (weighted mean age) *(in months)*
Infant	Karibu (44)	Kwera	Female	18- 49 (22.16)
	Klauce (16)	Kalema	Male	15-53 (29.13)
	Kox (17)	Kewaya	Female	15-25 (18.47)
Juvenile	Karo (21)	Kwera	Female	83-102 (93.19)
	Kasigwa (14)	Kutu	Male	60-64 (62.57)
	Night (53)	Nambi	Female	57-77 (63.21)
	Zak (38)	Zimba	Male	68-80 (77.63)
Adolescent	Kumi (11)	Kalema	Female	107-127 (115.73)
	Kwezi (13)	Kwera	Male	129-132 (166.92)
	Nora (19)	Nambi	Female	141-152 (145.68)
	Rose (18)	Ruhara	Female	154-175 (131.11)

Data collection

Original video data were collected by CH during three field periods between October 2007 and August 2009. For full details of the data collection method, see Hobaiter & Byrne (2011a). Data were collected using a *focal behaviour* sampling approach (Altmann, 1974). Any social interaction with the potential for communication was filmed, in practice all occasions in which two or more individuals were present and not occupied in solitary activities such as resting or self-grooming were considered an opportunity to film. When multiple opportunities to film were available, preference was given to those social contexts in which previous research suggested gestural communication was prolific (for example play) or to individuals for whom available data were more limited. A running record of the frequency with which individuals were observed was maintained in order to target infrequently sampled individuals. However, as East African chimpanzees are highly fission-fusion, with some individuals observed near daily, and others absent for several weeks or months, there remain differences in the frequency with which it was possible to observe particular individuals. As a result, some individuals are better represented, or represented over a longer period in the final dataset. Video data were recorded using a Sony Handycam (DCR-HC-55) and MiniDV tape. Videos were relatively short, typically lasting between 2

and 3 minutes, but a single communicative event often included several gestures towards the mother. 141 video clips were selected for further coding in this study (a mean of 12 recordings per individual).

Behaviour sampling/Coding procedure

Gestures were originally coded following the repertoire used in the original data collection, with interobserver coding conducted on 11% of the original data set by a 2nd coder with expertise in great ape gestural communication (full details in Hobaiter & Byrne, 2011a). To be considered for coding, all gestures must be produced during intentional communication by definition. In practice this meant that any potential cases of gesture must be accompanied by behavioural indications of intentional use (response waiting, audience checking, persistence). Gestures in which there was no clear evidence of intentional use, or where it was not possible to identify a specific recipient (for example more than one potential recipient) were excluded (interobserver reliability coding of Directedness, Cohen's kappa: K=0.69). In the current dataset, only those gestures that were marked as directed and where the recipient was the mother were considered for analysis.

The full repertoire included over 60 gesture types (Byrne et al., 2017; Hobaiter & Byrne, 2011a; interobserver reliability coding of Gesture Type, Cohen's kappa: K=0.86) gesture types were further categorised as a function of their predominant sensorial modality into three groups: auditory, contact, or silent-visual (Byrne et al., 2017). All gestures include a visual component. Gestures that additionally produce a sound as a consequence of the specific gesture action, were defined as *audible* gestures. Gestures that include physical contact as a consequence of the specific gesture action (which may or may not also produce a sound) were defined as *contact* gestures.

In this study, for each gesture, the following information was extracted from the original data: the main sensory modality (audible, silent-visual, or contact), the situational context, if the signal was produced as a single gesture or used in a sequence of gestures, and, if in a sequence, its position relative to other gestures. Given our focus on visual attention in assessing maternal attentional state, at times we considered the use of audible or contact gestures, both suitable for transmitting information to a visually inattentive recipient, as a combined category: audible-or-contact.

As in other studies (e.g. Hobaiter & Byrne, 2011b), we distinguish: *single gestures,* defined by the fact that a gesture is followed by a pause of 1 second or more from *gestures sequences* where 2 or more gestures are produced and separated by less than 1 second. The 1-second interval is taken to be the minimum time required for response-waiting to occur (e.g. Hobaiter & Byrne, 2011b), and so in this case gesture sequences are considered to be single communicative units, rather than reflecting persistence or modification of the communication by the signaller. Persistence would refer to the addition of further single gestures or gesture sequences after a pause of 1-second or more.

For the gesture cases extracted in this study the majority were recorded during social play (34%), feeding (33%), and grooming (23%); but data also included gestures recorded during traveling (5%), affiliation (2%), agonism (<1%), patrolling (<1%), resting (<1%), and unknown (<1%) contexts. To this coding, we added three new variables: maternal-attention, the *within-sequence modality*, and the *modal dimension* (unimodal, bimodal, or trimodal) of the sequence.

Maternal attention was defined by the mother's *visual attention* (original interobserver coding of the original variable, Attentional state: Cohen's kappa: K=0.63). We refined the coding of visual attention in the original data to employ a tighter definition of attention, so that in the current dataset a mother was only marked as 'attending' when she was looking directly at her offspring (marked using head orientation or gaze) immediately prior to the onset of the gesturing by the immature signaller. Mothers who had partial or peripheral view, or no view of their offspring were all marked as not attending. Thus, our attention variable had two conditions: 1) *visual-attentive*: the mother's face is fully oriented towards the infant; if the recipient's direct line of sight is considered to be at 0° the signaller could be located only within an arc 45° either side of this; 2) *visual-inattentive*: the mother's face is not fully oriented towards the infant; the signaller is located at an angle of 45° or greater from the recipient's direct line of sight. Examples of gestures produced by individual in each of the three developmental age categories and for each of the maternal attention conditions are available here: Dafreville et al_Sensitivity to the communicative partner's attention state.

We defined sequence modality for all sequences of two or more gestures. Sequence modality was described in two ways. 1) *within-sequence modality* indicated the succession of

gesture modalities that composed the sequence [for example a 3-gesture sequence composed of a *hit other – object move – stomp object* was coded as Contact-Audible-Audible (CAA); a 2-gesture sequence composed of *Dangle – Swing* was coded as Audible-Silent (AS)]. 2) *modal dimension,* which describes how many of the three modalities of information (silent-visual, audible, contact) were incorporated into the sequence (unimodal, bimodal, or trimodal).

Statistical analyses

We tested the effect of increasing age on the capacity for *unimodal adjustment* by fitting generalized linear mixed models (GLMM) using a Poisson error distribution and a log link function (Bolker et al., 2009) using the function *glmer* of the R package *lme4* version 1.1-23 with the optimizer function of *aictab* of the R package *AICcmodavg* to calculate AICc that is a transformation of the Akaike's Information Criterion (AIC) used for small sample. We coded a 'match' (0) when an immature chimpanzee produced a silent-visual gesture to a visually attentive mother, and a 'mismatch' (1) when they produced a silent-visual gesture to a visually inattentive mother. As fixed effects we included age of the signaler at the time of the gesture (in months), and we included sex of the signaler (male, female) as a fixed control factor. Signaler identity was included as a random effect. Interactions between fixed effects and random slopes were not included due to incomplete combination matrices. As an overall test of the predictor (age) on a decreasing unimodal mismatching, we compared the fit of the full model with that of a null model comprising only the fixed control effect (sex) and the random effect (signaler identity).

The sample size was small (11 individuals) and the data did not follow a normal distribution, so we performed nonparametric tests (Siegel & Castellan, 1988) and included a Bonferroni correction in the case of multiple comparisons. To test the capacity for *unimodal adjustment*, we compared the number of silent-visual gestures produced by immature chimpanzees when their mothers were visually attentive to when they were not. Note that visual attention was assessed immediately prior to the onset of gesture production, so after any adjustment of signalling location by the offspring had taken place. Firstly, we calculated a weighted mean age for each individual's data (in each maternal visual condition). To do so we calculated mean age weighted by the number of gesture cases recorded at each age point. For example: Karibu produced 3 gestures at 18-months, 10 at

19-months, 9 at 20-months, 9 at 21-months, 5 at 22-months, 7 at 29-months, and 1 at 49-months. The weighted mean age for her gesturing was (3*18+10*19+9*20+9*21+5*22+7*29+1*49)/44, or 22.16 months. Spearman rank order correlation coefficients were used to test the correlation between the proportion of silent-visual gestures and individual weighted mean age in each maternal visual condition.

In unimodal adjustment, as the silent-visual gestures use in attention and inattention conditions are symmetrical, we only tested the correlation between the proportion of silent-visual gestures and individual mean age when the mother shows visual inattention. We then used Fisher exact probabilities tests for independent samples to test the effect of age category on the distribution of silent-visual gestures by condition. To test the capacity for *cross-modal adjustment*, we compared the proportion of audible-or-contact gestures and silent-visual gestures across conditions of maternal visual attention, and we then specifically compared the use of the two types of audible-or-contact gestures in the case of maternal visual inattention. Spearman rank order correlation coefficients were used to test the correlation between either the proportions of silent-visual or audible-or-contact gestures, and individual weighted mean age for each maternal visual condition. Then we used Fisher exact probabilities tests for matched-samples within each age category to test the capacity to vary the production of audible-or-contact gestures and silent-visual gestures across the levels of maternal visual attention. To test the modal preference for *cross-modal adjustment*, we used Spearman rank order correlation coefficients and Fisher exact probabilities tests to test the effect of age category on the proportions of audible and contact gestures when the mother showed visual inattention.

To represent active adjustment of the different gesture modalities towards the maternal visual attention state, we calculated the percentage deviation in the variation in use of audible-or-contact as compared to silent-visual gestures for each condition of maternal attention and age category (as Hobaiter & Byrne, 2011a). The deviation was calculated by $(\beta/\alpha-1)$ x100 with α = number of audible-or-contact gestures/ total number of gestures used in the age-range subgroup, and β = number of audible-or-contact gestures/ total number of gestures used in the condition and age-range subgroup.

As we measured visual attention immediately prior to the onset of gesturing, the impact of these states of attention should be most profound on the first gesture produced (whether as a single gesture, or the first in a sequence). Restricting our dataset to these cases

would have substantially reduced our statistical power; however, whenever possible we provide a matching analysis replicated with this smaller dataset in the supplementary materials, to give an indication of the patterns of gesture use in the case of first or single gestures only (see Table S1, S2 and S3 in *supplementary material*). All tests were conducted using R v3.6.1 software (http://cran.r-project.org) with p-value equal or lower than 0.05 required for significance. All statistical tests were two-tailed.

Results

Individuals from all three age categories produced (total n=264) gestures, including gesture types in each modality (silent-visual, audible, contact), and across both states of maternal visual attention (with and without full visual-attention; see Table 2 and S4 in *supplementary material*).

TABLE 2. Gesture cases given visual attention of the mother. Number of gestures produced by infant, juvenile, and adolescent chimpanzees in each modality, according to the visual attention of the mother immediately prior to gesture production.

Age category	Gestural channel	Visual attentive	Visual inattentive
Infant (N=3)	Audible	17	13
	Contact	8	29
	Silent-visual	3	7
	Total	***28***	***49***
Juvenile (N=4)	Audible	13	17
	Contact	11	31
	Silent-visual	31	23
	Total	***55***	***71***
Adolescent (N=4)	Audible	17	17
	Contact	4	2
	Silent-visual	21	0
	Total	***42***	***19***
Combined categories (N=11)	Audible	47	47
	Contact	23	62
	Silent-visual	55	30
	Total	***125***	***139***

How are gestures of different modalities combined within gesture sequences?

Within the 208 communication acts coded, single gestures were more prominent than sequences (single gestures n=164; sequences n=44), which ranged from 2 to 4 gestures in length. Infants produced a much larger proportion of their communications as sequences (total communications n=47; sequences n=21, 45%), than juveniles (total communications n=109; sequences n=16, 15%) or adolescents (total communications n=52; sequences n=7, 13%).

The use of multiple modalities of gesture within a sequence occurred at all age categories, and the use of unimodal or bimodal sequences was similarly distributed across all communications (unimodal n=21, bimodal n=22) and within each age category (infant: unimodal n=9, bimodal n=11; juvenile: unimodal n=8, bimodal n=8; adolescent: unimodal n=4, bimodal n=3). We did not observe any sequence with all three modalities of gesture (trimodal). Gesture sequences that included both audible-or-contact and silent-visual gestures (n=16) were not more likely to start with an audible-or-contact gesture (n=9) than with a silent-visual gesture (n=7; see Table S5 in *Supplementary material*).

Unimodal adjustment: does the use of unimodal matching vary with signaller age?

There was a clear influence of age on a decreasing unimodal mismatching (null model comparison: $X^2=11.584$, df=1, $p<0.001$). The production of unimodal mismatching, here the use of silent-visual gestures to a visually inattentive recipient, decreased with increasing age (see Table 3). The same patterns held for the subset of gestures that were produced individually or as the first gesture in a sequence (see Table S1 in *supplementary material*).

We tested the correlation between the use of silent-visual gestures and weighted mean age in each maternal visual condition (see Fig.1). We found that the use of silent-visual gestures correlated negatively with age when their mother shows visual inattention (Spearman rank order correlation coefficient, $r = -0.90$, $p = 0.001$, N=11). As silent-visual gestures use in attention and inattention conditions are symmetrical, the correlation is exactly inverse when mother shows visual inattention and p-value is the same. The same

patterns held for the subset of gestures that were produced individually or as the first gesture in a sequence (see Table S2 in *supplementary material*).

When we considered age as separate developmental categories (see Fig. S1 in *supplementary material*), infants produced relatively few silent-visual gestures (n= 10 out of 77; Table 2) and when they did so there was no evidence that they adjusted these to their mother's visual attention (mother visually inattentive: n= 7; mother visually attentive: n=3). Juveniles produced a larger number of silent-visual gestures (n=54 of total n=126; Table 2) and were more likely to do so when their mother showed prior visual attention (mother visually attentive: n=31; mother visually inattentive: n= 23). Adolescent chimpanzees did not produce silent-visual gestures (n=21 out of 62, Table 2) when their mother did not show prior visual attention (mother visually inattentive: n= 0; mother visually attentive: n=21). We found no difference in the pattern of use between infant and juvenile chimpanzees (Fisher Test, effect size calculated as the odd ratio of log-likelihoods, OR = 1, p=0.68, N=7), but adolescents differed in their use from both infants (Fisher Test, OR=1, p<0.001, N=7) and juveniles (Fisher Test, OR=1, p<0.001, N=8; see Fig. S1 in *supplementary material*). The same patterns held for the subset of gestures that were produced individually or as the first gesture in a sequence (see Table S2 in *supplementary material*).

Individually, none of the infants showed reduced use of silent-visual gestures when their mother did not show full visual attention; whereas half of the juveniles (2 on 4) and all adolescents did so (see Fig. 2).

Cross-modal adjustment: Does the use of different gestural modalities vary according to maternal visual attention?

We tested the correlation between the use of silent-visual gestures or audible-or-contact gestures and weighted mean age in each maternal visual condition (see Fig.S2). We found that the use of audible-or-contact gestures correlated positively with age when the mother showed visual inattention (Spearman rank order correlation coefficient, r = 0.69, p = 0.033, N=11) while there was no significant correlation when the mother showed visual attention (Spearman rank order correlation coefficient, r = - 0.42, p = 0.248, N=11). As the use of silent-visual gestures and audible-or-contact ones were symmetrical, the results of correlation are exactly inverse for silent-visual gestures and p-values are the same within

each condition. The same patterns held for the subset of gestures that were produced individually or as the first gesture in a sequence (see Table S2 in *supplementary material*).

As the different modalities are not represented with equal frequency in the overall repertoire of gesture types, or in the subsets of gesture types used by each age category (see Table S4 in *supplementary material*), there may be biases in their use because of the availability of gesture types in an appropriate modality. To address this, we explored variation in use from the overall average gesture use within an age category according to the mother's state of visual attention. We found no evidence for *cross-modal* adjustment of silent-visual or audible-or-contact gesture types in infant chimpanzees (mother visually attentive: audible-or-contact gestures n=25, silent-visual gestures n=3; mother visually inattentive: audible-or-contact gestures n=42, silent-visual gestures n=7; Fisher Test, OR=1, p=1, N=3; see Fig.3). However, both juvenile and adolescent chimpanzees increased their use of silent-visual gestures and decreased their use of audible-or-contact gestures when their mothers were attentive, and vice-versa when they were inattentive prior to gesturing (<u>in juveniles</u>: mother visually attentive: audible-or-contact gestures n=24, silent-visual gestures n=31; mother visually inattentive: audible-or-contact gestures n=48, silent-visual gestures n=23; Fisher Test, OR=1, p=0.043, N=4; <u>in adolescents</u>: mother visually attentive: audible-or-contact gestures n=21, silent-visual gestures n=21; mother visually inattentive: audible-or-contact gestures n=19, silent-visual gestures n=0; Fisher Test, OR=1, p<0.001, N=4). When we replicated these tests on the subset of gestures that were produced individually or as the first gesture in a sequence, the same patterns were found for adolescents, but neither juveniles nor infants showed any evidence of *cross-modal* adjustment (see Table S2 and S3 in *supplementary material*).

How are audible-or-contact gestures used when the mother does not have visual attention?

We tested the correlation between the proportions of audible-or-contact gestures and weighted mean age when mother showed visual inattention. We did not find any significant correlation (Spearman rank order correlation coefficient, r = 0.03, p = 0.883, N=11). The same patterns held for the subset of gestures that were produced individually or as the first gesture in a sequence (see Table S2 in *supplementary material*).

When the mother did not show full visual attention towards their offspring, we found a difference in the relative use of different types of audible-or-contact gestures across the three age categories (infant: audible gestures n=13, contact gestures n=29; juvenile: audible gestures n=17, contact gestures n=31; adolescent: audible gestures n=17, contact gestures n=2; Fisher Test, OR=2, p<0.001, N=11; Fig. 4). Both infants and juveniles employed more contact gestures and did so in a similar way (Fisher Test, OR=1, p=1, N=7), whereas adolescents employed more audible gestures, and did so to an extent that differed from both infants (Fisher Test, OR=1, p<0.001, N=7) and juveniles (Fisher Test, OR=1, p<0.001, N=8; Fig. 4). The same pattern was found when we analysed only the subset of gestures produced either singly or as the first in a sequence (see Table S1 in *supplementary material*).

Discussion

This study investigated the development of the ability of immature wild chimpanzees to adjust their gesturing to the visual attention of a recipient, here their mother's. *Sensitivity to the recipient's attention state* is considered a hallmark of intentional communication and one of the core criteria used to infer that a gesture is produced intentionally (e.g. Fröhlich et al., 2018; Leavens et al., 2005; Pika et al., 2005; Tomasello et al., 1997). We find that as immature chimpanzees aged, they showed a decline in their use of silent-visual gestures when the mother's full visual attention was not available, indicating increasing inhibitory abilities that support *unimodal adjustment*. None of the infants showed adjustment of their silent-visual signals given the attentional state of their mother, while two out of the four juveniles and all the adolescents did so. Immature chimpanzees employed *cross-modal adjustment* from the juvenile period onwards, using more audible-or-contact gestures than silent-visual gestures when their mothers did not show full visual attention prior to gesturing, and vice-versa when they did. All immature chimpanzees produced gestures in sequences comprising one or two modalities, but did not initiate their sequences with a particular modality.

Juvenile and adolescent, but not infant, chimpanzees displayed cross-modal adjustment to their mother's visual attention, favouring audible-or-contact gestures over

silent-visual gestures when their mothers were visually inattentive. When considering the subset of gestures produced individually or as the first gesture in a sequence, this pattern of adjustment was only found in adolescents. Taken together, these results suggest that infant chimpanzees gesturing was not sensitive to their mother's full visual attention, that of adolescents was, and juveniles displayed an intermediate pattern (Fig. 5).

We also found that infant and juvenile chimpanzees favoured contact over audible gestures to address a visually inattentive mother, whereas adolescents favoured audible gestures in these situations. These results are consistent with previous studies reporting a preference for tactile communication in younger primates (Schneider et al., 2012; Tomasello et al., 1997), and reflect patterns of increasing physical distance between mothers and their offspring with age (Bard, 2019; Fröhlich et al., 2016; Lonsdorf et al., 2014), which may favour the use of audible over tactile gestures, which requires moving to within arms' reach. Our juvenile chimpanzees again showed an intermediate pattern, displaying the same preference for contact gestures as infants, but using these contact gestures to address their mother specifically when her full visual attention was not available.

Juvenile chimpanzees in the Sonso community regularly employ the widest variety of their gestural repertoire, as compared to younger or older individuals (Hobaiter & Byrne, 2011a), and similar results have been found in other primate species (e.g., *Gorilla gorilla*: Genty et al., 2009; *Papio anubis*: Molesti et al., 2020). Here, too, we found that juveniles employed a greater variety of silent-visual gestures as compared to infants and adolescents (see table S4 in *supplementary materials*). Thus, juveniles expand their use of silent-visual gesture types as they begin to display more sensitivity to their partners' visual attention, suggesting that the juvenile period is critical to exploring the repertoire of communicative gestures available and the appropriate ways of using gestures of different modalities.

Juvenile chimpanzees' use of cross-modal adjustment in their gesturing prior to unimodal adjustment was unexpected. From the perspective of executive functioning, unimodal adjustment relies on inhibitory control over the production of signals when a recipient is unable to perceive them. In contrast, cross-modal adjustment relies on flexibility in the deployment of signals across various sensory modalities. Inhibitory control is

amongst the earliest executive functions to emerge in human infants, while cognitive flexibility develops more slowly (Anderson, 2002). Following this framework, we would expect to observe unimodal adjustment preceding cross-modal adjustment (at least in human infants); however, empirical support for this developmental trajectory in humans (*Homo sapiens sapiens*) is also mixed. Liszkowski et al. (2008) showed that infants at 12 months old decreased their pointing gestures towards inattentive adults as compared to attentive adults, thereby showing unimodal adjustment. The onset of cross-modal adjustment has been reported prior to this age (~10 months; Wu & Gros-Louis, 2017), at the same age (~12 months; Igualada et al., 2015), and after this age (at ~18 months; Liszkowski et al., 2008). To our knowledge, there is no other developmental study directly comparing unimodal and cross-modal adjustment in non-human primates. However, the broader capacity to show *sensitivity to the recipient's attention state* has been investigated across diverse primates species, both in conspecific interactions and in interaction with human experimenters (e.g., captive population with an experimental design: *Pongo pygmaeus*, *Gorilla gorilla*: (adults) Botting & Bastian, 2019; (adults and adolescents) Poss et al., 2006; *Papio anubis*, mixed ages from 6 to 16 years old: Bourjade et al., 2014; *Macaca tonkeana,* mixed ages from 5 to 13 years old: Canteloup et al., 2015; *Cercocebus torquatus torquatus,* mixed ages from 2 to 23 years old: Maille et al., 2012; wild population with an observational method : *Macaca radiate*, mixed ages: Deshpande et al., 2018); but also across non-primates species in interaction with conspecifics (e.g., *Canis familiaris*, alduts and puppies: Horowitz, 2009; *Corvus corax*, ages from 1 to 2 years old: Pika & Bugnyar, 2011; several coral reef fishes species, mixed ages: Vail et al., 2013) and in interaction with humans (e.g., *Gymnorhina tibicen*, mixed ages: Kaplan, 2011; *Equus caballus*, mixed ages: Proops & McComb, 2010; *Canis familiaris*, adults: Horowitz, 2009; Kaminski et al., 2012; Savalli et al., 2014; Topál et al., 2014) . All these studies tested and found evidence for unimodal adjustment to the recipient's visual attention state but only a few tested and reported cross-modal adjustment (*Papio anubis,* mixed ages from 6 to 16 years old: Bourjade et al., 2014; *Canis familiaris*, adults and puppies: Horowitz, 2009) . All these studies tested and found evidence for unimodal adjustment to the recipient's visual attention state but only a few tested and reported cross-modal adjustment (*Papio anubis,* mixed ages from 6 to 16 years old: Bourjade et al., 2014; *Canis familiaris*, adults and puppies: Horowitz, 2009). Moreover, studies on executive function in baboons (*Papio papio*) have shown that subadults outperformed adults in

cognitive flexibility (Bonté et al., 2014), whereas the opposite pattern was found for inhibitory control (Fagot et al., 2008). Further investigation of a non-human primate-typical developmental pattern of executive control and flexibility is required to advance our understanding of the cognitive underpinnings of unimodal and cross-modal adjustment in chimpanzees.

In the present study, juveniles as a group demonstrated both aspects of sensitivity to attention but individual patterns differed sharply for unimodal adjustment, with Night and Zak (one of the youngest and one of the oldest) showing a clear adjustment of their silent-visual gestures to their mother's visual attention, while Karo and Kasigwa did not. Interestingly, Night and Zak also displayed a sharper cross-modal adjustment than the two others (see Figure S3 in *Supplementary material*) again suggesting that cross-modal adjustment may be a developmental precursor to unimodal adjustment in chimpanzees. Fröhlich et al. (2018) found no improvement in sensitivity to the recipient's attention when comparing infant and juvenile chimpanzees. However, each individual showed progression in adjusting its signal modality to the recipient's attention (i.e., within-age effect, see Fröhlich et al., 2018). Although the present sample was insufficient to perform a within-age study and required that we parse continuous development into artificially discreet age-categories, our results still suggest that attention-sensitive signalling may develop at a different pace across individuals.

One potential explanation of the relatively late expression of adjustment capacities in our data is that our definition of maternal visual attention was very restrictive. In our study, maternal visual attention was limited to where the mother had *full view* of the infant and excluded cases where the mother had *partial* or *peripheral view* (as well as those where she had no view at all). However, chimpanzee mothers likely continue to monitor their younger offspring in their peripheral vision (so called 'eyes in the back of their head'). Maternal visual attention may also be difficult to consistently detect; Kano and Tomonaga (2011) showed that chimpanzee eye saccades are shorter and more frequent than humans' and permit faster scans of the environment. If chimpanzee mothers reliably monitor their infants with peripheral vision and respond appropriately to their infants' gestures in these conditions, younger chimpanzees would experience relatively little learning pressure to

discriminate subtle differences between full and peripheral attention, and may take longer to fully develop this sensitivity. Once proximity to their mother increases as juveniles and then adolescents, both the mother's ability and need to visually monitor their offspring closely may decline, increasing the pressure on their offspring to refine modal adjustments in their signalling. Interestingly, this may also provide an explanation for why infants were found to prefer 'attending' to non-maternal, rather than maternal, partners (Fröhlich et al., 2018) – if 'mum is always watching' then the early onset of unimodal or cross-modal adjustment may be more easily detected in gesturing towards non-maternal partners. Similarly, differences in maternal attentiveness or mothering 'style' (Stanton et al., 2015) may contribute to individual variation in the onset of modal adjustments in gesturing.

Another possible explanation for the weak effect of maternal visual attention on the gesturing of infant and juvenile chimpanzees in our data may be that the majority of these gestures were recorded in the context of either social play (34%) or feeding (33%). In play, individuals may receive less pressure towards the efficient expression of their communication (Heesen et al., 2019). Fröhlich et al. (2018) examined the effect of communicative context on behavioural markers of intentional communication and while they found no difference between gesturing during play, travelling, and feeding on the sensitivity to recipient attention, they did find an effect on the occurrence of audience checking and persistence.

Conclusion and perspectives

This cross-sectional study on the development of the sensitivity to others' visual attention in chimpanzee gestural communication shows that immature chimpanzees may take several years to refine these skills. Infant chimpanzee did not display either unimodal or cross-modal adjustment and there was substantial variability in the age of onset, with only some juveniles and, finally, all adolescents showing appropriate adjustments. It remains unclear to what extent these findings reflect an inability to do so in infancy, or an absence of the need to do so, given mothers' tendency to monitor young infants closely. Our data suggest that cross-modal adjustment may be a developmental precursor of unimodal adjustment. Future studies should explore how the onset and expression of these abilities

are impacted by the identity and behaviour of social partners as well as the individual characteristics of the signaller.

Acknowledgements

We would like to thank all the staff of the Budongo Conservation Field Station. We thank the Uganda National Council for Science and Technology, the President's Office, the Uganda Wildlife Authority and the Uganda Forest Authority for permission to work in Uganda. Ethical approval for the original data collection was also provided by the University of St Andrews Animal Welfare and Ethics Committee. Fieldwork of CH was supported by grants from the Wenner-Gren Foundation (http://wennergren.org) and the Russell Trust. We also thank all the collaborators and students who have made this research possible in the Wild Minds Lab at St Andrews and in the MODAL Lab at Toulouse.

The authors declare that there are no conflict of interests.

References

Altmann, J. (1974). Observational study of behaviour: sampling methods. *Behaviour*, *49*, 227–265.

Anderson, P. (2002). Assessment and Development of Executive Function (EF) During Childhood. *Child Neuropsychology*, *8*(2), 71–82. https://doi.org/10.1080/09297049508400221

Arbib, M. A., Liebal, K., & Pika, S. (2008). Primate vocalization, gesture, and the evolution of human language. *Current Anthropology*, *49*(6), 1053–1076. https://doi.org/10.1086/593015

Armstrong, D., & Wilcox, S. (2007). *The Gestural Origin of Language* (Oxford Uni). Retrieved from https://doi.org/10.1093/acprof:oso/9780195163483.001.0001

Bard, K. (2019). Parenting in Nonhuman Primates. In *Handbook of Parenting*. https://doi.org/10.4324/9780429401459-3

Bates, E., Camaioni, L., & Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill-Palmer Quarterly of Behavior and Development*, *21*(3), 205–226.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology and Evolution*, *24*(3), 127–135. https://doi.org/10.1016/j.tree.2008.10.008

Bonté, E., Kemp, C., & Fagot, J. (2014). Age effects on transfer index performance and executive control in baboons (Papio papio). *Frontiers in Psychology*, *5*(MAR), 1–7. https://doi.org/10.3389/fpsyg.2014.00188

Botting, J., & Bastian, M. (2019). Orangutans (Pongo pygmaeus and Hybrid) and Gorillas (Gorilla gorilla gorilla) Modify Their Visual, but Not Auditory, Communicative Behaviors, Depending on the Attentional State of a Human Experimenter. *International Journal of Primatology*, *40*(2), 244–262. https://doi.org/10.1007/s10764-019-00083-0

Bourjade, M., Meguerditchian, A., Maille, A., Gaunet, F., & Vauclair, J. (2014). Olive baboons, Papio anubis, adjust their visual and auditory intentional gestures to the visual attention of others. *Animal Behaviour*, *87*, 121–128. https://doi.org/10.1016/j.anbehav.2013.10.019

Byrne, R. W., Cartmill, E., Genty, E., Graham, K. E., Hobaiter, C., & Tanner, J. (2017). Great ape gestures:

intentional communication with a rich set of innate signals. *Animal Cognition*, *20*(4), 755–769. https://doi.org/10.1007/s10071-017-1096-4

Canteloup, C., Bovet, D., & Meunier, H. (2015). Do Tonkean macaques (Macaca tonkeana) tailor their gestural and visual signals to fit the attentional states of a human partner? *Animal Cognition*, *18*(2), 451–461. https://doi.org/10.1007/s10071-014-0814-4

Cartmill, E. A., & Byrne, R. W. (2007). Report Orangutans Modify Their Gestural Signaling According to Their Audience ' s Comprehension. *Current Biology*, *17*, 1345–1348. https://doi.org/10.1016/j.cub.2007.06.069

Cartmill, E. A., & Byrne, R. W. (2010). Semantics of primate gestures: Intentional meanings of orangutan gestures. *Animal Cognition*, *13*(6), 793–804. https://doi.org/10.1007/s10071-010-0328-7

Christiansen, M. H., & Kirby, S. (2003). *Language evolution* (Oxford University Press, ed.). https://doi.org/10.1093/acprof:oso/9780199244843.001.0001

Clark, C. . (1977). A preliminary report on weaning among Chimpanzees of the Gombe National Park, Tanzania. *Primate Bio-Social Development: Biological, Social and Ecological Determinants*, 235–260.

Corballis, M. . (2002). *From Hand to Mouth: The Origins of Language* (P. U. Press, ed.). Retrieved from http://hdl.handle.net/10822/951670

Dennett, D. C. (1983). Intentional systems in cognitive ethology: The "Panglossian paradigm" defended. *Behavioral and Brain Sciences*, *6*(3), 343–355. https://doi.org/10.1017/S0140525X00016393

Deshpande, A., Gupta, S., & Sinha, A. (2018). Intentional communication between wild bonnet macaques and humans. *Scientific Reports*, *8*(1), 1–12. https://doi.org/10.1038/s41598-018-22928-z

Eggeling, W. . (1947). Observations on the Ecology of the Budongo Rain Forest, Uganda. *Journal of Ecology*, *34*(1), 20–87. https://doi.org/10.1017/CBO9781107415324.004

Fagot, J., Bonté, E., & Hopkins, W. D. (2008). Age-Dependant Behavioral Strategies in a Visual Search Task in Baboons (Papio papio) and Their Relation to Inhibitory Control. *J Comp Psychol.*, *127*(2), 194–201. https://doi.org/10.1038/jid.2014.371

Falk, D. (2004). Prelinguistic evolution in early hominins: Whence motherese? *Behavioral and Brain Sciences*, *27*(04). https://doi.org/10.1017/S0140525X04000111

Fröhlich, M., Wittig, R. M., & Pika, S. (2016). Should I stay or should I go? Initiation of joint travel in mother-infant dyads of two chimpanzee communities in the wild. *Animal Cognition*, *19*. https://doi.org/10.1007/s10071-015-0948-z

Fröhlich, M., Wittig, R. M., & Pika, S. (2018). The ontogeny of intentional communication in chimpanzees in the wild. *Developmental Science*, *22*(1), 1–14. https://doi.org/10.1111/desc.12716

Genty, E., Breuer, T., Hobaiter, C., & Byrne, R. W. (2009). Gestural communication of the gorilla (Gorilla gorilla): repertoire , intentionality and possible origins. *Animal Cognition*, *12*, 527–546. https://doi.org/10.1007/s10071-009-0213-4

Genty, E., & Byrne, R. W. (2010). Why do gorillas make sequences of gestures? *Animal Cognition*, *13*(2), 287–301. https://doi.org/10.1007/s10071-009-0266-4

Harrison, X. A. (2015). A comparison of observation-level randomeffect and Beta-Binomial models for modelling overdispersion in Binomial data in ecology & evolution. *PeerJ*, *2015*(7). https://doi.org/10.7717/peerj.1114

Heesen, R., Hobaiter, C., Ferrer-I-Cancho, R., & Semple, S. (2019). Linguistic laws in chimpanzee gestural communication. *Proceedings of the Royal Society B: Biological Sciences*, *286*(20182900). https://doi.org/10.1098/rspb.2018.2900

Hewes, G. W. (1992). Primate Communication and the Gestural Origin of Language. *Current Anthropology*, *33*(S1), 65–84. https://doi.org/10.1086/204019

Hobaiter, C., & Byrne, R. W. (2011a). The gestural repertoire of the wild chimpanzee. *Animal Cognition*, *14*(5), 745–767. https://doi.org/10.1007/s10071-011-0409-2

Hobaiter, C., & Byrne, R. (2011b). Serial gesturing by wild chimpanzees: Its nature and function for communication. *Animal Cognition*, *14*(6), 827–838. https://doi.org/10.1007/s10071-011-0416-3

Horowitz, A. (2009). Attention to attention in domestic dog (Canis familiaris) dyadic play. *Animal Cognition*, *12*(1), 107–118. https://doi.org/10.1007/s10071-008-0175-y

Hostetter, A. B., Cantero, M., & Hopkins, W. D. (2001). Differential use of vocal and gestural communication by chimpanzees (Pan troglodytes) in response to the attentional status of a human (Homo sapiens). *Journal of Comparative Psychology*, *115*(4), 337–343. https://doi.org/10.1037/0735-7036.115.4.337

Hostetter, A. B., Russell, J. L., Freeman, H., & Hopkins, W. D. (2007). Now you see me, now you don't: Evidence that chimpanzees understand the role of the eyes in attention. *Animal Cognition*, *10*(1), 55–62. https://doi.org/10.1007/s10071-006-0031-x

Igualada, A., Bosch, L., & Prieto, P. (2015). Language development at 18 months is related to multimodal communicative strategies at 12 months. *Infant Behavior and Development*, *39*, 42–52. https://doi.org/10.1016/j.infbeh.2015.02.004

Kaminski, J., Call, J., & Tomasello, M. (2004). Body orientation and face orientation: Two factors controlling apes' begging behavior from humans. *Animal Cognition*, *7*(4), 216–223. https://doi.org/10.1007/s10071-004-0214-2

Kaminski, J., Schulz, L., & Tomasello, M. (2012). How dogs know when communication is intended for them. *Developmental Science*, *15*(2), 222–232. https://doi.org/10.1111/j.1467-7687.2011.01120.x

Kano, F., & Tomonaga, M. (2011). Species difference in the timing of gaze movement between chimpanzees and humans. *Animal Cognition*, *14*(6), 879–892. https://doi.org/10.1007/s10071-011-0422-5

Kaplan, G. (2011). Pointing gesture in a bird- merely instrumental or a cognitively complex behavior? *Current Zoology*, *57*(4), 453–467. https://doi.org/10.1093/czoolo/57.4.453

Leavens, D. A., Hopkins, W. D., & Thomas, R. K. (2004). Referential Communication by Chimpanzees (Pan Troglodytes). *Journal of Comparative Psychology*, *118*(1), 48–57. https://doi.org/10.1037/0735-7036.118.1.48

Leavens, D. A., Hostetter, A. B., Wesley, M. J., & Hopkins, W. D. (2004). Tactical use of unimodal and bimodal communication by chimpanzees, Pan troglodytes. *Animal Behaviour*, *67*(3), 467–476. https://doi.org/10.1016/j.anbehav.2003.04.007

Leavens, D. A., Russell, J. L., & Hopkins, W. D. (2005). Intentionality as measured in the persistence and elaboration of communication by chimpanzees (Pan troglodytes). *Child Development*, *76*(1), 291–306.

Leavens, D. A., Russell, J. L., & Hopkins, W. D. (2010). Multimodal communication by captive chimpanzees (Pan troglodytes). *Animal Cognition*, *13*(1), 33–40. https://doi.org/10.1007/s10071-009-0242-z

Liebal, K., Call, J., & Tomasello, M. (2004). Use of gesture sequences in chimpanzees. *American Journal of Primatology*, *64*(4), 377–396. https://doi.org/10.1002/ajp.20087

Liebal, K., Call, J., Tomasello, M., & Pika, S. (2004). To move or not to move: How apes adjust to the attentional state of others. *Interaction StudiesInteraction Studies Social Behaviour and Communication in Biological and Artificial Systems*, *5*(2), 199–219. https://doi.org/10.1075/is.5.2.03lie

Liebal, K., Pika, S., & Tomasello, M. (2006). Gestural communication of orangutans (Pongo pygmaeus). *Gesture*, *6*(1), 1–38. https://doi.org/10.1075/gest.6.1.02lie

Liszkowski, U., Albrecht, K., Carpenter, M., & Tomasello, M. (2008). Infants ' visual and auditory communication when a partner is or is not visually attending. *Infant Behavior & Development*, *31*, 157–167. https://doi.org/10.1016/j.infbeh.2007.10.011

Lonsdorf, E. V., Markham, A. C., Heintz, M. R., Anderson, K. E., Ciuk, D. J., Goodall, J., & Murray, C. M. (2014). Sex differences in wild chimpanzee behavior emerge during infancy. *PLoS ONE*, *9*(6), 1–9. https://doi.org/10.1371/journal.pone.0099099

Maille, A., Engelhart, L., Bourjade, M., & Blois-Heulin, C. (2012). To Beg, or Not to Beg? That Is the

Question: Mangabeys Modify Their Production of Requesting Gestures in Response to Human's Attentional States. *PLoS ONE*, (7), e41197. https://doi.org/10.1371/journal.pone.0041197

Mehr, S. A., & Krasnow, M. M. (2017). Parent-offspring conflict and the evolution of infant-directed song. *Evolution and Human Behavior*, *38*(5), 674–684. https://doi.org/10.1016/j.evolhumbehav.2016.12.005

Molesti, S., Meguerditchian, A., & Bourjade, M. (2020). Gestural communication in olive baboons (Papio anubis): repertoire and intentionality. *Animal Cognition*, *23*(1), 19–40. https://doi.org/10.1007/s10071-019-01312-y

Nishida, T. (1988). Development of Social Grooming between Mother and Offspring in Wild Chimpanzee. *Folia Primatologica*, *50*, 109–123. https://doi.org/10.1159/000156335

Nishida, T. (2012). *Chimpanzees of the lakeshore: Natural history and culture at Mahale* (Cambridge University Press, ed.). https://doi.org/https://doi.org/10.1017/CBO9781139059497

Pika, S., & Bugnyar, T. (2011). The use of referential gestures in ravens (Corvus corax) in the wild. *Nature Communications*, *2*(1), 560–565. https://doi.org/10.1038/ncomms1567

Pika, S., Liebal, K., & Tomasello, M. (2003). Gestural communication in young gorillas (Gorilla gorilla): Gestural repertoire, learning, and use. *American Journal of Primatology*, *60*(3), 95–111. https://doi.org/10.1002/ajp.10097

Pika, S., Liebal, K., & Tomasello, M. (2005). Gestural communication in subadult bonobos (Pan paniscus): Repertoire and use. *American Journal of Primatology*, *65*(1), 39–61. https://doi.org/10.1002/ajp.20096

Poss, S. R., Kuhar, C., Stoinski, T. S., & Hopkins, W. D. (2006). Differential Use of Attentional and Visual Communicative Signaling by Orangutans (Pongo pygmaeus) and Gorillas (Gorilla gorilla) in Response to the Attentional Status of a Human. *Expedition Field Techniques BATS*, *7*(September 2006), 1–6. https://doi.org/10.1002/ajp

Povinelli, D. J., Eddy, T. J., Hobson, R. P., & Tomasello, M. (1996). What young chimpanzees know about seeing. *Monographs of the Society for Research in Child Development*, *61*(3). https://doi.org/10.2307/1166159

Povinelli, D. J., Theall, L. A., Reaux, J. E., & Dunphy-Lelii, S. (2003). Chimpanzees spontaneously alter the location of their gestures to match the attentional orientation of others. *Animal Behaviour*, *66*(1), 71–79. https://doi.org/10.1006/anbe.2003.2195

Proops, L., & McComb, K. (2010). Attributing attention: The use of human-given cues by domestic horses (Equus caballus). *Animal Cognition*, *13*(2), 197–205. https://doi.org/10.1007/s10071-009-0257-5

Pusey, A. E. (1990). Behavioural Changes at Adolescence in Chimpanzees. *Behaviour*, *115*(3–4), 203–246. https://doi.org/10.1163/156853990X00581

Reynolds, V. (2006). The Chimpanzees of the Budongo Forest. *International Journal of Primatology*, *27*(5). https://doi.org/10.1007/s10764-006-9075-7

Savalli, C., Ades, C., & Gaunet, F. (2014). Are dogs able to communicate with their owners about a desirable food in a referential and intentional way? *PLoS ONE*, *9*(9). https://doi.org/10.1371/journal.pone.0108003

Schneider, C., Call, J., & Liebal, K. (2012). Onset and Early Use of Gestural Communication in Nonhuman Great Apes. *American Journal of Primatology*, *74*(1), 102–113. https://doi.org/10.1002/ajp.21011

Stamps, J. (2003). Behavioural processes affecting development : Tinbergen 's fourth question comes of age. *Animal Behaviour*, *66*(1), 1–13. https://doi.org/10.1006/anbe.2003.2180

Stanton, M. A., Heintz, M. R., Lonsdorf, E. V., Santymire, R. M., Lipende, I., & Murray, C. M. (2015). Maternal Behavior and Physiological Stress Levels in Wild Chimpanzees (Pan troglodytes schweinfurthii). *Int J Primatol.*, *36*(3), 473–488. https://doi.org/10.1007/s10764-015-9836-2.Maternal

Tempelmann, S., Kaminski, J., & Liebal, K. (2011). Focus on the essential: All great apes know when others are being attentive. *Animal Cognition*, *14*(3), 433–439. https://doi.org/10.1007/s10071-011-0378-5

Tomasello, M. (2008). *Origins of Human Communication* (The MIT Press, ed.).

https://doi.org/https://doi.org/10.7551/mitpress/7551.001.0001
Tomasello, M., Call, J., Nagell, K., Olguin, R., & Carpenter, M. (1994). The learning and use of gestural signals by young chimpanzees: A trans-generational study. *Primates*, *35*(2), 137–154. https://doi.org/10.1007/BF02382050
Tomasello, M., Call, J., Warren, J., Frost, G. T., Carpenter, M., & Nagell, K. (1997). The ontogeny ofchimpanzee gestural signals: A comparison across groups and generations. *Evolution of Communicati*, *1*(2), 223–259. https://doi.org/10.1075/eoc.l.2.04tom
Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A New Look at Infant Pointing. *Child Development*, *78*(3), 705–722. https://doi.org/https://doi.org/10.1111/j.1467-8624.2007.01025.x
Tomasello, M., Gust, D., & Frost, G. T. (1989). A longitudinal investigation of gestural communication in young chimpanzees. *Primates*, *30*(1), 35–50. https://doi.org/10.1007/BF02381209
Topál, J., Kis, A., & Oláh, K. (2014). Dogs' Sensitivity to Human Ostensive Cues: A Unique Adaptation? A Unique Adaptation? In *The Social Dog: Behavior and Cognition*. https://doi.org/10.1016/B978-0-12-407818-5.00011-5
Townsend, S. W., Koski, S. E., Byrne, R. W., Slocombe, K. E., Bickel, B., Boeckle, M., … Manser, M. B. (2017). Exorcising Grice's ghost: an empirical approach to studying intentional communication in animals. *Biological Reviews*, *92*(3), 1427–1433. https://doi.org/10.1111/brv.12289
Vail, A. L., Manica, A., & Bshary, R. (2013). Referential gestures in fish collaborative hunting. *Nature Communications*, *4*, 1765–1767. https://doi.org/10.1038/ncomms2781
West, M. J., King, A. P., & White, D. J. (2003). The case for developmental ecology. *Animal Behaviour*, *66*, 617–622. https://doi.org/10.1006/anbe.2003.2221
Wu, Z., & Gros-Louis, J. (2017). The Value of Vocalizing: 10-Month-Olds' Vocal Usage Relates to Language Outcomes at 15 Months. *Infancy*, *22*(1), 23–41. https://doi.org/10.1111/infa.12150

TABLE 3. Parameter estimates for the tested model – GLMM with the number of silent-visual gestures (n=85) mismatching maternal visual attention as dependent variable and age (in months) as fixed effect. There was a clear influence of age on unimodal matching (null model comparison: X^2=11.584, df=1, p<0.001).

Parameter	Unimodal mismatching ~ Age					
	Estimate	**SE**	**Lower CI**	**Upper CI**	**Z-value**	**P-value**
(Intercept)	1,029	0,379	0.286	1.773	(1)	(1)
Age(2)	-0,018	0,005	0.029	0.008	-3.428	p<0.001

(1) Not indicated because of limited interpretive value

(2) Age: refers to the individual age of the number of unimodal matching (see methods)

Fig. 1. Effect of age on the distribution of silent-visual gestures according to the maternal visual attention for each individual (N=11). The effect of weighted mean age on the use of silent-visual gestures according to the maternal visual attention was significant (Spearman rank order correlation coefficient, p-value = 0.0001, r= +/-0.90).

Fig. 2. Number of silent-visual gestures for each individual with respect to maternal visual attention for infant (A; N=3), juvenile (B; N=4), and adolescent (C; N=4) chimpanzees. The lines illustrate individual shifts in behaviour across the two discrete levels of the attention condition.

Fig. 3. Variation in use of audible-or-contact signals and silent-visual gestures with respect to maternal visual attention by age category (**N=11,** *P<0.05. ***P<0.001). The deviations above and below the zero-line show changes (plus standard error bar) in the use of each modality, according to the maternal state of attention prior to gesturing, from the overall average use of that modality in gesturing.

Fig. 4. Proportion of audible and contact gestures used when the mother was visually inattentive (**N=11,** *P<0.05, **P<0.01, ***P<0.001). Large black circles represent mean proportion per subject. Median (horizontal lines), quartiles (boxes), percentiles (2.5% and 97.5%. vertical lines) and outliers (small black circles) are indicated. Differential use patterns are compared from one age category to another.

Fig. 5. Developmental trajectories of sensitivity to the recipient's attention state by chimpanzees' age category on the full sample of gestures.

Fig 1.

Percentage of silent-visual gestures
100%
50%
0%
20
40
60
80
100
120
140
160
Age mean
Mother shows visual attention
Mother shows visual inattention

Fig 2.

A
Number of signals
20
15
10
5
0
Karibu
Klauce
Kox
Visual inattention
Visual attention
Signaller
Karibu
Klauce
Kox
B
Night
Kasigwa
Zak
Karo
Signaller
Karo
Kasigwa
Night
Zak
C
Nora
Kwezi
Rose
Kumi
Signaller
Kumi
Kwezi
Nora
Rose

Fig. 3

*
***
100
50
0
−50
−100
Percentage variation in use
Attentive
Inattentive
Attentive
Inattentive
Attentive
Inattentive
Infant
Juvenile
Adolescent
Audible-or-contact gestures
Silent-visual gestures

Fig 4.

***
***
1.00
0.75
0.50
0.25
0.00
Percentage of gestures
Audible gestures
Contact gestures
Infant
Juvenile
Adolescent

Fig 5.

Unimodal adjustment
Cross-modal adjustment
Acquired
In process of acquisition
Non-acquired
ADULT PATTERN
Developmental axis
INFANT
JUVENILE
ADOLESCENT

Article accessible sur le lien suivant :
https://research-repository.st-andrews.ac.uk/bitstream/handle/10023/26170/Dafreville_2021_AJoP_Sensitivity_communicative_partner_AAM.pdf;jsessionid=98FC1B31AD8864723C3A9720B5EB0C2E?sequence=1

Article accessible sur le lien suivant :
https://research-repository.st-andrews.ac.uk/bitstream/handle/10023/26170/Dafreville_2021_AJoP_Sensitivity_communicative_partner_AAM.pdf;jsessionid=98FC1B31AD8864723C3A9720B5EB0C2E?sequence=1

Article accessible sur le lien suivant :
https://research-repository.st-andrews.ac.uk/bitstream/handle/10023/26170/Dafreville_2021_AJoP_Sensitivity_communicative_partner_AAM.pdf;jsessionid=98FC1B31AD8864723C3A9720B5EB0C2E?sequence=1

Article accessible sur le lien suivant :
https://research-repository.st-andrews.ac.uk/bitstream/handle/10023/26170/Dafreville_2021_AJoP_Sensitivity_communicative_partner_AAM.pdf;jsessionid=98FC1B31AD8864723C3A9720B5EB0C2E?sequence=1

Article accessible sur le lien suivant :
https://research-repository.st-andrews.ac.uk/bitstream/handle/10023/26170/Dafreville_2021_AJoP_Sensitivity_communicative_partner_AAM.pdf;jsessionid=98FC1B31AD8864723C3A9720B5EB0C2E?sequence=1

Article accessible sur le lien suivant :
https://research-repository.st-andrews.ac.uk/bitstream/handle/10023/26170/Dafreville_2021_AJoP_Sensitivity_communicative_partner_AAM.pdf;jsessionid=98FC1B31AD8864723C3A9720B5EB0C2E?sequence=1

Article accessible sur le lien suivant :
https://research-repository.st-andrews.ac.uk/bitstream/handle/10023/26170/Dafreville_2021_AJoP_Sensitivity_communicative_partner_AAM.pdf;jsessionid=98FC1B31AD8864723C3A9720B5EB0C2E?sequence=1

Article accessible sur le lien suivant :
https://research-repository.st-andrews.ac.uk/bitstream/handle/10023/26170/Dafreville_2021_AJoP_Sensitivity_communicative_partner_AAM.pdf;jsessionid=98FC1B31AD8864723C3A9720B5EB0C2E?sequence=1

Article accessible sur le lien suivant :
https://research-repository.st-andrews.ac.uk/bitstream/handle/10023/26170/Dafreville_2021_AJoP_Sensitivity_communicative_partner_AAM.pdf;jsessionid=98FC1B31AD8864723C3A9720B5EB0C2E?sequence=1

Article accessible sur le lien suivant :
https://research-repository.st-andrews.ac.uk/bitstream/handle/10023/26170/Dafreville_2021_AJoP_Sensitivity_communicative_partner_AAM.pdf;jsessionid=98FC1B31AD8864723C3A9720B5EB0C2E?sequence=1

Article accessible sur le lien suivant :
https://research-repository.st-andrews.ac.uk/bitstream/handle/10023/26170/Dafreville_2021_AJoP_Sensitivity_communicative_partner_AAM.pdf;jsessionid=98FC1B31AD8864723C3A9720B5EB0C2E?sequence=1

Article accessible sur le lien suivant :
https://research-repository.st-andrews.ac.uk/bitstream/handle/10023/26170/Dafreville_2021_AJoP_Sensitivity_communicative_partner_AAM.pdf;jsessionid=98FC1B31AD8864723C3A9720B5EB0C2E?sequence=1

Supplementary material

Sensitivity to the communicative partner's attentional state: a developmental study on mother-infant dyads in wild chimpanzees (*Pan troglodytes schweinfurthii*)

Mawa Dafreville[1]*, Catherine Hobaiter[2], Michèle Guidetti [1], David Sillam-Dussès[3], Marie Bourjade[1,4]

[1] CLLE, Université de Toulouse, CNRS, Toulouse, France
[2] Origins of Mind, School of Psychology and Neuroscience, University of St Andrews, St Andrews, Scotland
[3] Laboratoire d'Ethologie Expérimentale et Comparée UR 4443, Université Sorbonne Paris Nord, Villetaneuse, France
[4] Institut Universitaire de France

We replicated all our statistical analysis tests on a restricted sample of only single gesture and first gesture in the sequence in Table S1, S2 and this subset sample is detailed in Table S3. All the other tables and figures are related to the entire sample.

TABLE S1. Parameter estimates for the tested model – GLMM with the number of silent-visual gestures (n=76) mismatching maternal visual attention as dependent variable and age (in months) as fixed effect. There was a clear influence of age on unimodal matching (null model comparison: X^2=12.491, df=1, p<0.001).

Parameter	Unimodal matching ~ Age					
	Estimate	**SE**	**Lower CI**	**Upper CI**	**Z-value**	**P-value**
(Intercept)	1.072	0.383	0.321	1.823	(1)	(1)
Age(2)	-0.019	0.006	-0.031	-0.009	-3.548	p<0.001

(1) Not indicated because of limited interpretive value

(2) Age: refers to the individual age of the number of unimodal matching (see methods)

TABLE S2: Summary of p-values of the statistical tests on the subset of the sample that contains only gestures produced as either *single* gestures or as first in sequence gestures.

	Spearman rank order correlation	Fisher's exact tests'	
Tested modalities variation across states of maternal visual attention	**Correlation with age (N=11)**	***Age categories included in the statistical test***	**p-value without considering age effect**
1. Unimodal adjustment	r = +/-0.92, p <0.001	*Infant – Juvenile*	0.22
		Juvenile – Adolescent	<0.001
		Infant – Adolescent	<0.001
2. Cross-modal adjustment	Attention : r = +/-0.56, p = 0.173	*Infant*	0.9
		Juvenile	0.13
	Inattention: r = +/-0.78. p = 0.009	*Adolescent*	<0.001
3. Audible-or-contact gestures preference when the mother shows visual inattention	r = +/-0.052, p =0.819	*Infant – Juvenile*	1
		Juvenile – Adolescent	<0.01
		Infant – Adolescent	0.01

TABLE S3: Number of gestures produced by signaller by gesture modality, visual condition, and age category when considering only *single* gestures and first in sequence gestures. A = audible gestures; C = contact gestures; S = silent-visual gestures.

Condition and gesture modality / Age category and signaller	Inattentive mother				Attentive mother				
	A	**C**	**S**	**Total**	**A**	**C**	**S**	**Total**	**Total # gestures**
Infant	8	16	7	31	11	4	1	16	47
Karibu	4	12	5	21	6	1	1	8	29
Klauce	4	3	1	8	2	1	0	3	11
Kox	0	1	1	2	3	2	0	5	7
Juvenile	13	29	22	64	10	10	25	45	109
Karo	3	3	4	10	2	1	4	7	17
Kasigwa	0	3	5	8	0	1	5	6	14
Night	1	12	11	24	4	5	13	22	46
Zak	9	11	2	22	4	3	3	10	32
Adolescent	12	2	0	14	13	4	21	38	52
Kumi	3	0	0	3	4	0	3	7	10
Kwezi	2	0	0	2	2	2	6	10	12
Nora	0	2	0	2	6	1	7	14	16
Rose	7	0	0	7	1	1	5	7	14
Total # gestures	33	47	29	109	34	18	47	99	208

TABLE S4: Repertoire size, number and name of gestures produced by young chimpanzees by age category. *Grey part* indicates sum of produced gestures by age category and gesture modality.

		Infant	Juvenile	Adolescent	Total # gestures
Audible gestures		**30**	**30**	**34**	**94**
	Bigloudscratch	2	21	22	45
	Dangle	16	1		17
	Dangle:withshake	3	1		4
	Drum			1	1
	Hitobject/ground			2	2
	Leafclipping		1	1	2
	Locomote:gallop		1		1
	Objectmove	1	4		5
	Objectshake	1			1
	Stomp2-feetobject/ground	4			4
	Stompobject/ground	2			2
	Stomping2-feetobject/ground	1	1	8	10
Contact gestures		**37**	**42**	**6**	**85**
	Bite		2	1	3
	Grab	6			6
	Grab-pull	2			2
	Hitother	7			7
	Hitwithobject	1			1
	Hittingother	2	1		3
	Push	4	1		5
	Push(directed)	4	16		2
	Stomp2-feetother	3			3
	Stompother	2			2
	Stomping2-feetother	1			1
	Stroke:mouth		7	2	9
	Tapother		1		1

	Touch	2	13	1	16
	Touch:longother	3	1	2	6
Silent-visual gestures		**10**	**54**	**21**	**85**
	Beckon		1		1
	Headstand	1	1		2
	Look		1		1
	Objectinmouth	1			1
	Present(directed)	1	11	14	26
	Present:genitals		1		1
	Raisearm	1	10		11
	Reach	2	15	7	24
	Reach-palm		6		6
	Rollover		2		2
	Shakearm		1		1
	Spin:sideroulade		1		1
	Spin:somersault	2	1		3
	Swing	2	3		5
Total # gestures		**77**	**126**	**61**	**264**
Repertoire size		27	28	11	66

TABLE S5: Number of sequences and nature of them across visual condition by age category. A= audible gestures; C= contact gestures and S= silent-visual gestures; Modal dimension refers to the number of incorporated modalities within the gestures sequence: only one (unimodal), two of them (bimodal), all of them (trimodal).

Age category	Nature of the sequence	Modal dimension	Total # sequences	Total # sequences by age category
Infant (n=3)	AA	Unimodal	1	20
	AAA	Unimodal	1	
	AACA	Bimodal	1	
	AC	Bimodal	1	
	ACC	Bimodal	2	
	AS	Bimodal	1	
	AS-A	Bimodal	2	
	CAA	Bimodal	1	
	CAC	Bimodal	1	
	CC	Unimodal	5	
	CCC	Unimodal	2	
	SA	Bimodal	1	
	SC	Bimodal	1	
Juvenile (n=16)	AA	Unimodal	5	16
	AAA	Unimodal	1	
	AC	Bimodal	1	
	AS	Bimodal	3	
	CC	Unimodal	1	
	CS	Bimodal	3	
	SC	Bimodal	1	
	SS	Unimodal	1	
Adolescent		Unimodal		7
(n=7)	AA		3	
	AAAA	Unimodal	1	
	SA	Bimodal	3	
Total # gestures				43

Fig. S1. Distribution of silent-visual gestures according to the maternal visual attention for each individual (N=11, ***P<0.001**).** Large black circles represent mean proportion per subject. Median (horizontal lines), quartiles (boxes), percentiles (2.5% and 97.5%. vertical lines) and outliers (small black circles) are indicated. Differential use patterns are compared from one age category to another.

Fig. S2. Effect of age on the proportion use of silent-visual and audible-or-contact gestures according to the maternal visual attention for each individual (N=11). The effect of weighted mean age on the use of gestures type was significant when mother shows visual inattention (Spearman rank order correlation coefficient, S = 68.517, p-value = 0.039*, rho= +/-0.69) while it is not the case when she shows visual attention (Spearman rank order correlation coefficient, S = 328.25, p-value = 0.248, rho= +/-0.492).

Fig. S3. Number of silent-visual gestures and audible-or-contact gestures for each individual according to the maternal visual attention for each juvenile chimpanzee (N=4). The lines illustrate individual shifts in behaviour across the two discrete levels of the attention condition.

4.
COMMUNICATION SENSIBLE A L'ATTENTION MATERNELLE CHEZ L'ENFANT PREVERBAL (*HOMO SAPIENS*)

4.1. ARTICLE 2: ATTENTION-SENSITIVE SIGNALLING BY 7-TO-20-MONTH-OLD INFANTS IN A COMPARATIVE PERSPECTIVE

Mawa Dafreville[1*], Michèle Guidetti [1], Marie Bourjade[1,2]

[1] CLLE, Université de Toulouse, CNRS, Toulouse, France

[2] Institut Universitaire de France

Article soumis à Child Development en 2022

Cette étude a fait l'objet d'une communication affichée à l'International Congress of Infant Studies (ICIS) en juillet 2022 à Ottawa, Canada.

Synthèse de l'article 2

Contexte : Les hypothèses d'Evo-devo suggèrent que le contrôle vocal ait pu apparaître dans la lignée humaine comme un moyen d'attirer et de maintenir l'attention et l'investissement parental. Cette étude visait donc à déterminer si et comment les enfants préverbaux de 7 à 20 mois ajustaient la modalité sensorielle de perception de leurs signaux à l'attention maternelle (i.e. l'attention visuelle, le contact physique et le Langage Adressé à l'Enfant (LAE)).

Méthode : A partir d'observations en condition écologique de trente dyades mère-enfants humaines, ont d'abord été identifiés les principaux indices maternels qui influencent la signalisation de l'enfant. Puis, la *communication sensible à l'attention* a été testée à partir de descripteurs d'ajustement *unimodal* et *intermodal*, respectivement définis par la capacité de l'enfant à traiter les mères visuellement inattentives (i) en évitant de produire des signaux *silencieux-visuels* et/ou (ii) en favorisant la production de signaux perceptibles par l'utilisation de signaux *audibles* ou *tactiles*.

Résultats : Les enfants présentaient un ajustement *unimodal* avant l'ajustement *intermodal* à partir de respectivement, 8 et 11 mois. Le LAE influençait de manière concurrente ces deux ajustements ; l'ajustement *unimodal* était plus marqué lorsque la mère s'adressait à l'enfant, tandis que c'était l'inverse pour l'ajustement *intermodal*. De plus, pendant l'ajustement intermodal, les enfants préféraient les signaux oraux aux gestes audibles seulement lorsque la mère leur parlait.

Conclusion : Concernant l'ajustement *unimodal*, la présence de LAE inciterait les enfants à inhiber leurs gestes silencieux-visuels lorsque la mère est visuellement inattentive. Concernant l'ajustement *intermodal*, l'absence de LAE inciterait les enfants à privilégier les moyens de communication audibles pour regagner l'attention de leur mère mais la présence de LAE les inciterait à le faire avec des signaux oraux plus que gestuels.

Attention-sensitive signalling by 7-to-20-month-old infants in a comparative perspective

Mawa Dafreville[1*], Michèle Guidetti [1], Marie Bourjade[1,2]

[1] CLLE, Université de Toulouse, CNRS, Toulouse, France

[2] Institut Universitaire de France

*Corresponding author: mawa.dafreville@univ-tlse2.fr

Abstract:

Evo-devo hypotheses suggest that vocal control may have appeared in the human lineage as a means to attract and maintain parental attention and investment. This study aimed at exploring whether and how 7-to-20-month-old infants adjust the sensory modality of their communicative signals to their mother's attentional cues. Three dimensions of maternal attention were examined: visual attention, infant-directed speech (IDS) and physical contact. We studied 30 mother-infant dyads in naturalistic settings, sampled according to three developmental periods, [7-10], [11-14] and [15-20], in order to document developmental effects. Infants' attention-sensitive signalling was tested through *unimodal* and *cross-modal* adjustment, respectively defined by the capacity of infants to address visually inattentive mothers (i) by avoiding visual communication mismatches and/or (ii) favouring communication matches through audible-or-contact signals. The infants displayed *unimodal* adjustment before *cross-modal* adjustment from respectively 8 and 11 months of age. Maternal *Infant-directed speech* (IDS) affected both descriptors in concurrent directions; *unimodal* adjustment was steeper and appeared earlier when the mother was talking to the infant while *cross-modal* adjustment was steeper when the mothers were silent. Furthermore, during

cross-modal adjustment, infants preferred oral signals over audible gestures only when the mother was talking to them. The presence of IDS may prompt infants to inhibit their silent-visual gestures when the mother is visually inattentive while the absence of IDS may prompt infants to favour audible communication means to regain their mother's attention. These results are discussed in line with evo-devo hypotheses and comparative studies on the evolution of communication within the primate lineage.

Keywords: sensory modality, infant-directed speech, visual attention, mother–infant dyads, development.

Introduction

Evolutionary developmental psychology bridges the gap between developmental and evolutionary processes on the strength of the adaptiveness of developmental stages (Bjorklund & Pellegrini, 2000; Frankenhuis & Tiokhin, 2018). In regards to language acquisition, infancy and childhood are crucial stages to develop communicative behavior and related socio-cognitive skills that can be compared to the communicative behavior of other *non-talkers* such as non-human primates (for reviews see Arbib et al., 2008; Fröhlich et al., 2019; Tomasello, 2008). One missing link that is common to both language *evolution* and *acquisition* is the reorganization of the visual and audible modalities from separate communication systems into an integrated multimodal stream including the finely-tuned control of the vocal cords. Evo-devo hypotheses suggest that vocal control may have appeared in the human lineage as a means to attract and maintain parental attention and investment (Falk, 2004; Locke & Bogin, 2006; Mehr & Krasnow, 2017). Although infant communication has been extensively studied, the question of whether infants are sensitive to their caregiver's attention when selecting their communication modalities has been somewhat overlooked. The present study is framed within this evo-devo perspective and uses an ethological coding to address

attention-sensitive signalling in 7-to-20-month-old infants communicating with their mother.

Some authors suggested that parental attention belongs to the selective forces potentially involved in the emergence of proto-linguistic vocal substrates (Falk, 2004; Locke, 2006, 2017; Mehr & Krasnow, 2017). Specifically, following the acquisition of the bipedal posture, increasing physical distance between mothers and infants may have prompted mothers to vocalize (Falk, 2004) or sing (Mehr & Krasnow, 2017) to soothe their dependent offspring. Another hypothesis (Locke, 2006, 2017) suggests that the selection process may have favoured infants producing sophisticated and tactical vocalizations. Thus, babblings may progressively have constituted an indicator of the infant's good health and may have encouraged more parental investment (Locke, 2006, 2017; Locke & Bogin, 2006; Oller et al., 2016). According to these theoretical accounts, the prominence of audible, vocal communication in the human lineage may have emerged within caregiver-infant communication and attentional exchanges.

Interestingly, non-human primates also communicate using distinct sensory modalities, as has been shown for the great apes (Byrne et al., 2017; Genty et al., 2009; Hobaiter & Byrne, 2011) and monkeys (Gupta & Sinha, 2019; Molesti et al., 2020). The study of non-human primate communication is based on repertoires of signals classified by the dominant sensory modality of perception, stemming from *audible*, *silent-visual* and *contact* signals (Fröhlich & van Schaik, 2020). Most of these signals are visible and show a visual component, but when they additionally make sound or involve physical contact with the recipient, they are respectively considered as *audible* and *contact* signals. In addition, the sensory channel conveying information is also relevant for studying communication and language in an evolutionary perspective (Frölich & Van Schaik 2018). Non-voiced oral signals have for example been considered as potential precursors of speech (Meguerditchian et al 2014). To our knowledge, no study on human infants has

investigated the contexts of use of these non-voiced oral signals (see Oller, 2000). The use of an ethological coding will fill an important gap and complement the few human studies that have recently undertaken comparisons with non-human primate gestural communication (Kersken et al., 2019; Rodrigues et al., 2021).

Many studies, however, have investigated the acquisition of specific communication signals in preverbal infants, showing that the sensory modalities used to communicate had differential developmental trajectories. In terms of silent-visual signals, 5-to-9-month-old infants may use *body movements* to delineate early forms of interaction with an adult (Scola et al., 2015). In addition, many studies investigated the communicative function of shared reference by considering *deictic gestures* (i.e. *giving*, *showing* the so-called *index-finger pointing* gestures) around 12 months of age (Liszkowski et al., 2006). In terms of audible signals, most studies investigated the use of vocalisations, emphasizing the importance of non-speech vocalisations in the development of speech (Oller, 2000). They include proper vocalisations that activate the vocal cords (e.g., *babbling and vowel-like sounds)* but also some non-voiced sounds made without activation of the vocal cords (Oller, 2000).On the other hand, only a few studies have investigated the use of audible gestures in 7-to-12-month-old infants (Murillo et al., 2021; Rodrigues et al., 2021) but the use of the audible aspect of the gesture was not always the purpose of the study (Murillo et al., 2021).

A few studies examined the use of pointing gestures and vocalisations by prelinguistic children according to the adult recipient's attention in declarative pointing tasks. The authors addressed the question of whether infants were able to adjust their communicative means to the attention of the audience. This sensitivity to the attentional state of recipient is considered a hallmark of intentionality and stems from a perceptive constraint whereby silent-visual signals are not noticeable when the recipients do not show visual attention (Townsend et al., 2017). Wu and Gros-Louis (2017) showed that 10-month-old infants who produced more

vocalisations (not combined with pointing gestures) when their mother did not show visual attention than when she did, exhibited higher language skills at 15 months of age. Liszkowski et al. (2008) showed that 12-month-old infants decreased the frequency of pointing gestures towards inattentive adults as compared to attentive ones. By comparing infants of 12 and 18 months of age, the authors found that only 18-month-old infants produced more vocalisations when the adults were not visually attentive and not responsive than when they were visually attentive and responsive (Liszkowski et al., 2008). Igualada et al. (2015) found that 12-month-old infants used more pointing–vocal coupling when the adults showed visual attention than when they did not while some non-experimental studies found the opposite pattern, with more pointing–vocal coupling when the adults did not show visual attention than when they did in 10-to-13-month-old infants (Gros-Louis & Wu, 2012; Wu & Gros-Louis, 2014). Although these studies reported mixed results, they pointed out the link between vocal communication and parental attention put forward by *evo-devo* hypotheses on language evolution.

These results have recently been deepened by comparative observational studies conducted on both human infants and immature chimpanzees. Rodrigues et al. (2021) investigated the adjustment of audible, silent-visual and contact gestures by 7-to-12-month-old infants to their recipient's visual attention, in an ecological context of free play. Infants used more audible and contact gestures than silent-visual gestures when their recipients did not show visual attention and conversely when they did (Rodrigues et al., 2021). However, this study focused only on gestures and did not consider vocalisations. Using a similar coding, some authors recently investigated the development of attention-sensitive gestural communication in wild chimpanzees (Dafreville et al., 2021). Using two descriptors of attention-sensitive communication, they found that immature chimpanzees preferentially used silent-visual gestures in front of a visually attentive mother, i.e., *unimodal adjustment*, and switched communication modalities when her visual

attention was unavailable; juveniles preferred using contact gestures and adolescents preferred using audible gestures, i.e., *cross-modal adjustment* (Dafreville et al., 2021). *Unimodal adjustment* hence refers to the capacity of the signaller to deploy fewer silent-visual signals when the recipient does not show visual attention as compared to when she does, i.e. the capacity for avoiding communication mismatch (Dafreville et al., 2021). *Cross-modal adjustment* refers to the capacity of the signaller to not only inhibit the production of silent-visual signals, but also to deploy more audible-or-contact signals that could be perceived when the recipient does not show visual attention as compared to when she does, i.e., the capacity for switching modalities to favour communication match (Dafreville et al., 2021).

The present study seeks to expand these collective findings by conducting the first systematic investigation of the sensory modalities used by 7-to-20-month-old infants to communicate with their mother in naturalistic conditions. More precisely, the aim was to quantify the infants' attention-sensitive signalling across spontaneous 'conditions' of maternal attention, operationalized as *visual attention* (attentive/inattentive), *infant-directed speech* (talking/silent) and *physical contact* (yes/no) from video footages filmed in the home. We first examined how infants used signal modalities across maternal contexts in order to identify salient maternal cues. Then, we tested infants' capability for *unimodal adjustment* by investigating the distribution of visual signals between the conditions where the mother's visual attention was available and unavailable. Thirdly, we tested infants' capability for *cross-modal adjustment* observed when infants used more audible-or-contact than visual signals when the mother was inattentive compared to when she was attentive. We compared the use of contact and audible signals in the case of maternal inattention to further document the specific usage of vocal communication in the human lineage. Finally, we examined the use of multimodal combinations of signals from at least two different modalities across maternal contexts. Each analysis was conducted on three age categories, [7-10], [11-14] and [15-20], in order to document developmental effects.

Method

Participants

This study stems from data collected from March to July 2019 by MD during fieldwork in the city of Toulouse (France) and its surroundings. Mother-infant dyads came mainly from urban areas. Participants were recruited via the BabyLab platform of the University Toulouse Jean Jaurès and via the distribution of flyers and emails to a network of nurseries of the Community Centre for Social Action (CCAS). A pre-home visit with the purpose of mutual familiarization allowed the observer to briefly inform the mother about the study and to ask her a few questions about her infant (family configuration, type of childcare, etc.) using a short demographic questionnaire. Mostly, they were families with first-born infants (n=18 out of 30 mother-infant dyads, 60%) and the average number of siblings per infant observed was 1.87. The mother's age varied between 27 and 40 years. The mother had a high level of formal education, as they all had graduate degrees (at least bachelor level). Most of the infants were totally breastfed from birth to 6 months of age (n=24 out of 30 mother-infant dyads, 80%). Nine individuals also experienced health difficulties at their birth (i.e. preterm; difficult childbirth; asthma).

In this cross-sectional study, thirty mother-infant dyads were allocated to three infant age categories spanning roughly *before*, *during* and *after* the emergence of intentional use of pointing gestures at the mean age of 11-12 months (Camaioni et al., 2004; Carpenter et al., 1998; Colonnesi et al., 2010). The three age groups defined were**:** 7 to 10 months of age, 11 to 14 months of age and 15 to 20 months of age (see table 1 for further details).

Table 1. Number of human infants included in the study (mean age ± SD) by sex and age category.

Age category (in months)	**[7-10]**	**[11-14]**	**[15-20]**	**Total**
Females	**6** (8.83 ± 1.03)	**3** (12.78 ± 0.99)	**2** (18.53± 1.59)	**11** (11.67 ± 3.95)
Males	**5** (8.80 ± 1.09)	**9** (13.11 ± 0.91)	**5** (16.41 ± 0.93)	**19** (12.84 ± 2.99)
Total	**11** (8.81 ± 1.00)	**12** (13.03 ± 0.90)	**7** (17.01 ± 1.44)	**30**(12.41 ± 3.36)

Procedure and design

Dyads were observed in naturalistic conditions at home during the context of meals (lunch, dinner, afternoon snack time). As visits aimed at accessing the most natural conditions, the observations took place whether other members of the family were present or not. During the observations, MD stayed as much as possible on the side-lines and was not engaged in any interaction with the mother or the infant during the observation. Each dyad was observed twice for on average 1 hour, one month apart. The recording started on average 15 minutes before the start of the meal (taking care to ensure that the infant was awake). Both visits were made so that the infant remained in the same age group during the two observations. The visits were filmed using two cameras: a camera (PANASONIC HC-V770) mounted on a tripod, providing an overview of the room and the movements of the mother in particular; a second camera (HANDYCAM HDR-CX625) hand-held by the observer, who focused only on the infant and followed him/her. All the mother-infant dyads were kept for further analysis, even if there was only one observation per infant (three infants were not observed twice).

Ethical procedure

During the first visit, the parents of infant participants completed a consent form on behalf of their infants and were questioned about their profession, the sleep rhythm and natural eating habits of their infant. During the last visit, the experimenter added an amendment to the initial signed consent in which the mother authorised the researchers to study her behaviour in addition to that of her

infant. The video and coding data were anonymised and stocked on external drives stored in safe places locked until 15 years after the end of the study, at which time they will be destroyed. The research design was approved by the ethical committee of the University of Toulouse (i.e., CERNI; n° IRB00011835-2019-03-19-141).

Data collection

Quantitative behavioural data were collected on videos using an individual focal sampling approach on the infant (Altmann, 1974). One infant was observed at a time and the infant's communicative signals were recorded continuously over 3 focuses of 5 min where each focus was randomly selected before, during and after feeding per visit (1 and 2). When it was impossible to have one focus before or after the meal because the observation did not cover these three times, we replaced the missed meal time (for example 'after meal') by a second one from the same meal time ('before meal' in our example).

Behavioural sampling and coding procedure

The infant's communicative signals were coded in ELAN software (Version 5.2) following the repertoire in *Supplementary material* (Table S1). The repertoire included about 23 behaviours classified into three categories based on their main sensory modalities, i.e. visual, audible and tactile. In the repertoire, we considered that *audible signals* encompassed oral and gestural signals, while *silent-visual* and *tactile* signals were exclusively gestural (see Table 2 and Table S1 for further details).

Table 2. Brief categorization of the communicative signals according to the signal modality and channel of production. The detailed repertoire of behaviours is available in the *Supplementary material* (see Table S1 in *Supplementary material*).

Perception modality	Production channel	Sub-groups of communicative signals
Silent-visual	Gestural	• Body movements (circular movement of the limbs; synchronized trunk/members movement; swagger) • Deictic gestures (whole-hand pointing gesture; index-pointing; reaching attempt; giving; receiving; showing) • Conventional gestures • Facial expressions (smiling)
Tactile	Gestural	• Gestures involving the infant touching his/her mother with at least one body part except with the feet)
Audible	Oral	• Non-voiced sounds (mouth sounds and laughs [1]) • Voiced sounds (vowel-like sounds; grunts; babbling; cries; pseudo-words; words; succession of words)
	Gestural	• Audible gesture (sound made by manual or bodily actions with or without an object)

[1] *laughs were included in non-voiced sounds because they were defined as being smiles with vocal or non-vocal sounds lasting more than 1 second, interspersed by silences during which the infant breathes in and so they are not necessarily produced with vocal cords.*

For each signal, we recorded the following information: (1) the perception sensory modality (audible, tactile or silent-visual), (2) the production channel (gestural or oral), (3) if the signal was combined with another signal of another sensory modality, (4) the presence of another recipient closer than the mother (i.e., non-maternal recipient), (5) the attention state of the mother and of the non-maternal recipient if applicable. To define the recipient attention state, we recorded the attentional variables below (for more details, see Table S2 in *Supplementary material*): *1) Visual attention* was recorded when the mother had her head directed at her infant from narrow to large angles, the geometric lines of sight of the two

individuals being able to cross each other; *2) Infant Directed Speech* was recorded when the mother was talking to her infant; *3) Physical contact* was recorded when the mother had at least one body part in physical contact with her infant. *Auditory attention*, i.e., the possibility that the mother can hear her infant, was considered always present as we could not determine its absence.

We also accounted for the various multimodal combinations defined as the combination of signals of different sensory modalities. Among these multimodal combinations, we distinguished between *distal combinations* (combination of audible and silent-visual signals without contact gestures) and *proximal combinations* (combination of at least a contact gesture and at least one other signal modality). A combination was recorded at each overlap between one signal and another one except if the time between the end of the first signal (over time) and the beginning of the second one was longer than 200 milliseconds. Because we focused on maternal attentional state in this study, all the infant signals potentially directed to a non-maternal recipient were removed.

Inter-rater reliability

The videos were entirely coded by one main coder (MD), who was therefore used as the reference for training a second blind coder. The second coder (NP) trained for one month, then blindly coded 20% of the dataset. Reliability was assessed through Cohen's Kappa statistics. The second coder reached a Cohen's Kappa coefficient of almost 80% with the first coder for the infants' visual communicative behaviours; Cohen's Kappa ranged from 0.72 to 0.83 across the infants' audible signals (Table S3 in *Supplementary material*). Agreement for maternal attention was high (Cohen's k > 0.80 Table S4 in *Supplementary material*).

Statistical analysis

Generalized Linear Mixed Models on proportion data (GLMM; Bolker et al., 2009; Harrison, 2015) were conducted with a binomial error distribution and the *logit link function* to determine which factors most affected the use of signal

modality, i.e., audible, silent-visual or contact signals. The dependent variable was the proportion of signals in a given modality for each infant. Fixed effects included *mean age* of each infant in months (range = 7-20), maternal visual attention state (two levels: attentive, inattentive), maternal IDS (two levels: silent mother, talking mother) and maternal physical contact (two levels: yes, no) as main fixed effects (key test predictors). To account for variation in the dependent variable caused by confounding factors, we ran the models with these factors in random effects (more details are given in the *Supplementary material* and in Table 3; Theobald, 2018). A GLMM using a Poisson error distribution and a log link function was also conducted to test the effect of increasing age on the capacity for *unimodal adjustment* (as in Dafreville et al., 2021). Thus, the dependent variable was the sum of mismatches (i.e. a silent- visual signal produced to a visually inattentive mother) for each infant. As fixed effects we included IDS, *mean age* (in months; Van de Pol & Wright, 2009) and the interaction between them. To achieve these GLMM analyses, the models were implemented using the *glmer* function of the R package *lme4* version 1.1- 23 with the optimizer function of *aictab* of the R package *AICcmodavg* to calculate AICc, which is a transformation of the Akaike Information Criterion (AIC) that is used for small and medium samples (more details are given in the *Supplementary material*). Because the sample size was medium (30 mother-infant dyads) and the data did not follow a normal distribution, we used non-parametric statistics to test the hypotheses (Siegal & Castellan, 1988), including Bonferroni corrections in the case of multiple comparisons. All the non-parametric tests detailed below were conducted twice, on data (i) with maternal IDS ('talking mother') and, (ii) without it ('silent mother'). Permutation tests and Fisher's exact tests on paired data (Pearce, 2016) were performed to test our hypotheses (see the *Supplementary material for further details).* All the tests were conducted using R v3.6.1 software (http://cran.r-project.org) with p-values equal to or lower than 0.05 required for significance. All statistical tests were two- tailed.

Results

Overall, we coded a total of 9,245 communicative signals. After removing all the signals potentially oriented towards non-maternal recipients, we were able to reliably analyse 8,367 signals: 192 tactile signals, 1,422 visual signals and 6,753 audible signals of which 4,658 were oral signals.

- **Proportions of signal modalities by maternal visual attention, IDS and physical contact**

GLMM analysis aimed at identifying which maternal attention cues, among maternal visual attention, IDS and physical contact, influenced the proportion of use of each signal modality most. The proportion of silent-visual signals was affected by maternal visual attention, IDS and mean age as evidenced by the unique best fitting model (AICc =2332.26; Chi-square test of the log-likelihood ratio with null model, $p < 0.001$; wi = 1; see Table S5). The proportion of audible signals was also affected by maternal visual attention, IDS and mean age as evidenced by the unique best fitting model (AICc =2461.90; Chi-square test of the log-likelihood ratio with null model, $p < 0.001$; $wi = 0.81$; see Table S5). The proportion of contact gestures was affected by maternal physical contact, visual attention and IDS as evidenced by the best fitting models. Indeed, one main factor model and two interactive models fitted equally well; infants showed differential tactile signaling according to maternal physical contact (AICc =731.46; Chi-square test of the log-likelihood ratio with null model, $p < 0.001$; $wi = 0.40$; see Table S5), and the same result was found with the additional effect of maternal visual attention (AICc =731.96; Chi-square test of the log-likelihood ratio with null model, $p < 0.001$; $wi = 0.31$; see Table S5) or IDS (AICc =733.37; Chi-square test of the log-likelihood ratio with null model, $p < 0.001$; $wi = 0.15$; see Table S5). Regardless of the sensory modality of the infant's signal, maternal visual attention and IDS were

the most likely determining factors. These factors were therefore kept for further analysis

- **Unimodal adjustment: Does the use of silent-visual gestures vary according to maternal visual attention and IDS?**

The results of this GLMM analysis showed a significant effect of mean-age and maternal IDS on unimodal mismatches (see Table S6 in *Supplementary material*; AICc = 2775.35; Chi-square test of the log-likelihood ratio with null model, $\chi 2$ = 86.338, df = 3, $p < 0.001$; *wi* = 1). The production of unimodal mismatches decreased with increasing age and IDS (see Table S7 in *Supplementary material*). With a talking mother (IDS), infants produced fewer silent-visual signals when the mothers did not show visual attention than when they did, regardless of age group ([7-10]: Permutation Test, p =0.006, N = 11; [11-14]: Permutation Test, p =0.001, N = 12; [15-20]: Permutation Test, p =0.047, N = 7; see Figure 1). With a silent mother (No-IDS), this same pattern was found only from [11-14] months of age ([7-10]: Permutation Test, p =0.076, N = 11; [11-14]: Permutation Test, p =0.003, N =12; [15-20]: Permutation Test, p =0.047, N = 7; see Figure 1).

Figure 1. Distribution of silent-visual gestures produced by maternal condition and age range (*: p-value<0.05; **: p-value<0.01)**.** Large black circles represent the mean proportion per subject. Median (horizontal lines), quartiles (boxes), percentiles (2.5% and 97.5%), vertical lines, and outliers (small black circles) are indicated. Differential use patterns are compared from one age category to another. *Silent mother* refers to the condition without IDS; *Talking mother* refers to the condition with IDS.

In addition, unimodal adjustment was significantly better when the mother was talking than when the mother was silent in the [7-10] and [11-14] months of age groups ([7-10]: Fisher Test, OR = 2.452, p <0.001, N = 11; [11-14]: Fisher Test, OR = 3.154, p <0.001, N = 12; [15-20]: Fisher Test, OR = 1.092, p =1, N = 7, Figure S2 in *Supplementary material*).

We suspected that infants aged [7-10] months may constitute a heterogeneous group regarding unimodal adjustment. We therefore looked separately at infants younger than 8 months and those older than 8 months. Significantly fewer silent-visual signals were produced when the mothers did not

show visual attention than when they did only by infants older than 8 months (before 8: Permutation Test, p =1, N = 4; from 8 to 10: Permutation Test, p =0.031, N = 5; see Figure 2).

Figure 2. Distribution of silent-visual gestures produced by individuals before and after 8 months of age in front of silent mothers (*: p-value<0.05). Large black circles represent the mean proportion per subject. Median (horizontal lines), quartiles (boxes), percentiles (2.5% and 97.5%), vertical lines, and outliers (small black circles) are indicated. *Note: Data from the first observation only are represented. Infants not exposed to attentive and inattentive mother during the first observation were removed from the analysis.*

- **Cross-modal adjustment: how do signal modalities vary according to maternal visual attention and IDS?**

Only infants of [11-14] and [15-20] months of age showed capacities of cross-modal adjustment by producing significantly more audible-or-contact signals than silent-visual signals when the mothers did not show visual attention as

compared to when they did whether the mother was talking ([7-10]: Fisher Test, OR = 0.955, p =1, N =11; [11-14]: Fisher Test, OR = 0.543, p =0.007, N =12; [15-20]: Fisher Test, OR = 0.554, p =0.036, N =7; see Figure 3) or silent ([7-10]: Fisher Test, OR = 0.786, p =0.348, N =11; [11-14]: Fisher Test, OR = 0.524, p<0.001, N =12; [15-20]: Fisher Test, OR = 0.168, p<0.001, N =7; see Figure 3).

Figure 3. Proportion of audible-or-contact signals and silent-visual gestures produced by maternal visual attention condition and age range when the mothers were silent (*: p-value<0.05; **: p-value<0.01; see Figure S1 in *Supplementary material* for the figure when the mothers were talking). Large black circles represent the mean proportion per subject. Median (horizontal lines), quartiles (boxes), percentiles (2.5% and 97.5%), vertical lines, and outliers (small black circles) are indicated. Cross-modal adjustment occurs when the mean proportion difference between the proportions of audible-or-contact and silent-visual signals is greater when the mother is not visually attentive than when she is.

In addition, cross-modal adjustment was better when the mother was silent than when she was talking, regardless of the age group ([7-10]: Fisher Test, OR = 0.454, p <0.001, N = 11; [11-14]: Fisher Test, OR = 0.285, p <0.001, N = 12; [15-20]: Fisher Test, OR = 0.084, p <0.001, N = 7).

- **How are audible and contact gestures used when maternal visual attention is not available?**

Infants produced more audible than contact signals when the mothers did not show visual attention regardless of their age and whether they faced a talking mother ([7-10]: Permutation Test, p =0.006, N = 11; [11-14]: Permutation Test, p =0.001, N = 12; [15-20]: Permutation Test, p =0.047, N = 7) or a silent mother ([7-10]: Permutation Test, p =0.006, N = 11; [11-14]: Permutation Test, p =0.001, N = 12; [15-20]: Permutation Test, p =0.047, N = 7).

Infants facing a talking mother produced more oral than gestural audible signals when the mothers did not show visual attention; there was no age effect ([7-10]: Permutation Test, p =0.035, N = 11; [11-14]: Permutation Test, p =0.006, N = 12; [15-20]: Permutation Test, p =0.047, N = 7; see Figure 4). This pattern did not reach significance for infants facing a silent mother (No-IDS) ([7-10]: Permutation Test, p =0.469, N = 11; [11-14]: Permutation Test, p =0.195, N = 12; [15-20]: Permutation Test, p =1, N = 7; see Figure 4).

Figure 4. Proportions of audible gestures and oral signals produced by individual and age range when the mothers did not show visual attention (*: p-value<0.05; **: p-value<0.01). Large black circles represent the mean proportion per subject. Median (horizontal lines), quartiles (boxes), percentiles (2.5% and 97.5%), vertical lines, and outliers (small black circles) are indicated. Differential use patterns are compared from one age category to another. *Silent mother* refers to the condition without IDS; *Talking mother* refers to the condition with IDS.

- **How are multimodal combinations used according to maternal visual attention and IDS?**

Within the 987 multimodal combinations coded, regardless of the maternal condition and age range, infants produced more *distal* than *proximal* combinations (distal combinations *n=883*, 89,46%; for more details, see Table 4).

Table 4. Amount of multimodal combinations by age range and maternal attention condition

Maternal visual attention		**Attentive**				**Inattentive**			
Main category		**Distal**	**Proximal**			**Distal**	**Proximal**		
Combined modalities		**AV**	**AC**	**AVC**	**VC**	**AV**	**AC**	**AVC**	**VC**
Talking mother	[7-10]	96	11	1	6	20	0	1	0
	[11-14]	155	12	2	7	18	4	1	0
	[15-20]	164	4	2	5	18	1	0	0
	Subtotal	*415*	*27*	*5*	*18*	*56*	*5*	*2*	*0*
	Global total	***415***		***50***		***56***		***7***	
Silent mother	[7-10]	111	13	1	2	47	2	0	0
	[11-14]	120	4	1	1	38	1	0	0
	[15-20]	76	10	5	5	20	2	0	0
	Subtotal	*307*	*27*	*7*	*8*	*105*	*5*	*0*	*0*
	Global total	***307***		***42***		***105***		***5***	
	Overall total	***722***		***92***		***161***		***12***	

Note: the different categories of multimodal combinations are mutually exclusive. 'Combined modalities' indicates the different sensory modalities included in the signal combination: 'A' refers to one or more audible signals, 'V' refers to one or more silent-visual gestures and 'C' refers to one or more tactile gestures.

Infants produced significantly more *distal combinations* when the mothers showed visual attention than when they did not, whether the mother was talking (Permutation test, $p < 0.001$; see Figure 5) or silent (Permutation test, $p < 0.001$; see Figure 5). We did not find any age effect on this pattern (talking mother: Fisher Test, $p = 0.136$, $N = 30$; silent mother: Fisher Test, $p = 0.262$, $N = 30$). Infants produced *proximal combinations* independently from maternal visual condition and IDS (talking mother: Permutation test, $p=1$, $N=30$; silent mother: Permutation test, $p=1$, $N=30$; see Figure 5). We did not find any age effect on this pattern (talking mother: Fisher Test, $p = 0.451$, $N = 30$; silent mother: Fisher Test, $p = 1$, $N = 30$). In addition, maternal IDS did not affect the *proximal combinations* (Fisher

Test, OR = 0.856, p =1, N = 30) produced by infants. However, we found a non-significant trend for the *distal combinations* (Fisher Test, OR = 2.439, p =0.061, N = 30; Figure 5).

Figure 5. Distribution of *distal* and *proximal* multimodal combinations produced by individual and maternal visual condition (*: p-value<0.05; **: p-value<0.01; ***: p-value<0.001)**.** Large black circles represent the mean proportion per subject. Median (horizontal lines), quartiles (boxes), percentiles (2.5% and 97.5%), vertical lines, and outliers (small black circles) are indicated. *Silent mother* refers to the condition without IDS; *Talking mother* refers to the condition with IDS.

Discussion

This study aimed at exploring whether and how 7-to-20-month-old infants adjust the sensory modality of their communicative signals to the attentional cues of their mother. Three dimensions of maternal attention were examined: visual attention, infant-directed speech (IDS) and physical contact. The results showed that maternal visual attention and IDS were the most influential factors on infants' selection of signal modality. Moreover, infants displayed *unimodal adjustment* from 8 months onwards before being able to deploy audible-or-contact signals preferentially when maternal visual attention was unavailable from 11 months

onwards, hence showing *cross-modal* adjustment. When mothers did not show visual attention, infants of all ages preferred audible over contact signals to address their mother. Within audible signals, they preferentially selected oral signals such as vocalizations or mouth sounds over audible gestures only when the mother was talking to them. IDS also seemed to mediate the ease of unimodal adjustment as the mismatching of silent-visual signals towards visually inattentive mothers decreased sharply when the mothers were talking. IDS affected cross-modal adjustment in different ways. Infants showed better performances in selecting audible-or-contact over visual signals when the mother was neither visually attentive, nor talking. However, they favoured oral signals above all others when she was visually inattentive and talking. Finally, infants did not use multimodal combinations for cross-modal adjustment to maternal visual inattention. Instead, they produced most combinations in front of a visually attentive mother, whether she was talking or not.

In our study, infants showed *unimodal* adjustment from 7 or 8 months of age, respectively with a talking or a silent mother. Infants were actually able to adjust their silent-visual signals to their mother's visual attention from early on, and this is an early onset compared to other studies. Indeed, previous studies examining the use by 10-to-18-month-old infants of pointing gestures and vocalizations according to the adult's visual attention did not show *unimodal* adjustment before 12 months of age (Igualada et al., 2015; Liszkowski et al., 2008; Wu & Gros-Louis, 2017). However, these studies were conducted in laboratories and focused only on pointing gestures whereas the present study used an ethological approach relying on a repertoire of silent-visual gestures extended to other communicative signals (e.g. *body movements* or *smiles*). These latter communicative signals may come first and may be adjusted to maternal attention before the intentional use of pointing gestures at 12 months of age (Liszkowski et al., 2004; Rodríguez et al., 2015). This suggests that not all silent-visual signals become adjusted to the recipient's attention at the same time during development.

We found a developmental effect on *cross-modal* adjustment. From 11-14 months of age, infants produced more audible-or-contact signals than silent-visual ones when their mothers were visually inattentive as compared to when they were attentive. Although previous studies conducted their investigations with a smaller repertoire than ours, our result is globally consistent with this previous work on infants' coupling of pointing gestures and vocalisations in declarative pointing tasks (Gros-Louis & Wu, 2012; Wu & Gros-Louis, 2014, 2017). Interpreting their results in the light of our operationalization of *cross-modal* adjustment, it can be said that infants did not show *cross-modal* adjustment on average before 12 months of age (Gros-Louis & Wu, 2012; Wu & Gros-Louis, 2014, 2017). However, a recent study with an ethological coding comparable to ours showed an early *cross-modal* adjustment with contact gestures from 7 months of age (Rodrigues et al., 2021). This study was conducted in a free-play context of interaction in a nursery in which infants interacted with both other infants and with caregivers (Rodrigues et al., 2021). Free-play may favour the use of contact gestures and explain why the authors obtained a different result from ours. In our sample, infants were observed in a context around meal times during which they were usually settled in a chair or baby bouncer during the feeding and left 'on tour' on the floor while the mothers went about their daily activities. Our observation context was therefore more favourable to distal communication (infants' use of more audible and visual signals). This methodological choice may have affected the onset of cross-modal adjustment in the present study and further investigations are needed to clarify if tactile communication is preferred by younger infants over audible communication to address an inattentive mother.

We found that infants of all ages combined audible and visual signals more often when the mother was visually attentive than when she was inattentive. We quantified distal and proximal combinations of signals to examine the extent to which this precursor of language might be better adjusted to maternal attention than isolated signals. Contrary to our expectations related to maternal inattention,

infants did not combine signals of different modalities to regain their mother's attention. The present study did not discriminate *between* signals within each modality. Further studies should examine if certain combinations are preferentially used in specific attentional contexts and compare qualitatively the levels of sophistication of these audio-visual combinations across conditions. For instance, we may consider a coupling of '*pointing gestures* and *babblings*' more sophisticated than a coupling of '*body movements* and *mouth sounds*' because babbling involves *consonant vowel* vocalisations (i.e., consisting of at least one syllable containing at least one consonant and vowel; D'odorico & Franco, 1991; Grünloh & Liszkowski, 2015) whereas *mouth sounds* do not. Furthermore, our results may also be interpreted in terms of maternal *scaffolding* in communication development (Bernicot et al., 1993; Bernicot & Marcos, 1993; Bruner, 1975); full maternal attention may stimulate infants to display more complex communication by the use of multimodal combinations (Aureli et al., 2017). Our results are partially consistent with the findings of previous studies on infants' coupling of pointing gestures and vocalisations in experimental declarative pointing tasks (Igualada et al., 2015; Liszkowski et al., 2008; Wu & Gros-Louis, 2017). One of these studies found the same pattern in 12-month-old infants as our own (Igualada et al., 2015) while the others did not find any significant differences in the use of pointing-vocal couplings at 10 and 12 months of age across conditions (Liszkowski et al., 2008; Wu & Gros-Louis, 2017). Again, these discrepancies can be explained by the inclusion, in the present study, of communicative signals beyond pointing and vocalization.

We observed mediating effects of IDS that went in concurrent directions according to whether unimodal or cross-modal adjustment was considered. The pattern of *unimodal* adjustment was steeper and appeared earlier with talking mothers. Many studies support the association between IDS and infant communication outcomes (see reviews: Golinkoff et al., 2015; Spinelli et al., 2017). To visualize the effect of IDS on the capacity of *unimodal* adjustment, we compared

the active adjustment towards a talking mother and a silent mother during *unimodal* adjustment (see Figure S2 in *Supplementary material*). Regardless of the age group, we can see that the presence of maternal IDS prompts infants to *inhibit* their silent-visual gestures when the mother is visually inattentive rather than to produce more silent-visual gestures when she is visually attentive. Previous studies showed that infants decreased their body movements when their mother was singing compared to when she was silent and this was interpreted as an increased attention paid to the mother (Scola et al., 2015). Here, the talking mothers may exert a comparable effect on infants' attentional focus leading them to decrease their silent-visual gestures. Another compatible explanation is that of Falk (2004) according to which IDS may soothe infants at a distance leading to fewer silent-visual gestures. These effects may come in synergy with the visual inattention from the mother and favour a sharper adjustment of infant communication.

On the other hand, from 11-to-20-months of age the pattern of *cross-modal* adjustment was steeper when the mothers were silent. One first interpretation is that the absence of IDS in addition to visual inattention may constitute a situation of global maternal inattention that is sufficiently strong to put a strain on the learning infant. In other words, full maternal inattention (neither visual attention, nor speech) may prompt infants to favour a communication match to regain their mother's attention by using *cross-modal* adjustment (see figure S2 in *Supplementary material*). Another, more artefactual, interpretation is that infants avoid vocalising when their mother is talking as they follow conversation rules, i.e., turn-taking. Several studies showed that infants avoid mutual vocalisations from early on (Gratier, 2003; Hillbrink, 2015; Northrup & Iverson, 2020b). A longitudinal study showed that 9-to-18-month-old infants gave a vocal response between 1150 and 650 ms after the end of the maternal vocal utterance (Hillbrink, 2015). However, our methodology of scoring maternal IDS, in a one-second window before and after infant signals, does not allow to disentangle between the two hypotheses.

Interestingly, we also found that infants favoured oral signals against audible gestures when the mothers were visually inattentive, but this effect was significant only with talking mothers. Even if the *cross-modal* adjustment pattern was more pronounced when the mothers were silent than when they were talking, only conditions with IDS elicited a preferential use of oral against audible gestures. This finding supports the idea that preverbal oral signals serve more communicative functions than simply an attention-getting function. Many previous studies investigating the coordination of communicative signals within mother-infant dyads showed that from very early on, when infants use vocalisations towards the mother, the mothers look at their infants and use IDS significantly more in return and vice versa (Northrup & Iverson, 2020b). Furthermore, other previous studies showed the effect of IDS on the infant's communication outcomes (Cristià, 2011; Goldstein & Schwade, 2008; Ma et al., 2011; Warlaumont et al. , 2015). In terms of vocal exchange, our results corroborate the results of some previous studies showing a temporal contingency between the mother and the infant during communicative exchanges (Northrup & Iverson, 2020b). In sum, the absence of IDS may *quantitatively* improve the capacity of *cross-modal* adjustment but its presence may *qualitatively* change the process of adjusting to attention, namely using oral signals preferentially. Further qualitative analysis is needed on this issue. Our finding supports the interest of considering both oral and gestural signals to investigate the key milestones of modal transition preceding the verbal phase.

We found differential developmental trajectories for the two strategies of adjusting to maternal inattention. The infants showed *unimodal* adjustment in the period of 7-10 months of age before showing *cross-modal* adjustment at 11-14 months. In a comparative perspective, the attention-sensitive signalling capacity has been investigated across diverse primate and non-primate species (Gaunet & Deputte, 2011; Kaplan, 2011; Maille et al., 2012; Pika & Bugnyar, 2011). While most of these studies found evidence for *unimodal* adjustment (e.g. Gaunet & Deputte, 2011; Maille et al., 2012; Pika & Bugnyar, 2011), only a few tested and

reported *cross- modal* adjustment (Bourjade et al., 2014; Horowitz, 2009). Furthermore, too few of them have investigated the attention-sensitive signalling capacity in a developmental approach (Fröhlich, et al., 2018; Horowitz, 2009; Tomasello et al., 1997) and only one investigated both *unimodal* and *cross-modal* adjustment in a developmental approach (Dafreville et al., 2021). Dafreville et al. (2021) showed that (i) immature chimpanzees showed *unimodal* adjustment after *cross-modal* adjustment and (ii) juveniles favoured contact gestures while adolescents preferred audible gestures when the mothers did not show visual attention. Thus, human infants present unimodal then cross-modal adjustment, while the opposite pattern is found in chimpanzees. Although the chimpanzee study included only gestural signals without considering oral signals, a first comparison with the present study already points to interesting modal differences in terms of developmental trajectories. One hypothesis is that the prominent channel of communication observed in mature individuals of each species is the latest to be adjusted to the mother's attention. This hypothesis is supported by recent results on chimpanzees showing that the species-specific mature forms of communication was longer to develop than other behavioural domain (Fröhlich & Hobaiter, 2018).

Conclusion

By using an ethological design and an extended behavioural repertoire, our study is the first to give a comprehensive overview of the development of attention-sensitive signalling. By dissociating *unimodal* and *cross-modal* adjustment, we evidenced that unimodal adjustment comes ahead of schedule on the one hand and before cross-modal adjustment on the other hand. Thus, our results showed that attention-sensitive signalling takes place in two stages over time and evidence the role of IDS in the development of this sociocognitive skill. In a comparative perspective, human infants' developmental trajectory differs from that of chimpanzees (Dafreville et al., 2021). This calls for further comparative analyses with other species as well as other human populations to depict the 'multi-modalic'

orchestration of the various communication means into a single stream (Kendon, 2011).

References

Altmann, J. (1974). Observational study of behaviour: sampling methods. *Behaviour*, *49*, 227–265.

Arbib, M. A., Liebal, K., & Pika, S. (2008). Primate vocalization, gesture, and the evolution of human language. *Current Anthropology*, *49*(6), 1053–1076. https://doi.org/10.1086/593015

Aureli, T., Spinelli, M., Perucchini, P., & Odorico, L. D. (2017). *The Pointing – Vocal Coupling Progression in the First Half of the Second Year of Life*. *22*(6), 801–818. https://doi.org/10.1111/infa.12181

Bernicot, J., Comeau, J., Feider, H., European, S., Bernicot, J., Comeau, J., & Feider, H. (1993). Speech Acts by Mother and Child : Determining their Nature and Form. *European Journal of Psychology of Education*, *8*(1), 35–50.

Bernicot, J., & Marcos, H. (1993). La demande prélinguistique et linguistique: qu'apporte le langage? *Journal International de Psychologie*, *28*(3), 291–306.

Bjorklund, D. F., & Pellegrini, A. D. (2000). Child development and evolutionary psychology. *Child Development*, *71*(6), 1687–1708. https://doi.org/10.1111/1467-8624.00258

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology and Evolution*, *24*(3), 127–135. https://doi.org/10.1016/j.tree.2008.10.008

Bourjade, M., Meguerditchian, A., Maille, A., Gaunet, F., & Vauclair, J. (2014). Olive baboons, Papio anubis, adjust their visual and auditory intentional gestures to the visual attention of others. *Animal Behaviour*, *87*, 121–128. https://doi.org/10.1016/j.anbehav.2013.10.019

Bruner, J. (1975). The ontogenesis of speech acts. *Journal of Child Language*, *2*(1), 1–19. https://doi.org/10.1017/S0305000900000866

Byrne, R. W., Cartmill, E., Genty, E., Graham, K. E., Hobaiter, C., & Tanner, J. (2017). Great ape gestures: intentional communication with a rich set of innate signals. *Animal Cognition*, *20*(4), 755–769. https://doi.org/10.1007/s10071-017-1096-4

Camaioni, L., Perucchini, P., Bellagamba, F., & Colonnesi, C. (2004). The Role of Declarative Pointing in Developing a Theory of Mind. *Infancy*, *5*(3), 291–308. Retrieved from https://doi.org/10.1207/s15327078in0503_3

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social Cognition , Joint Attention , and Communicative Competence from 9 to 15 Months of Age. In *Monographs of the Society for Research in Child Development* (Vol. 63, p. (Serial No.255)).

Colonnesi, C., Stams, G. J. J. M., Koster, I., & Noom, M. J. (2010). The relation between pointing and language development: A meta-analysis. *Developmental Review*, *30*(4), 352–366. https://doi.org/10.1016/j.dr.2010.10.001

Cristià, A. (2011). Fine-grained variation in caregivers' /s/ predicts their infants' /s/ category. *The Journal of the Acoustical Society of America*, *129*(5), 3271–3280. https://doi.org/10.1121/1.3562562

D'odorico, L., & Franco, F. (1991). *Selective production of vocalization types in different communication contexts*. *18*(1991), 475–499. https://doi.org/10.1017/S0305000900011211

Dafreville, M., Hobaiter, C., Guidetti, M., Sillam-Dussès, D., & Bourjade, M. (2021). Sensitivity to the communicative partner's attentional state: A developmental study on mother–infant dyads in wild chimpanzees (Pan troglodytes schweinfurthii). *American Journal of Primatology*, *83*(12). https://doi.org/10.1002/ajp.23339

Falk, D. (2004). Prelinguistic evolution in early hominins: Whence motherese? *Behavioral and Brain Sciences*, *27*(04). https://doi.org/10.1017/S0140525X04000111

Frankenhuis, W. E., & Tiokhin, L. (2018). Evolutionary Developmental Psychology: 2017 Redux. *Child Development*, *89*(6), 2282–2287. https://doi.org/10.1111/cdev.13018

Fröhlich, M., & Hobaiter, C. (2018). The development of gestural communication in great apes. *Behavioral Ecology and Sociobiology*, *72*(12). https://doi.org/10.1007/s00265-018-2619-y

Fröhlich, M., Sievers, C., Townsend, S. W., Gruber, T., & van Schaik, C. P. (2019). Multimodal communication and language origins: integrating gestures and vocalizations. *Biological Reviews*, *1820*, 1809–1829. https://doi.org/10.1111/brv.12535

Fröhlich, M., & van Schaik, C. P. (2020). Must all signals be evolved? A proposal for a new classification of communicative acts. *Wiley Interdisciplinary Reviews: Cognitive Science*, (February), 1–13. https://doi.org/10.1002/wcs.1527

Fröhlich, M., Wittig, R. M., & Pika, S. (2018). The ontogeny of intentional communication in chimpanzees in the wild. *Developmental Science*, *22*(1), 1–14. https://doi.org/10.1111/desc.12716

Gaunet, F., & Deputte, B. L. (2011). Functionally referential and intentional communication in the domestic dog: Effects of spatial and social contexts. *Animal Cognition*, *14*(6), 849–860. https://doi.org/10.1007/s10071-011-0418-1

Genty, E., Breuer, T., Hobaiter, C., & Byrne, R. W. (2009). Gestural communication of the gorilla (Gorilla gorilla): repertoire , intentionality and possible origins. *Animal Cognition*, *12*, 527–546. https://doi.org/10.1007/s10071-009-0213-4

Goldstein, M., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, *19*(5), 515–523. https://doi.org/10.1111/j.1467-9280.2008.02117.x

Golinkoff, R. M., Can, D. D., Soderstrom, M., & Hirsh-Pasek, K. (2015). (Baby)Talk to Me: The Social Context of Infant-Directed Speech and Its Effects on Early Language Acquisition. *Current Directions in Psychological Science*, *24*(5), 339–344. https://doi.org/10.1177/0963721415595345

Gratier, M. (2003). Expressive timing and interactional synchrony between mothers and infants: Cultural similarities, cultural differences, and the immigration exprerience. *Cognitive Development*, *18*(4), 533–554. https://doi.org/10.1016/j.cogdev.2003.09.009

Gros-Louis, J., & Wu, Z. (2012). Twelve-month-olds' vocal production during pointing in naturalistic interactions: Sensitivity to parents' attention and responses. *Infant*

Behavior and Development, *35*(4), 773–778. https://doi.org/10.1016/j.infbeh.2012.07.016

Grünloh, T., & Liszkowski, U. (2015). Prelinguistic vocalizations distinguish pointing acts. *Journal of Child Language*, *42*(06), 1312–1336. https://doi.org/10.1017/s0305000914000816

Gupta, S., & Sinha, A. (2019). Gestural communication of wild bonnet macaques in the Bandipur National Park, Southern India. *Behavioural Processes*, *168*(May 2018). https://doi.org/10.1016/j.beproc.2019.103956

Harrison, X. A. (2015). A comparison of observation-level randomeffect and Beta-Binomial models for modelling overdispersion in Binomial data in ecology & evolution. *PeerJ*, *2015*(7). https://doi.org/10.7717/peerj.1114

Hillbrink, E. (2015). Early developmental changes in the timing of turn-taking: a longitudinal study of mother–infant interaction. *Frontiers in Psychology*, *6*. https://doi.org/doi: 10.3389/fpsyg.2015.01492

Hobaiter, C., & Byrne, R. W. (2011). The gestural repertoire of the wild chimpanzee. *Animal Cognition*, *14*(5), 745–767. https://doi.org/10.1007/s10071-011-0409-2

Horowitz, A. (2009). Attention to attention in domestic dog (Canis familiaris) dyadic play. *Animal Cognition*, *12*(1), 107–118. https://doi.org/10.1007/s10071-008-0175-y

Igualada, A., Bosch, L., & Prieto, P. (2015). Language development at 18 months is related to multimodal communicative strategies at 12 months. *Infant Behavior and Development*, *39*, 42–52. https://doi.org/10.1016/j.infbeh.2015.02.004

Kaplan, G. (2011). Pointing gesture in a bird- merely instrumental or a cognitively complex behavior? *Current Zoology*, *57*(4), 453–467. https://doi.org/10.1093/czoolo/57.4.453

Kendon, A. (2011). Some modern considerations for thinking about language evolution: A discussion of The Evolution of Language by Tecumseh Fitch. *Public Journal of Semiotics*, *3*(1), 79–108. https://doi.org/10.37693/pjos.2011.3.8832

Kersken, V., Gómez, J. C., Liszkowski, U., Soldati, A., & Hobaiter, C. (2019). A gestural repertoire of 1- to 2-year-old human children: in search of the ape gestures. *Animal Cognition*, *22*(4), 577–595. https://doi.org/10.1007/s10071-018-1213-z

Liszkowski, U., Albrecht, K., Carpenter, M., & Tomasello, M. (2008). Infants ' visual and auditory communication when a partner is or is not visually attending. *Infant Behavior & Development*, *31*, 157–167. https://doi.org/10.1016/j.infbeh.2007.10.011

Liszkowski, U., Carpenter, M., Henning, A., Striano, T., & Tomasello, M. (2004). Twelve-month-olds point to share attention and interest. *Developmental Science*, *7*(3), 297–307. https://doi.org/10.1111/j.1467-7687.2004.00349.x

Liszkowski, U., Carpenter, M., Striano, T., & Tomasello, M. (2006). 12- and 18-Month-Olds Point To Provide Information for Others. *Journal of Cognition and Development*, *7*(2), 173–187. https://doi.org/10.1207/s15327647jcd0702_2

Locke, J. L. (2006). Parental selection of vocal behavior. *Human Nature*, *17*(2), 155–168. https://doi.org/10.1007/s12110-006-1015-x

Locke, J. L. (2017). Development of the Capacity for Spoken Language. In P. Fletcher & B. MacWhinney (Eds.), *The Handbook of Child Language*. https://doi.org/10.1111/b.9780631203124.1996.x

Locke, J. L., & Bogin, B. (2006). Language and life history: A new perspective on the development and evolution of human language. *Behavioral and Brain Sciences*, *29*(3), 259–280. https://doi.org/10.1017/S0140525X0600906X

Ma, W., Golinkoff, R. M., Houston, D. M., & Hirsh-Pasek, K. (2011). Word learning in infant- and adult-directed speech. *Language Learning and Development*, *7*(3), 185–201. https://doi.org/10.1080/15475441.2011.579839

Maille, A., Engelhart, L., Bourjade, M., & Blois-Heulin, C. (2012). To Beg, or Not to Beg? That Is the Question: Mangabeys Modify Their Production of Requesting Gestures in Response to Human's Attentional States. *PLoS ONE*, (7), e41197. https://doi.org/10.1371/journal.pone.0041197

Mehr, S. A., & Krasnow, M. M. (2017). Parent-offspring conflict and the evolution of infant-directed song. *Evolution and Human Behavior*, *38*(5), 674–684. https://doi.org/10.1016/j.evolhumbehav.2016.12.005

Molesti, S., Meguerditchian, A., & Bourjade, M. (2020). Gestural communication in olive baboons (Papio anubis): repertoire and intentionality. *Animal Cognition*, *23*(1), 19–40. https://doi.org/10.1007/s10071-019-01312-y

Murillo, E., Montero, I., & Casla, M. (2021). On the Multimodal Path to Language: The Relationship Between Rhythmic Movements and Deictic Gestures at the End of the First Year. *Frontiers in Psychology*, *12*(February), 1–8. https://doi.org/10.3389/fpsyg.2021.616812

Northrup, J. B., & Iverson, J. M. (2020). Multimodal coordination of vocal and gaze behavior in mother–infant dyads across the first year of life. *Infancy*, *25*, 952–972.

Oller, D. K. (2000). *The Emergence of the Speech Capacity*.

Oller, D. K., Griebel, U., & Warlaumont, A. S. (2016). Vocal Development as a Guide to Modeling the Evolution of Language. *Topics in Cognitive Science*, *8*(2), 382–392. https://doi.org/10.1111/tops.12198

Pearce, N. (2016). Analysis of matched case-control studies. *BMJ (Online)*, *352*, 1–4. https://doi.org/10.1136/bmj.i969

Pika, S., & Bugnyar, T. (2011). The use of referential gestures in ravens (Corvus corax) in the wild. *Nature Communications*, *2*(1), 560–565. https://doi.org/10.1038/ncomms1567

Rodrigues, E. D., Marôco, J., & Frota, S. (2021). Communicative gestures in 7–12-month infants: A phylogenetic comparative approach. *Infant and Child Development*, *30*(5), 1–11. https://doi.org/10.1002/icd.2262

Rodríguez, C., Moreno-Núñez, A., Basilio, M., & Sosa, N. (2015). Ostensive gestures come first: Their role in the beginning of shared reference. *Cognitive Development*, *36*, 142–149. https://doi.org/10.1016/j.cogdev.2015.09.005

Scola, C., Bourjade, M., & Jover, M. (2015). Social interaction is associated with changes in infants' motor activity. *Socioaffective Neuroscience & Psychology*, *5*(1), 28256. https://doi.org/10.3402/snp.v5.28256

Siegal, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences (2nd ed)* (McGraw Hil). Boston.

Spinelli, M., Fasolo, M., & Mesman, J. (2017). Does prosody make the difference? A meta-

analysis on relations between prosodic aspects of infant-directed speech and infant outcomes. *Developmental Review*, *44*, 1–18. https://doi.org/10.1016/j.dr.2016.12.001

Theobald, E. (2018). Students are rarely independent: When, why, and how to use random effects in discipline-based education research. *CBE Life Sciences Education*, *17*(3), 1–12. https://doi.org/10.1187/cbe.17-12-0280

Tomasello, M. (2008). *Origins of Human Communication* (The MIT Press, ed.). https://doi.org/https://doi.org/10.7551/mitpress/7551.001.0001

Tomasello, M., Call, J., Warren, J., Frost, G. T., Carpenter, M., & Nagell, K. (1997). The ontogeny ofchimpanzee gestural signals: A comparison across groups and generations. *Evolution of Communicati*, *1*(2), 223–259. https://doi.org/10.1075/eoc.l.2.04tom

Townsend, S. W., Koski, S. E., Byrne, R. W., Slocombe, K. E., Bickel, B., Boeckle, M., … Manser, M. B. (2017). Exorcising Grice's ghost: an empirical approach to studying intentional communication in animals. *Biological Reviews*, *92*(3), 1427–1433. https://doi.org/10.1111/brv.12289

Van de Pol, M., & Wright, J. (2009). A simple method for distinguishing within- versus between-subject effects using mixed models. *Animal Behaviour*, *77*(3), 753–758. https://doi.org/10.1016/j.anbehav.2008.11.006

Warlaumont, A. S., Richards, J. A., Gilkerson, J., & Oller, D. K. (2015). A social feedback loop for speech development and its reduction in autism. *Psychol Sci.*, *25*(7), 1314–1324. https://doi.org/10.1177/0956797614531023.A

Wu, Z., & Gros-Louis, J. (2014). Infants' prelinguistic communicative acts and maternal responses: Relations to linguistic development. *First Language*, *34*(1), 72–90. https://doi.org/10.1177/0142723714521925

Wu, Z., & Gros-Louis, J. (2017). The Value of Vocalizing: 10-Month-Olds' Vocal Usage Relates to Language Outcomes at 15 Months. *Infancy*, *22*(1), 23–41. https://doi.org/10.1111/infa.12150

Supplementary material

Attention-sensitive signalling by 7-to-20-month-old infants in a comparative perspective

Mawa Dafreville[1*], Michèle Guidetti [1], Marie Bourjade[1,2]

[1] CLLE, Université de Toulouse, CNRS, Toulouse, France
[2] Institut Universitaire de France

Table S1. Repertoire of human 7-to-20-month-**old infants' multimodal communication.**

Visual gestures			*Onset*	*Offset*
Body movement	Circular movement of the limbs (VAM)	Repetitive movements of the arms and/or legs and/or hands and/or fingers performed two or more times. Included: the infant holds a silent object during it; when the infant swings his/her arms while walking. Excluded : the infant holds a noisy object during it (code AO)	It starts at the image on which a movement is visible.	A new occurrence is recorded after a break longer than 1 second.
	Synchronised Trunk/Members movement (VSTM)	The infant leans back against the chair or against his/her mother and raises the pelvis up while stretching the trunk and the legs. The arms can be raised up or not. Included: when the infant is sitting on his/her mother's lap in an embrace position for example and leans back against her.	It starts when the trunk is stretching.	A new occurrence is recorded after a break longer than 1 second.
	Swagger (VS)	The infant moves his/her trunk along an axis (back and forth, from side to side, or up and down). An occurrence is recorded if at least a two-way movement is achieved (e.g. one movement back followed by one movement forth).	It starts when the movement begins.	When the infant stops moving.

Deictic gestures	Whole-hand pointing gesture (VWHPO)	The infant points with the whole hand and fingers extended in the direction of the object or another individual.	It starts when the infant begins his/her movement towards the object/individual	It ends when the infant releases his/her arm (even a little) or changes the target
	Index-pointing gesture (VIPO)	The infant points with his/her index finger towards an external object or individual (triadic pattern).	It starts when the infant begins his/her movement towards the object/individual.	It ends when the infant releases his/her arm or changes the target.
	Reaching attempt (VRAO)	The infant tries to reach out for one (or more) object(s)/or another individual out of range, without pointing, but leaning the trunk towards the object in a time window of 1 second before or after stretching his/her arm with his/her hand open. The infant can repeatedly open and close his/her fingers in the direction of the object (e.g., grasping movements). Excluded: when the infant makes the same movement but with an object in his/her hand already.	It starts when the infant begins his/her movement towards the object/individual.	It ends when the infant releases his/her arm or changes the target.
	Giving (VGO)	The infant holds the object with at least one hand towards the partner or he/she makes the object accessible to the recipient.	It starts when the infant begins his/her movement towards the object.	It ends when the infant no longer has the object in his/her hand.
	Receiving (VRO)	The infant takes/touches, with at least one of his/her hands, the object that the partner has handed or made accessible to him/her. Included: when the infant holds food that is not intended for him/her (e.g. when the partner is eating). Excluded: when it is included in the feeding (for example when the mother gives some pasta) if the infant was not holding it before.	It starts when the infant touches the object that the partner is handing to him/her.	It ends 1 second after the first contact between the infant and the object.
	Showing (VSO)	The infant without outstretched arm, holds in his/her hands an object in the direction of the partner, looking at the partner for more than 1 second while keeping the posture. Included: pointing towards himself/herself or the partner's body (dyadic pattern). For example, when the mother asks him/her to show her a part of his/her body ("where's your nose?").	It starts when the child looks at his/her partner with the object in his/her hands.	It ends when the child relaxes his/her arms or changes posture.

Conventional Gesture	Conventional Gesture (CVG)	The infant performs a conventional gesture, i.e. one that requires some forms of common ground to be understood. Included: yes-no with the head; Hi with the hand; bravo with his/her hands; Gestures from an ontogenetic ritualization (e.g. when the infant reaches both arms upwards towards his/her communication partner to be carried aloft by the partner or the infant hides his/her face to play *PeekaBoo*); We also included gestures from sign language; Also, if the infant touches his mother's breast with his/her hands and the mother makes a reference to breastfeeding.	It starts when the infant begins the conventional gesture.	It ends when the infant stops the movement during more than 1 second; a new occurrence is recorded after this interval.
Facial Expression	Facial Expression- Smiling (FX)	Any movement of the mouth likened to a smile, that results in an increased distance between the two corners of the mouth and any movement of the eyebrows that rise while looking at his/her mother, father or siblings.	It starts when the infant begins producing facial expression (the infant starts smiling) while looking at his/her partner.	A new occurrence is recorded after 1 second spent without smiling or looking at his/her partner.
Tactile gestures			*Onset*	*Offset*
Touching	Physical contact (TM)	The infant punctually engages a gesture with a part of his/her body (exclusively hands or head) that leads to physical contact with his/her mother (e.g. grab, bite, push, clap…). This has to last less than 5 seconds. If there are repetitive contacts less than 1 second apart by the same body part (for example the right hand), it is considered the same occurrence. If the contact made by the baby lasts more than a second, the behaviour is automatically coded as a behaviour lasting 1 second (segment of 1 second on ELAN). Included: when the infant touches his/her mother's breast (referring to breastfeeding VCG); When the infant touches his/her mother's clothes while embracing her (e.g. the infant grabs his/her mother dress).	It starts when an infant's body part touches his/her mother.	It ends when there is no physical contact between the infant and the mother or if the baby changes parts of body in contact.

		Excluded: when the infant is touching his/her mother because her/his mother is embracing or carrying him/her (indirectly touching); Also when the infant is touching the mother's breast while breastfeeding (except if he/she is touching his/her mother's face or hands); When the infant is touching his/her mother with an object.		
Audible Signals			*Onset*	*Offset*
Non-voiced sounds	Mouth sounds (VOMS)	Any non-voiced sound resulting from the air breathed in and out, without vibrating the vocal folds (e.g. sigh, tongue clicking, raspberry). Included: loud breathing sound that lasts less than 5 seconds. Excluded: if the infant has food in his/her mouth while producing mouth sounds (e.g. while breastfeeding).	It starts when the sound is first audible.	It ends only if the mother talks to her baby while the sound is stopped or after 1 second silent or if another type of oral sound replaces it (no acoustic gestures) while the sound is stopped
	Laughs (VOL)	Smiles with vocal or non-vocal sounds are defined as spontaneous laughs (Kawakami et al., 2007) continuing for more than 1 second, interspersed by silences during which he/she breathes in.	It starts when the sound is audible.	It ends only if the mother talks to her baby while the sound is stopped or after 1 second silent or if another type of oral sound replaces it (no acoustic gestures) while the sound is stopped.
Voiced sounds	Vowel-like sounds (VOP)	Vowel-like sounds due to the vibration of the vocal folds, as well as squeals and growls. Excluded: vegetative sounds (e.g., sneezes or burps).	It starts when the sound is first audible.	It ends only if the mother talks to her baby while the sound is stopped or after 1 second silent or if another type of oral sound replaces it (no

			acoustic gestures) while the sound is stopped.
Grunt (VOG)	The infant produces a vowel sound on breathing out with a rhythmic pattern where the impulse is on a strong beat and the resulting sound on a weak beat.	It starts when the sound is audible.	It ends only if the mother talks to her baby while the sound is stopped or after 1 second silent or if another type of oral sound replaces it (no acoustic gestures) while the sound is stopped.
Babbling (VOB)	The infant emits short or continuous sounds, composed of vowels and consonants forming isolated or repeated syllables, from the vibration of the vocal folds and the opening and closing of the mouth.	It starts when the sound is audible.	It ends only if the mother talks to her baby while the sound is stopped or after 1 second silent or if another type of oral sound replaces it (no acoustic gestures) while the sound is stopped.
Cries (VOCR)	The infant utters a continuous high-pitched expiratory sound intertwined with inspiratory sounds that can be voiced or not and accompanied by tears or not.	It starts when the sound is first audible.	It ends only if the mother talks to her baby while the sound is stopped or after 1 second silent or if another type of oral sound replaces it (no acoustic gestures) while the sound is stopped

	Pseudo-word (VONW)	The infant does not clearly say a word but utters a word which sounds like it. We recorded it only when the parent repeated the word to confirm it (e.g. when the child said, "Caaaaag" and his/her parent needs to confirm by asking him/her, "what do you want? Do you want some cake?").	It starts when the infant pronounces the first syllable of the pseudo-word.	It ends at the end of the pseudo-word.
	Word (VOW)	The infant says a single word of one of his/her parent languages that the parent understands.	It starts when the infant pronounces the first syllable of the word.	It ends at the end of the word.
	Word succession (VOWS)	The infant clearly says a clause comprising at least subject and an object that the parent understands.	It starts when the child pronounces the first syllable of the first word of the clause.	A new clause is recorded for each association of subject, verb and object.
Audible gestures	Sound made with an object (AO)	Any sound produced by an action in the environment with an object in the hands (e.g., hit, scratch, scrape). Included: if the infant pulls on a tree branch that makes a noise. Excluded: Oral sounds or if the sound results in a movement made by the mother (for example, she grasps her baby's arm in which he/she holds a maracas); Sounds made by an electronic toy.	It starts when the sound is audible.	It ends only if the mother talks to her baby while the sound is stopped or after 1 second silent or if another type of acoustic gesture replaces it (excluded : non-voiced sounds and vocalizations) while the sound is stopped.
	Sound made without an object (AWO)	Any sound produced by an action on the environment without an object (e.g., hit, scratch, scrape) or by a self-directed action (e.g., clap hands) and the baby is not moving in the room. Included: noise stirring the water (i.e. playing in a stream) Excluded: oral sounds	It starts when the sound is audible.	It ends only if the mother talks to her baby while the sound is stopped or after 1 second silent or if another type of acoustic gesture replaces it (excluded : non-voiced sounds and vocalizations) while the sound is stopped.

Table S2. Maternal attention variables used in the study

Maternal attention variables	**Definitions**
Physical contact	The mother has one or several body parts in physical contact with her infant.
Visual attention	The mother's head is directed towards her infant, the geometric lines of sight of the two individuals can cross or the infant is in the mother's line of sight. Visual attention is not considered if the mother has her eyes closed.
Auditory attention	The mother is sufficiently close to hear her infant vocalising. As the auditory perception of the mother cannot be measured, auditory attention was systematically considered as available. Warning: If the mother uses a food processor or other noisy equipment, please provide the information in comments. When the mother is out of sight, auditory attention was considered as available.
IDS	The mother is talking to her infant while interacting with her infant. Included: interjections (oh lala, hop, ah ahhh) Included: mouth sounds, interjections, onomatopoeia, babbling or protophone imitation. Excluded: mother's laughs.

Statistical analysis (detailed)

Firstly, Generalized Linear Mixed Models on proportion data (GLMM; Bolker et al., 2009; Harrison, 2015) were conducted with a binomial error distribution and the *logit link function* to determine which factors most affected the use of the different signal modalities, i.e., use of audible, silent-visual or contact signals. The dependent variable was the proportion of signals in a given modality for each infant. Fixed effects included *between-age* (mean age of each infant in months, range = 7-20), maternal visual attention state (two levels: attentive, inattentive), maternal IDS (two levels: silent mother, talking mother) and maternal physical contact (two levels: yes, no) as main fixed effects (key test predictors). To account for variation in the outcome variable caused by "nuisance" independent variables that were not considered in this study, we sought to identify which random effects could limit the variation for each signal modality (Theobald, 2018). To this end, many independent variables were integrated as potential random effects. As for the final model selection, the lowest AIC was considered the best-fitting model (that means here, limiting as much as possible the 'nuisance' of independent variables). In sum, for tactile signals, random effects included random slopes of *time* (referring to the time of the observation: before, during or after meal) and random intercepts for infant signaller. For silent-visual and audible signals, random effects included random slopes of *distance* with mother (referring to the distance between the mother and the signaller: close, far or out-of-sight, i.e. respectively a distance of less than one arm's length of the mother, more than one arm's length, and a distance at which the mother cannot be seen by her infant and/or the observer) and *time* and random intercepts for infant signaller. Some relevant interactive models were built on the basis of the theories of attentional conflict such as the interactive models where maternal IDS and physical contact variables interact either with each other or with other variables (e.g., maternal visual attention, between-age) but none of the interactive models involved involve more than three fixed effects to facilitate our interpretation. The models were implemented using the *glmer* function of the R package *lme4* version 1.1- 23 with the optimizer function of *aictab* of the R package *AICcmodavg* to calculate AICc , which is a transformation of the Akaike Information Criterion (AIC) used for a medium-sized sample. For the model selection, the models with the lowest AIC values

are considered the best-fitting models but models within a ΔAIC of 2 are considered to be equally good (Burnham & Anderson, 2004).

Secondly, we tested the effect of increasing age on the capacity for *unimodal adjustment* by fitting GLMM using a Poisson error distribution and a log link function (as in Dafreville et al., 2021). We coded a "match" (0) when the infants produced a silent- visual gesture to a visually attentive mother, a "mismatch" (1) when they produced a silent- visual gesture to a visually inattentive mother and then, we totalled the number of mismatches. Thus, the dependent variable was the sum of mismatches for each infant. As fixed effects we included IDS, *between-age* (in months; Van de Pol & Wright, 2009) and interaction between them. Infant identity was included as a random effect. The sample size was medium (30 mother-infant dyads) and the data did not follow a normal distribution, so we performed nonparametric tests (Siegal & Castellan, 1988) and included a Bonferroni correction in the case of multiple comparisons. All the non-parametric tests detailed below were conducted twice, on data (i) with maternal IDS ('talking mother') and, (ii) without it ('silent mother'). To test the capacity for *unimodal adjustment*, we compared the number of silent- visual gestures produced by the infants when their mothers were visually attentive to when they were not (as in Dafreville et al., 2021). We used a permutation test to test if the infants produced significantly fewer silent-visual gestures when the mothers did not show visual attention than when they did in each age group. In addition, it must be emphasised that at the age of 8-9 months old infants enter a stage of referential communication (Caselli, Rinaldi, & Volterra, 2019; Tomasello, 1995). Because this step may be hidden in the first age category of 7-10 months of age, we split it into two subgroups: younger and older than 8 months of age. We then replicated the test of *unimodal* adjustment when the mothers were silent in each sub-group, the only condition where the 7-10 age group did not show unimodal adjustment against the other age groups and conditions. **Thirdly**, to test the capacity for *cross- modal adjustment*, we compared the proportions of audible- or- contact signals (gestures and oral signals combined) and silent- visual gestures across conditions of maternal visual attention. We used a Fisher's exact test on paired data (Pearce, 2016) to determine if the infants produced significantly more audible- or- contact signals than silent-visual gestures when the mothers did not show visual attention than when they did in each age range (as in Dafreville et al., 2021) . **Fourthly**, to test the preferential use of audible signals (as compared to contact gestures) and oral signals (as

compared to audible gestures) when the mothers did not show visual attention, we used a permutation test to test the preferential uses of one type against the other one when the mothers did not show visual attention. **Finally**, to test the distribution of different multimodal combinations according to maternal visual attention, we used a Permutation test to compare the amount of multimodal combinations in one maternal visual condition as compared to the other one. After that, we used a Fisher's exact test to test the age effect on this distribution pattern. To test the dependence between the variables of maternal visual attention and the maternal IDS on the expression of *unimodal* and *cross-modal* adjustment and the distribution of multimodal combinations, we used a Fisher's exact test on each age group. All tests were conducted using R v3.6.1 software (http://cran.r-project.org) with p value equal or lower than 0.05 required for significance. All statistical tests were two- tailed.

Table S3. Summary of Cohen's Kappa coefficients for each infant's communication modality

Communicative modality	Infant's behaviour	Cohen's Kappa coefficient
Audible	Acoustic gestures	0.83
	Non-voiced sounds	0.72
	Vocalisations	
Silent-visual	Body movements	0.75
	Conventional gestures	
	Facial expression	
	Deictic gestures	
Tactile	Touch*	

** As tactile gestures were quite rare in the final database (in comparison to the other signals), we included them in the calculation of Cohen's Kappa coefficient of silent-visual gestures.*

Table S4. Summary of Cohen's Kappa coefficients for each mother's attention category

Mother's attention state	Cohen's Kappa coefficient
Visual and tactile (in)attention*	0.89
IDS	0.81

** As the auditory perception of the mother cannot be measured, auditory attention was systematically considered as equal to 1.*

Table S5. Summary of the models fitted with the proportion of each signal modality as dependent variable (n=8367). Interactions between two effects are represented by colons; bold characters indicate the best fitting models based on the lowest AICc; k is the number of parameters (Burnham & Anderson, 2004; Symonds & Moussalli, 2011) in the model; ΔAICc gives the difference in AICc between each model and the best fitting model; AICcWt is Akaike weight corresponding to the probability that a given model is the best approximating model.

Signals modality		**Silent-visual signals (random effect: VisualMot|ID)**				**Audible signals (random effect: DistanceMot,Time|ID)**				**Tactile signals (random effect: VisualMot,Time|ID)**			
Model type	**Fixed effects**	**K**	**AICc**	**ΔAICc**	**AICcWt(2)**	**K**	**AICc**	**ΔAICc**	**AICcWt(2)**	**K**	**AICc**	**ΔAICc**	**AICcWt(2)**
Null	Null	11	2459.63	137.76	0.00	22	2596.13	134.23	0.00	11	744.94	17.81	0.00
Single effect	Visual attention (VisualMot)	12	2420.82	98.95	0.00	23	2553.43	91.53	0.00	12	746.87	19.74	0.00
	Physical contact (TactileMot)	12	2455.54	133.67	0.00	23	2590.80	128.91	0.00	12	727.14	0.00	0.51
	IDS	12	2354.52	32.65	0.00	23	2492.59	30.70	0.00	12	745.80	18.66	0.00
	BetwAge	12	2459.14	137.27	0.00	23	2598.15	136.25	0.00	12	746.94	19.81	0.00
	Type of care(3)	13	2463.06	141.19	0.00	24	2596.07	134.17	0.00	13	745.43	18.29	0.00
	Sex(3)	12	2461.66	139.79	0.00	23	2598.12	136.22	0.00	12	746.98	19.84	0.00
	Breastfeeding(3)	12	2460.10	138.23	0.00	23	2595.73	133.83	0.00	12	743.58	16.44	0.00
	Mother parity (3)	12	2456.16	134.29	0.00	23	2596.66	134.76	0.00	12	744.70	17.56	0.00
	Family configuration(3)	12	2461.39	139.52	0.00	23	2597.47	135.58	0.00	12	745.70	18.56	0.00
Additive	VisualMot,TactileMot,IDS,BetwAge	15	2325.10	3.23	0.11	26	2465.42	3.52	0.14	15	733.02	5.88	0.03
Interactive	VisualMot:BetwAge	14	2417.82	95.95	0.00	25	2548.00	86.11	0.00	14	750.06	22.93	0.00
	TactileMot:BetwAge	14	2456.82	134.95	0.00	25	2594.72	132.83	0.00	14	730.98	3.84	0.07
	IDS:BetwAge	14	2357.76	35.89	0.00	25	2495.71	33.82	0.00	14	748.74	21.61	0.00
	VisualMot:TactileMot	14	2412.03	90.16	0.00	25	2549.37	87.48	0.00	14	729.08	1.95	0.19
	VisualMot:IDS	14	2323.12	1.25	0.31	25	2467.21	5.32	0.06	14	749.53	22.39	0.00
	IDS:TactileMot	14	2355.60	33.73	0.00	25	2490.13	28.23	0.00	14	729.37	2.23	0.17
	VisualMot,TactileMot:BetwAge	16	2416.59	94.72	0.00	27	2543.87	81.98	0.00	16	734.25	7.11	0.01
	VisualMot,IDS:BetwAge	16	2321.87	0.00	0.58	27	2461.90	0.00	0.81	16	751.74	24.61	0.00
	IDS,TactileMot:BetwAge	16	2359.71	37.84	0.00	27	2496.47	34.58	0.00	16	733.99	6.85	0.02

(1) *DistanceMot* refers to the distance between the mother and the infant (close: the mother is less than an arm's length from her infant; far: the mother is more than an arm's length from her infant; out-of-sight: the mother is totally out of sight of her infant), *ID* refers to the individual infant and *Time* refers to the time of the observation (before, during or after meal).
(2) Comparison and significance with the null model (***: p-value<0.001)
(3) As control predictors and further fixed effects, we included the *type of care* (three levels: nursery, nanny, stay-at-home parent), the *sex* of the infant (two levels: female, male), the weaning (two levels: breastfed, not-breastfed), the *mother parity* (number of siblings) and *family configuration* (two levels: two parents, single parent).

Table S6. Summary of the models fitted with the proportion of silent-visual gesture mismatching (n=1422). Interactions between two effects are represented by colons; bold characters indicate the best fitting models based on the lowest AICc; k is the number of fixed effects in the model; ΔAICc gives the difference in AICc between each model and the best fitting model; Wi is Akaike weight corresponding to the probability that a given model is the best approximating model.

		Unimodal mismatching (random effect: ID)			
Model type	**Fixed effects**	**K**	**AICc**	**Delta_AICc**	**AICcWt**[(1)]
Null	Null	2	4495.61	80.29	0
Single effect	IDS	3	4439.72	24.41	0
	Mean age	3	4496.66	81.35	0
Interactive	DS:Mean age	5	4415.31	0.00	1***

[(1)] Comparison and significance with the null model (***: p-value<0.001)

Table S7. Parameter estimates for the best fitting model – GLMM with the number of silent-visual gestures (n=1422) mismatching as dependent variable and mean age (in months), IDS and maternal visual attention as fixed effect. There was a clear influence of age on unimodal matching (null model comparison: X^2=609.97. df=1. p<0.001).

Parameters	**Unimodal mismatching ~ IDS: Mean age**					
	Estimate	**SE**	**Lower CI**	**Upper CI**	**z-value**	**P-value**
(Intercept)	-2,888	0,271	-3,4184	-2,3568	(1)	(1)
IDS	-0,739	0,229	-1,1878	-0,2907	-3,230	p<0.01
Between-age	-0,068	0,021	-0,1089	-0,0268	-3,242	p<0.01
IDS: Between-age	0,090	0,017	0,0559	0,1240	5,180	**p<0.001**

[(1)] Not indicated because of limited interpretive value

Figure S1. Proportion of audible-or-contact signals and silent-visual gestures produced by maternal visual attention condition and age range <u>when the mothers were talking</u> (*: p-value<0.05; **: p-value<0.01; ***: p-value<0.001). The proportion of audible-or-contact signals is calculated like "number of audible-or-contact signals divided by the sum of all signals" and the proportion of silent-visual signals in the same way. Thus, when the mean proportion difference between the proportions of audible-or-contact and silent-visual signals is larger when the mother is not visually attentive than when she is, this indicates cross-modal adjustment. *Silent mother* refers to the condition without IDS; *Talking mother* refers to the condition with IDS. Large black circles represent

the mean proportion per subject. Median (horizontal lines), quartiles (boxes), percentiles (2.5% and 97.5%), vertical lines, and outliers (small black circles) are indicated.

Figure S2. Variation in use of silent- visual gestures with respect to IDS by age range and maternal visual attention (N = 11, *p < 0.05. ***p < 0.001). The deviations above and below the zero- line show changes (plus standard error bar) in the distribution of silent- visual gestures across maternal visual condition and IDS, from the overall average use of silent- visual gestures in this condition regardless of IDS.

Note: To represent the IDS effect on the adjustment of the silent-visual gestures towards the maternal visual attention state, we calculated the percentage deviation in the variation in the use of silent-visual gestures in attentive condition as compared with the use of silent-visual gestures in inattentive condition for each IDS condition and age range. The deviation was calculated for each individual by $(\beta/a - 1) \times 100$ *with* a *= number of silent-visual gestures in attentive condition (regardless of IDS) / total silent-visual gestures in the sample (regardless of IDS), and* β*= number of silent-visual gestures in attentive and IDS condition / total number of silent-visual gestures used in the condition IDS (regardless of maternal visual attention). Then, we calculated the mean by age group.*

Figure S3. Variation in use of audible- or- contact and silent- visual gestures with respect to maternal visual attention by age range when the mothers were silent (N = 11, *p < 0.05. ***p < 0.001). The deviations above and below the zero- line show changes (plus standard error bar) in the use of each modality, according to the maternal state of attention before signalling, from the overall average use of that modality in signalling.

Note: To represent active adjustment of the different gesture modalities towards the maternal visual attention state, we calculated the percentage deviation in the variation in the use of audible- or- contact as compared with silent- visual gestures for each condition of maternal attention and age category (as Hobaiter & Byrne, 2011a). The deviation was calculated by $(\beta/a - 1) \times 100$ *with* a *= number of audible- or-contact signals/ total number of signals used in the age- range subgroup, and* β*= number of audible- or- contact signals / total number of signals used in the condition and age- range subgroup.*

4.2. ARTICLE 3: BEYOND POINTING GESTURES: ATTENTION-SENSITIVE COMMUNICATION IN 7-TO-20-MONTH-OLD INFANTS DEVELOPS PRAGMATIC SKILLS

Mawa Dafreville[1*], Michèle Guidetti [1], Marie Bourjade[1,2]

[1] CLLE, Université de Toulouse, CNRS, Toulouse, France

[2] Institut Universitaire de France

Article soumis à Developmental Psychology en 2022

Synthèse de l'article 3

Contexte : Toujours dans l'investigation de la *communication sensible à l'attention du destinataire*, cette étude visait à étudier la capacité des enfants préverbaux de 7 à 20 mois à ajuster leurs signaux *canoniques* (i.e. des comportements prédicteurs du langage) ou autres signaux issus d'un *répertoire plus large* (i.e. présentant un intérêt dans le développement communicationnel de l'enfant mais non considérés comme prédicteurs langagiers) qui seraient également perçus par la mère dans l'interaction quotidienne.

Méthode : A partir de ces mêmes observations en condition écologique de trente dyades mère-enfants humaines, il a été question de tester les ajustements *unimodal* et *intermodal* sur des signaux *canoniques* et issus d'un *répertoire plus large*. Les signaux *canoniques* incluaient les gestes déictiques *distaux* (incluant le geste de pointage) et le babillage. Les signaux issus d'un *répertoire plus large* incluaient les gestes déictiques *proximaux* (e.g. *showing* et *giving*), les gestes audibles *avec* et *sans* objet et les signaux oraux hors babillage (i.e. les bruits de bouche et les vocalisations non-consonantiques).

Résultats : Les enfants ajustaient leurs signaux *canoniques* mais aussi de nombreux signaux issus d'un répertoire plus large de communication à l'attention visuelle de leur mère (i.e. les *mouvements corporels*, les gestes audibles *avec* des objets et les bruits de bouche). Dans l'ensemble, les signaux coordonnés au regard vers le destinataire n'étaient pas significativement mieux ajustés à l'attention maternelle sauf pour quelques signaux silencieux-visuels spécifiques à certains âges seulement.

Conclusion : Dans l'ensemble, ces résultats indiquent que la *communication sensible à l'attention* est pertinente pour le développement de compétences pragmatiques précoces et que la fonction illocutoire des signaux pourrait être plus fidèlement appréhendée par cette capacité que par la coordination du regard avec ces signaux.

Beyond pointing gestures: attention-sensitive communication in 7-to-20-month-old infants develops pragmatic skills

Mawa Dafreville[1*], Michèle Guidetti [1], Marie Bourjade[1,2]

[1] CLLE, Université de Toulouse, CNRS, Toulouse, France

[2] Institut Universitaire de France

*Corresponding author: mawa.dafreville@univ-tlse2.fr

Abstract:

The present study aimed at investigating the ability of 7-to-20-month-old infants to display *attention-sensitive communication* using either canonical markers of language acquisition (e.g. pointing gestures, canonical babblings) or other signals based on the physical features *actually perceived* by the mother in everyday interaction (e.g. body movements, mouth sounds). We studied 30 French mother-infant dyads in naturalistic settings. We assessed the infants' *attention-sensitive communication* through *unimodal* and *cross-modal* adjustment, defined as the capacity of infants to address visually inattentive mothers by avoiding visual communication mismatches and/or favouring communication matches through audible-or-contact signals. *Unimodal* and *cross-modal* adjustments were tested for specific signals across spontaneous 'conditions' of maternal visual attention (attentive/inattentive) from video footage filmed in the home. Both canonical markers of language development and signals belonging to an extended repertoire of communication were used by infants to adjust to their mother's visual attention. Gaze-coordinated canonical distal pointing gestures and canonical babbling were used for respectively *unimodal* and *cross-modal* adjustment by 11-to-14-month-old infants. However, many other signals belonging to the extended communicative repertoire were used in order to adjust to maternal attention such as *body movements*, *audible gestures with objects*

or *mouth sounds*. Gaze-coordinated signals were overall not significantly better adjusted to maternal attention than non-gaze-coordinated signals, except for specific silent-visual signals at certain ages. Overall, these results indicate that attention-sensitive communication is relevant to the development of early pragmatic skills and that the illocutionary use of signals may be more reliably approximated by this capacity than by gaze-coordination with signals.

Keywords: perception modality, production channel, canonical pointing, canonical babbling, mother–infant interaction

Introduction

Infants aged eleven to sixteen months transition from the use of gestures to the use of spoken words to communicate with surrounding adults (Aureli et al., 2017; Bates et al., 1975; Igualada et al., 2015; Mccune & Zlatev, 2015; Murillo et al., 2021). Speech onset results in the coexistence of the two communication modalities, although the vocal modality progressively becomes prominent (Guidetti et al., 2014; Iverson & Goldin-Meadow, 2005). According to *evo-devo* theories, this human-specific vocal prominence may have emerged over the course of evolution from the selective pressures on both infants and parents to provide mutual attention (Falk, 2004; Locke, 2017; Locke & Bogin, 2006; Mehr & Krasnow, 2017; Oller et al., 2016). Relatedly, some lab-based developmental studies found that the use of vocalizations coupled with pointing gestures was a function of the visual attention state of the recipient (Igualada et al., 2015; Liszkowski et al., 2008; Wu & Gros-Louis, 2017). More comprehensive studies on attention-sensitive communication have begun to use comparative data with an extended ethological repertoire to provide evidence of infants' abilities to adjust the modality of their communicative signals to the adult's attention (Bourjade et al., *submitted*.; Dafreville et al., *submitted*; Dafreville et al., 2021; Rodrigues et al., 2021). However, the extent to which these attention-sensitive abilities are driven by the adjustment of canonical markers of language acquisition such as pointing

gestures or babbling, or by the pragmatic use of other visual and audible signals has not been established. To date, gestures have been considered hallmarks of pragmatic development while the use of prelinguistic vocalizations for their signal value, i.e., *illocutionary* function (Bates et al 1975; Bernicot 1993, Bruner 1975), has been somewhat overlooked.

Indeed, many studies have investigated the acquisition of specific *gestural* signals in relationship with subsequent language skills. Infants acquire the use of deictic gestures such as pointing at the end of their first year (see Table 1). Infants of this age point to distal entities and start directing the adult's attention to them with accompanying gaze alternation between the object and the adult (Beuker et al., 2013; Camaioni et al., 2004; Carpenter et al., 1998; Liszkowski et al., 2012). Pointing gestures are universally produced across cultures (Liszkowski, Brown, Callaghan, Takada, et al., 2012; Salomo & Liszkowski, 2013) and predict language onset and vocabulary size later in development (Cochet & Byrne, 2016; Colonnesi et al., 2010; Kita, 2003; Liszkowski et al., 2012) although with differences across cultures (Kita, 2003; Mastin & Vogt, 2016; Salomo & Liszkowski, 2013). While canonical pointing is considered a distal finger-pointing gesture and has been attributed a declarative illocutionary function (Bates et al., 1975; Bourjade et al., 2020; Camaioni, 1997; Cochet & Byrne, 2016), recent studies have begun to investigate proximal deictic gestures, in which a referent object remains in contact with the hand, e.g. *showing* and *giving* gestures (Murillo et al., 2019), as a likely transition to distal deictic gestures (Liszkowski et al., 2008; Liszkowski et al., 2006) that can even predict language skills at 18 months of age (Choi et al., 2021). In line with these recent investigations, the study of the pragmatic usage of gestural signals has extended to bodily gestures such as body or rhythmic movements. Scola et al. (2015) showed that 5-to-9-month-old infants produced more *body movements* when the partners were singing than when they were not and thus, exhibited in this way early forms of interaction with an adult. In addition, Murillo et al. (2021) showed that 9-to-12-month-old infants frequently used *rhythmic movements with objects* in their

repertoire and that when they were combined with vocalisations at 9 months of age, they were positively related with the use of *proximal deictic* gestures at 12 months of age.

Prelinguistic vocalisations have long been considered as speech precursors leading to extensive studies of their acoustic and phonological characteristics (Oller et al., 2013; Oller et al., 1998). Protophones are speech precursors excluding vegetative sounds (e.g. coughing, sneezing, burping) and fixed vocal signals (e.g. crying, laughter; Oller, 2000; Oller et al., 1999). Quasivowel sounds emerge over the first 2 months of life and refine over the 3rd to the 8th months of life during which infants produce mainly vowel-like sounds. Between 5 and 10 months of age, infants progressively acquire canonical babbling, consisting of well-formed syllables, i.e., a rapid transition between at least one full vowel-like element and one consonant-like element (Oller et al., 1999). *Babbling* constitutes the most sophisticated prelinguistic vocalization and arouses particular interest because it elicits more responses from caregivers (Bloom & Beaumont, 1993; Bloom & Lo, 1990; Gros-Louis et al., 2014; Locke, 2006), predicts consonant repertoire size (Jensen et al., 1988), speech onset (Oller et al., 1999) and language skills, both expressive and receptive (Jensen et al., 1988; Oller et al., 1998; Stoel-Gammon, 2011). Interestingly, the other protophones, also called non-consonant-vowel (*non-CV*) vocalizations (e.g. Donnellan et al., 2019) or cooing (e.g. Locke, 2006) also represent significant speech precursors although much less is known about their signal value, that is their social use by infants for their (perlocutionary) effects on- and (illocutionary) recognition as signals by- parents, i.e., their pragmatic use (but see Donnellan et al., 2019). In line with this, the perlocutionary effects of a signal on the receiver depend on the sensory modalities of *perception*, more than on those of *production*. This line of investigation has long been applied in comparative and evolutionary psychology that relies on ethological signal coding of perceived modalities; i.e., silent-visual, audible or tactile signals (Byrne et al., 2017; Tomasello & Call, 2018). Recent studies in 6- or 7-to-20-month-old infants investigated

audible gestures, but also mouth sounds for their signal value (Bourjade et al., *submitted*; Dafreville et al., *submitted*; Rodrigues et al., 2021). Mouth sounds, i.e., non-voiced sounds made with the mouth (e.g., a sigh), correlated positively with the time spent gazing at the mother in 6-month-old infants (Bourjade et al., *submitted*). Comparative work also highlighted non-voiced proto-consonants in orangutans (Lameira & Call, 2018) and a language-like hemispheric specialization of the chimpanzee raspberry, i.e., a bilabial fricative made with air expired between protruded lips (Reynolds Losin, Russell, Freeman, Meguerditchian, & Hopkins, 2008; Wallez et al., 2012). Although the raspberry is considered a protophone (Oller, 2000), to the best of our knowledge no study has investigated its real contribution to human language-speech development.

Whether infants *intend* to communicate with silent-visual signals such as pointing or audible signals such as vocalizing, hence displaying *illocutionary communication*, was initially evidenced with infants' gaze coordination to the signal (Bates et al., 1975; Csibra & Volein, 2008). Thus, most studies have so far considered gaze-coordination to a signal as being a sufficient criterion to evidence at a minimum, a *first-order* intentionality (Dennett, 1983). Many studies therefore retained only gaze-coordinated signals in their analyses (Aureli et al., 2017; Gros-Louis & Wu, 2012; Igualada et al., 2015; Iverson & Goldin-Meadow, 2005; Rodrigues et al., 2021; Wu & Gros-Louis, 2014, 2017). In line with this, Donnellan et al. (2019) tested the gaze-coordination of different types of signals addressed to the caregiver. They found more gaze-coordinated silent-visual gestures than gestures not coordinated with gaze. This held true for audible-visual coupling of signals but not for audible signals alone in 11-month-old infants. However, they also showed that gaze-coordinated consonant vocalisations constituted a good predictor of language skills. Collectively, these results point up the limits of using gaze-coordination as the only criterion for establishing intentional signal use. Relatedly, Liszkowski (2006,2008) suggested that gaze-coordination may not be the best criterion to evidence a communicative intent in pointing gestures. First, an

infant may look at the recipient without communicative intent. Second, non-gaze-coordinated signals can be used with a communicative intent when the signaller relies on an auditory cue rather than a visual one or when gaze is part of a conversation and alternates with communicative signals (Liszkowski, 2006; Liszkowski et al., 2008; Rossano, 2013).

In line with this proposition, Liszkowski and collaborators (2008) investigated *attention-sensitive communication* in 12- and 18-month-old infants. This capacity stems from the basic perceptive constraint whereby silent-visual signals are not noticeable when the recipients do not show prior visual attention. They showed that 12-month-old infants decreased the frequency of pointing gestures towards inattentive adults as compared to attentive ones. By comparing infants of 12 and 18 months of age, the authors found that only 18-month-old infants produced more vocalizations when the adults were not visually attentive and not responsive than when they were visually attentive and responsive (Liszkowski et al., 2008). Further lab-based studies evidenced that attention-sensitive communication relates to subsequent language skills. Wu and Gros-Louis (2017) showed that 10-month-old infants who produced more vocalizations when their mother did not show visual attention than when she did, exhibited higher language skills at 15 months of age. Likewise, infants who increased their combinations of vocalizations with pointing gestures when the parent was attending to them but not to the target of their point had enhanced subsequent language skills (Igualada et al., 2015; Wu & Gros-Louis, 2014). There is also a recent endeavour to study attention-sensitive communication with an ethological approach based on an extended repertoire of communicative signals (Bourjade et al., *submitted*.; Dafreville et al., *submitted*; Dafreville et al., 2021; Rodrigues et al., 2021). In a comparative study on immature chimpanzees, were used two complementary descriptors of attention-sensitive communication (Dafreville et al., 2021). *Unimodal adjustment* refers to the capacity of the signaller to deploy fewer silent-visual signals when the recipient does not show visual attention as compared to when she does. *Cross-modal*

adjustment refers to the capacity of the signaller to deploy more audible-or-contact signals than silent-visual ones when the recipient does not show visual attention as compared to when she does. Bourjade et al. (*submitted*) experimentally addressed this ability in 6-month-old infants. They found that infants displayed a rudimentary form of unimodal adjustment in that they adjusted their silent-visual gestures to maternal visual attention and interaction. In an ecological context of free play, 7-to-12-month-old infants display cross-modal adjustment of their communicative gestures (Rodrigues et al., 2021), which is earlier than the onset found in lab-based experiments, namely 10 months of age for cross-modal (Gros-Louis & Wu, 2012; Wu & Gros-Louis, 2014) and 12 months of age for unimodal adjustment (Liszkowski et al., 2008; Wu & Gros-Louis, 2017). In contrast, Dafreville et al. (*submitted*) provided evidence that *unimodal* adjustment emerges before *cross-modal* adjustment in naturalistic settings, from respectively 8 and 11 months of age. However, all communicative signals of a given sensory modality were clumped together in the latter study. It raises the question of whether canonical markers of language acquisition are preferred by infants to adjust their communication modality to the visual attention of the recipient or whether other signals are also adjusted (Table 1).

Table 1. Brief categorization of the canonical markers of language development with corresponding possible extensions of the repertoire. A *canonical marker* was defined here as a well-known developmental milestone predicting subsequent language skills in more than three published articles. The *extended repertoire* was composed of other signals presenting growing interest for early pragmatic skills in recently published articles. *Non-consonant-vowel vocalisations (Non-CV voc.)* are voiced protophones (i.e. produced with vocal cords) different from babbling (e.g. grunts, squeals, growls; Oller, 2000).

Perception modality	Canonical marker	References	Extended repertoire	References
Silent-visual	**Distal deictic gestures**	(Cochet & Byrne, 2016; Colonnesi et al., 2010; Goldin-Meadow & Alibali, 2014; Iverson & Goldin-Meadow, 2005; Liszkowski et al., 2012; Mcgillion et al., 2017).	**Body movements**	(Murillo et al., 2019; Scola et al., 2015)
			Proximal deictic gestures	(Choi et al. , 2021)
Audible	**babbling (vowel-consonant vocalization)**	(Donnellan et al., 2019; Jensen et al., 1988; Oller et al., 1998; Oller et al., 1999; Stoel-Gammon, 2011; Stolt et al., 2012)	**Mouth sounds**	(Lameira et al., 2013; Wallez et al., 2012)
			Non-CV voc	(Donnellan et al., 2019; Fasolo & D'Odorico, 2012)
			Pseudo-words and words	(Fasolo & D'Odorico, 2012)
			Audible gestures *with* and *without* object	(Murillo et al., 2019; Rodrigues, 2021)

The present study aims at bridging this gap by investigating the ability of 7-to-20-month-old infants to display *attention-sensitive communication* using either canonical markers of language acquisition or other signals based on the physical features *actually perceived* by the mother in everyday interaction. *Unimodal* and *cross-modal* adjustments were tested for specific signals across spontaneous 'conditions' of maternal visual attention (attentive/inattentive) from video footage filmed in the home. The infants' capability for *unimodal* adjustment was tested on *deictic gestures* (canonical marker) and *body movements* (extended repertoire). The infants'

capability for *cross-modal* adjustment was tested on babbling (canonical marker) but also on mouth sounds, non-CV vocalizations, pseudo-words and words (verbal content) and audible gestures with and without objects (extended repertoire). Each analysis was conducted on three age categories, [7-10], [11-14] and [15-20], in order to document developmental effects and was replicated on a smaller data set consisting of gaze-coordinated signals only so as to give an indication of the patterns of gaze-coordinated signal use.

Method

Participants

This study stems from data collected from March to July 2019 by MD in the city of Toulouse and its surroundings (France). Mother-infant dyads came mainly from urban areas. Participants were recruited via the BabyLab platform of the University of Toulouse Jean Jaurès and via nurseries (i.e. distribution of flyers and emails). A pre-home visit with the purpose of mutual familiarization allowed the observer to briefly inform the mother about the study and to ask her a few questions about her infant (family configuration, type of childcare, etc.) using a short demographic questionnaire. Most of the participants were families with first-born infants (n=18 out of 30, 60%) and the average number of siblings per infant observed was 1.87. Mothers' ages varied between 27 and 40 years. The mothers had a high level of formal education, as they all had graduate degrees (at least bachelor level). Most of the infants were totally breastfed (n=24 on 30 mother-infant dyads, 80%). Nine individuals had also experienced health difficulties at birth (i.e. preterm; difficult childbirth; asthma).

In this cross-sectional study, the thirty mother-infant dyads were allocated to three infant age categories spanning roughly *before*, *during* and *after* the emergence of intentional pointing gestures at the mean age of 11-12 months (Camaioni et al., 2004; Carpenter et al., 1998; Colonnesi et al., 2010). The three infant age groups were: 7 to 10 months of age, 11 to 14 months of age and 15 to 20 months of age (more details in table 2).

Table 2. Age of infants included in the study by session and mean age (± SD) by age range group.

Age range in months	ID	Sex	Session 1	Session 2	Mean age in months (± SD)
7-10 (N = 11)	TH23	M	7	8	*Females*: 8.83 (± 1.03) *Males*: 8.80 (± 1.09) *All*: 8.81 (± 1.00)
	SA8	M	7,3	8,2	
	CH19	F	7,5	8,3	
	AD3	F	7,7	9,1	
	ZO13	F	8,1	*NA*	
	AP26	F	8,5	9,5	
	CL7	M	8,8	9,8	
	EL15	F	9,1	*NA*	
	AL30	M	9,2	10,1	
	PA4	F	9,2	10,2	
	TH16	M	9,8	*NA*	
11-14 (N = 12)	CL17	M	11,3	12,5	*Females*: 12.78 (± 0.99) *Males*: 13.11 (± 0.91) *All*: 13.03 (± 0.90)
	AR11	M	11,4	12,4	
	LO6	F	11,4	12,6	
	AL24	M	11,8	13,6	
	CO15	M	11,8	12,9	
	MA26	F	12	12,9	
	AL20	M	12,9	14,2	
	MA12	M	13	13,9	
	LA16	F	13,4	14,4	
	AI23	M	13,5	14,5	
	GA3	M	13,5	14,4	
	AX15	M	13,6	14,8	
15-20 (N = 7)	AU30	M	14,9	16,1	*Females*:18.53± 1.59 *Males*: 16.41 (± 0.93) *All*: 17.01 (± 1.44)
	AR13	M	15	16,1	
	LO3	M	15,9	17,4	
	AN13	M	16	17,2	
	VI13	F	16,9	17,9	
	NI7	M	17	18,5	
	MA25	F	19	20,3	

Procedure and design

Dyads were observed in naturalistic conditions at home during meals (lunch, dinner, afternoon snack time). As visits aimed at accessing the most natural conditions, the observations took place whether other members of the family were present or not. During the observations, the observer stayed as much as possible on the side-lines and did not interact with individuals until 5 minutes before the end of each observation. During on average 55 minutes of observation, the observer was not engaged in any interaction with the mother or the infant. During

the last 5 minutes of observation, semi-structured interviews related to parenting practices, non-parental care patterns, weaning, breastfeeding, etc., were performed. Thus, each dyad was observed twice for on average 1 hour, one month apart. The recording started on average 15 minutes before the start of the meal (taking care that the infant was awake). Both visits were made so that the infant remained in the same age group during the two observations. The visits were filmed using two cameras: a camera (PANASONIC HC-V770) mounted on a tripod that provided an overview of the room (mainly the living room and/or dining room) and the movements of the mother in particular; and a camera (HANDYCAM HDR-CX625) held by the observer, who focused only on the infant and followed him/her. All individuals were kept for further analysis, even if there was only one observation per infant (three infants were not observed twice).

Ethical procedure

During the first visit, the parents of infant participants completed a consent form on behalf of their infants and were questioned about their profession, the sleep-rhythm and diet habits of their infant. During the last visit, the experimenter added an amendment to the initial signed consent in which the mother authorized the researchers to study her behaviour in addition to that of her infant. The video and coding data were anonymized and stocked on external drives stocked in safe places locked until 15 years after the end of the study when they will be destroyed. The video data are exclusively used for the purpose of analysis. The parents also completed an image release agreement allowing us to use video data for the purpose of teaching or scientific communication. The research design was approved by the ethical committee of the University of Toulouse (i.e., CER-NI; n° IRB00011835-2019-03-19-141).

Behavioural sampling and coding procedure

Infants' communicative signals were coded in ELAN software (Version 5.2) following the repertoire in *Supplementary material*, Table S1 (and briefly summarized in Table 3). The repertoire included about 23 behaviours classified into the three categories based on their main sensory modalities tested here, i.e. visual and audible ones. In the repertoire we considered *audible signals*, which encompassed oral and gestural signals, while *silent-visual* and *contact* signals were exclusively gestural.

Table 3. Brief categorization of the communicative signals considered is this study**.** The detailed repertoire of behaviours is available in the *Supplementary material* (see **Table S1**).

Perception modality	Production channel	Category of communicative signals
Silent-visual	Gestural	Body movements
		Deictic gestures
Audible	Oral	Mouth sounds
		Non-CV vocalisations
		Babblings
		Verbal content
	Gestural	With object
		Without object

Signals were extracted from a larger database comprising recipients other than the mother and various maternal attention states, such as tactile contact or infant directed speech. Here we selected signals that were addressed to the mother only (by removing all the signals potentially directed towards a non-maternal recipient), and we removed the communicative events when she was talking to the child so as to better capture the infants' abilities to adjust to *visual attention*. All the signals analysed here were therefore addressed to a silent mother who was either visually attentive or not (see Dafreville et al., *submitted*). For each signal, we considered the following information: (1) the *sensory modality* (audible or silent-visual), (2) the *type of signal* (see Table 3), (3) if the signal was *gaze-coordinated,* i.e. associated to a gaze towards the mother (at least once in a window from 1 second

before the start of the signal to 1 second after the end of it) and (4) the *maternal visual attention* state. Visual attention was recorded when the mother had her head directed at her infant from narrow to large angles, the geometric lines of sight of the two individuals being able to cross each other.

Inter-rater reliability

The videos were entirely coded by one main coder, who was hence used as the reference for training a second blind coder. The second coder trained for one month on a small part of the video corpus, then blindly coded 20% of the video corpus. Reliability was assessed through Cohen's Kappa statistics. The second coder reached a Cohen's Kappa coefficient of almost 80% with the first coder for infants' visual communicative behaviours. For infants' audible signals, Cohen's Kappa ranged from 0.72 for oral sounds to 0.83 for audible gestures *with* and *without* object. Agreement for maternal attention was high (Cohen's k > 0.80) and also for infant gaze direction (Cohen's k= 0.92).

Statistical analysis

Because the sample size was medium (30 mother-infant dyads) and the data did not follow a normal distribution, we used non-parametric statistics to test defined hypotheses (Siegal & Castellan, 1988), including Bonferroni corrections in cases of multiple comparisons. One-sample permutation tests were performed to determine if the infants produced significantly fewer silent-visual signals when the mothers did not show visual attention than when they did, i.e. *unimodal adjustment* (as in Dafreville et al., 2021). Fisher's exact tests on paired data (Pearce, 2016) were performed to determine if the infants produced significantly more audible-or- contact signals than silent-visual gestures when the mothers did not show visual attention than when they did, i.e. *cross-modal adjustment* (as in Dafreville et al., 2021). All tests were conducted using R v3.6.1 software (http://cran.r-project.org) with p-values equal to or lower than 0.05 required for significance. All statistical tests were two- tailed.

Results

We analysed 506 silent-visual signals, together with 4211 audible signals, including 2672 oral signals and 1539 audible gestures. The following analyses were entirely replicated on a subsample including only signals associated with a gaze-to-the-mother, i.e., hereafter gaze-coordinated signals (see details in Tables S3 and S4 in *Supplementary material*).

- **Unimodal adjustment**: **Are body movements and manual gestures equally sensitive to the mother's attention?**

Infants did not show unimodal adjustment of *body movements*. In other words, they did not produce significantly fewer *body movements* when the mothers did not show visual attention than when they did ([7-10]: Permutation Test, p =0.406, N = 11; [11-14]: Permutation Test, p =0.195, N = 12; [15-20]: Permutation Test, p =0.156, N = 7; Figure 1). However, infants showed unimodal adjustment of *deictic gestures* from [7-10] months of age ([7-10]: Permutation Test, p =0.008, N = 11; [11-14]: Permutation Test, p =0.008, N = 12; [15-20]: Permutation Test, p =0.031, N = 7; Figure 1). Taken separately, proximal deictic gestures were significantly used in unimodal adjustments by 11-to-14-month-old infants only ([7-10]: Permutation Test, p =0.063, N = 11; [11-14]: Permutation Test, p =0.002, N = 12; [15-20]: Permutation Test, p =0.25, N = 7) and distal deictic gestures were not significantly used in unimodal adjustments by infants ([7-10]: Permutation Test, p =0.086, N = 11; [11-14]: Permutation Test, p =1, N = 12; [15-20]: Permutation Test, p =0.125, N = 7).

Gaze-coordinated silent-visual signals showed unimodal adjustment for *body movements* in the 7-to-10-month-old infants only (Permutation Test, p =0.016, N = 11). Gaze-coordinated deictic gestures were no longer significant in 15-to-20-month-old infants. (Permutation Test, p =0.0.062, N = 7; see details in Table S3 in *Supplementary material*). Gaze-coordinated *distal* deictic gestures showed unimodal

adjustment in 11-to-14-month-old infants only (Permutation Test, p =0.008, N = 12).

Figure 1. Distribution of *body movements* and *deictic gestures* produced by individual, age range and maternal visual condition (p-value: *<0.05, **<0.01, ***<0.001). Large black circles represent the mean proportion per subject. Median (horizontal lines), quartiles (boxes), percentiles (2.5% and 97.5%), vertical lines, and outliers (small black circles) are indicated. Differential use patterns are compared from one maternal visual condition to another.

- **Cross modal adjustment: Are audible gestures *with* and *without* object equally sensitive to the mother's attention?**

At [7-10] months of age, infants did not show cross-modal adjustment of audible gestures, whether *with* or *without* objects. In other words, they did not produce significantly more audible gestures *with* and *without* object than silent-visual gestures when the mothers did not show visual attention as compared to when they did (with object: Fisher Test, OR = 0.653, p =0.107; without object: Fisher Test, OR = 0.828, p =1, N = 11; Figure 2). At [11-14] months of age, infants showed cross-modal adjustment of audible gestures only *with* object (with object:

Fisher Test, OR = 0.510, p =0.001; without object: Fisher Test, OR = 1.309, p = 1, N = 12; Figure 2). At [15-20] months of age, the same pattern was found (with object: Fisher Test, OR = 0.114, p <0.001; without object: Fisher Test, OR = 0.288, p =0.055, N = 7; Figure 2).

Figure 2. Mean (+ standard error bar) proportion of audible gestures against silent-visual gestures produced by maternal condition and age range. The proportions are calculated as "number of a given audible signal X divided by the number of this audible signal X plus the number of silent-visual signals". Thus, when the mean proportion is above the dashed line, infants produce more audible signals than silent-visual signals. Cross-modal adjustment occurs when the proportion of audible gestures is higher in the inattentive than in the attentive condition.

When only audible gestures associated with a gaze-to-the-mother were selected, infants did not show any cross-modal adjustment of the audible gestures (see details in Table S4 in *Supplementary material*).

- **Cross-modal adjustment: Are *mouth sounds*, *protophones*, *babblings* and *verbal contents* equally sensitive to the mother's attention?**

At [7-10] months of age, infants did not show any cross-modal adjustment of any oral signals and even produced significantly the opposite pattern with

babblings (mouth sounds: Fisher Test, OR = 0.984, p =1; non-CV vocalisations: Fisher Test, OR = 0.955, p = 1; babblings: Fisher Test, OR = 2.409, p =0.026; verbal contents: Fisher Test, OR = 0, p =1; N = 12; Figure 3). At [11-14] months of age, infants only showed cross-modal adjustment of *babblings* (mouth sounds: Fisher Test, OR = 0.635, p =0.166; non-CV vocalisations: Fisher Test, OR = 0.655, p = 0.095; babblings: Fisher Test, OR = 0.322, p <0.001; verbal contents: Fisher Test, OR = 0, p =1; N = 11; Figure 3). At [15-20] months of age, infants showed a cross-modal adjustment of *mouth sounds, non-CV vocalisations* and *babblings* but not *verbal content* (mouth sounds: Fisher Test, OR = 0.311, p =0.003; non-CV vocalisations: Fisher Test, OR = 0.419, p = 0.031; babblings: Fisher Test, OR = 0.361, p =0.026; verbal contents: Fisher Test, OR = 0.519, p =0.717; N = 7; Figure 3).

Figure 3. Proportions of *mouth sounds*, *non-CV vocalisations*, *babblings* and *verbal contents* against silent-visual gestures produced by individual, age range and maternal visual condition (p-value: *<0.05, **<0.01, ***<0.001). Large black circles represent the mean

proportion per subject. Median (horizontal lines), quartiles (boxes), percentiles (2.5% and 97.5%), vertical lines, and outliers (small black circles) are indicated. The use patterns are compared from one maternal visual condition to another. The proportions are calculated as "number of one oral signal X on the sum of this oral signal X and silent-visual signals". Thus, when the proportion is above the dashed line, infants produce more oral signals than silent-visual gestures and vice-versa. Cross-modal adjustment occurs when the proportion is higher in the inattentive condition than in the attentive one.

When only oral signals associated with a gaze-to-the-mother were selected, infants did not show any cross-modal adjustment of the oral sounds (see details in Table S4 in *Supplementary material*).

Discussion

This study examined the early pragmatic use of audible and visual communicative signals by 7-to-20-month-old infants. Both canonical markers of language development and signals belonging to an extended repertoire of communication were used by infants to adjust to their mother's visual attention. Gaze-coordinated distal deictic gestures (including canonical pointing gestures) were used for unimodal adjustment by 11-to-14-month-old infants, while canonical babbling was used for cross-modal adjustment from 11 months onwards. However, many other signals belonging to the extended communicative repertoire were used for adjusting to maternal attention, hence avoiding communication mismatches. The youngest infants showed *unimodal* adjustment of *body movements*, provided that they were gaze-coordinated. Infants aged 11-14 months showed unimodal adjustment of *proximal deictic gestures* and *cross-modal* adjustment of audible object gestures. Infants aged 15-20 months displayed cross-modal adjustment of mouth sounds and non-CV vocalizations. Gaze-coordinated signals were overall not significantly better adjusted to maternal attention than non-gaze-coordinated signals, except for specific silent-visual signals at certain ages.

Apart from gaze-coordination, (proximal and distal) deictic gestures were more often produced by infants towards attentive than inattentive mothers, from

7 months onwards, i.e. *unimodal adjustment*. When proximal and distal deictic gestures were considered separately, only 11-14-month-old infants still significantly adjusted deictic gestures. Proximal deictic gestures, including giving and showing, and gaze-coordinated distal deictic gestures were produced more towards visually attentive than inattentive mothers. This may come from the higher statistical power in this age group where deictic gestures have just emerged and are prominent among the overall number of silent-visual signals produced ([7-10]: 44 *deictic* gestures out of 225 silent-visual signals produced, 19,6 %; [11-14]: 138 *deictic* gestures out of 203 silent-visual signals produced, 68 %; [15-20]: 50 *deictic* gestures out of 78 silent-visual signals produced, 64,1 %). To our knowledge, this study is the first to include *proximal* deictic gestures and *body movements* in the investigation of attention-sensitive communication in preverbal infants. There is a dearth of studies on this topic in human infants, compared to non-human primates, and most studies have focused on (i) quantifying pointing gestures, vocalisations and pointing-vocalisation coupling (i.e. audio-visual combinations) only and (ii) exclusively gaze-coordinated signals (Gros-Louis & Wu, 2012; Igualada et al., 2015; Wu & Gros-Louis, 2014, 2017). These studies showed a unimodal adjustment of distal deictic gestures from 12 months of age. However, more recent studies using an ethological coding with an extended repertoire found the onset of *unimodal* adjustment at 7-8 months of age (Dafreville et al., *submitted*; Rodrigues et al., 2021). Indeed, 7-to-10-month-old infants avoided communication mismatches towards inattentive mothers by making gaze-coordinated body movements. Previous studies have provided evidence of the relevance of using body movements to delineate the early forms of interactive exchanges in infants. Some studies showed that 5-to-9-month-old infants reduced their body movements when the mothers were talking or singing in an experimental context of *face-to-face*, hence delineating episodes of interaction (Nakata & Trehub, 2004; Scola et al., 2015). Here, the body movements were used more frequently towards attentive than inattentive mothers, provided that they were gaze-coordinated. This result contrasts with a recent experiment in 6-month-old infants who showed rudimentary forms of unimodal

adjustment towards mothers who were both visually attentive and interactive compared to mothers who were neither visually attentive nor interactive (Bourjade et al., *submitted*). The ecological settings of the present study may explain this earlier onset.

One intriguing result of this study is the transitory adjustment to maternal attention of some silent-visual signals at a certain age. Our results also suggest that this pattern is not independent from gaze-coordination. Gaze-coordinated *body movements* in young infants are significantly adjusted to maternal attention because it is one of the primary ways to establish interaction very early (Beebe et al., 2016; Stern, 1999; Stern et al., 1977). This adjustment does not last in older infants. Then, proximal deictic gestures and gaze-coordinated distal deictic gestures are significantly adjusted to maternal attention in 11-14-month-olds, together with audible object gestures, whereas the older infants tend to show cross-modal *vocal* adjustment. Here it appears that adjustments to maternal attention are better reached with signals that are prominent in infants' communication at certain ages. Does this mean that older infants no longer point preferentially towards attentive than towards inattentive mothers? Competing hypotheses can be used to explain this pattern. First, infants may decrease the relative quantity of deictic gestures at the onset of multimodal utterances and an increasing vocal communication is expected at 15-20 months of age ([7-10]: 926 oral sounds against 225 deictic gestures, 95,46 %; [11-14]: 1010 oral sounds against 203 deictic gestures, 87,98 %; [15-20]: 736 oral sounds against 78 deictic gestures, 93,64 % ; e.g. Guidetti, 2005; Murillo & Casla, 2020). Second, many recent endeavours have provided evidence of internal (personal) and interpersonal rhythmicity in interactions (Beebe et al., 2016; Devouche & Gratier, 2001; Northrup & Iverson, 2020a). We suggest that gazing at the recipient may have its own rhythmicity that may change over the course of development from 12 months of age. Gaze may hence be coordinated at the scale of the whole 'conversation' rather than at the scale of each signal (Rossano, 2013) and therefore might have escaped our coding.

Several studies examined developmental trajectories for gaze-coordinated signals and found different trajectories according to the signal type and modality (D'Odorico et al., 1997; Donnellan et al., 2019; Franco & Butterworth, 1996). Gaze precedes vocalizations in 12-month-old infants, while it occurs after vocalizing in 20-month-old infants (D'odorico et al 1997). In contrast, 12-to-14-month-old infants do visual checking during or after producing a pointing gesture rather than before, while it is the reverse from 16 months of age (Franco & Butterworth, 1996; Liszkowski et al., 2004). The methodologies used to sample gaze-coordinated signals are highly variable, with the result that there does not appear to be a consistent reliable criterion to identify the illocutionary use of a signal (Liszkowski et al., 2008). Here, it is worth noting that none of the audible signals were better adjusted to maternal attention when gaze-coordinated. This is an important result on attention-sensitive communication because the audible sensory modality can still be perceived regardless of prior visual attention by the recipient. In contrast, our infants showed better unimodal adjustment of silent-visual body movements or distal deictic gestures at 7-10 and 11-14 months of age respectively; namely when they deliberately addressed those signals to visually attentive mothers. In sum, the use and refinement of gaze-coordinated signals may depend on different variables - interaction context, sensory modality, nature of the signal, and recipient response - and takes time to be fully mastered. That is why it may be relevant to consider other criteria than gaze-coordination to capture the developmental trajectory of the pragmatic usage of signals driving the transition to word use (Liszkowski, 2006; Liszkowski et al., 2008).

Furthermore, the integration of an object in the signalling may play a part in the development of adjustment capacities. Indeed, infants displayed *cross-modal* adjustment of signals integrating an object (i.e. audible gestures *with* object) at 11-14 months of age. As mentioned earlier, some authors (Liszkowski & Rüther, 2021; Rodríguez et al., 2015) suggested that the integration of an object in the social

activity (e.g. *giving* and *showing* gestures) may constitute a crucial step leading to object reference, conventionally approximated by the referential use of *pointing* gestures at 12 months of age (Liszkowski, 2006; Rüther & Liszkowski, 2020). Some recent studies support this idea by showing that the production of *proximal* deictic gestures (i.e., giving or showing objects) predicts the production of *distal* (canonical) pointing gestures (Cameron-Faulkner et al., 2015) and linguistic skills at 18 months of age (Choi et al., 2021), and are themselves predicted by the production of rhythmical movements *with* object combined with vocalisations at 9 months (Murillo et al., 2019). Regarding the hypothesis whereby the integration of an object in the interaction may predict subsequent prelinguistic milestones of language development (Liszkowski & Rüther, 2021; Rodríguez et al., 2015), our results highlight the importance of considering the multimodal facet of the signals produced with objects.

Cross-modal adjustment was also reached with vocalizations and oral sounds. From an early stage, *babblings* stand out from the other oral sounds. Infants in fact showed a differential use of *babblings* from 7-10 months of age but in the direction opposite to our hypothesis of *cross-modal* adjustment, since they deployed fewer *babblings* than silent-visual signals when the mothers did not show visual attention compared to when they did. However, from 11 months of age, infants showed a drastic cross-modal adjustment with babblings, while mouth sounds and non-CV vocalizations were significantly preferred to address a visually inattentive mother from 15 months of age. Interestingly, although *babblings* emerge relatively late as compared to the other prelinguistic oral sounds (Esteve-Gibert & Guellaï, 2018; Oller et al., 1999; Stoel-Gammon, 2011), they are sensitive to visual attention earlier than the other oral sounds. One explanation for this pattern is that non-canonical prelinguistic vocalizations are mostly endogenous and therefore independent of the presence, attention and interaction of the recipient (Harold & Barlow, 2013; Long et al., 2020; Oller et al., 2021). The onset of babbling at 7-8 months of age generates a higher responsiveness from caregivers ,which could

motivate the learning of the perlocutionary effects of these signals (Bloom & Beaumont, 1993; Bloom & Lo, 1990; Gros-Louis et al., 2014; Locke, 2006). This may be why the youngest infants produced more babblings towards attentive mothers. We suggest that as babbling becomes prominent in the repertoire and generates consistent parental responses, these responses represent multiple learning opportunities for developing the illocutionary value of babbling that may then be used strategically by 11-14-month-old infants to address visually inattentive mothers. Furthermore, many studies showed that *babblings* consitute a good predictor of linguistic production (Donnellan et al., 2019; Jensen et al., 1988; Oller et al., 1998; Oller et al., 1999; Stoel-Gammon, 2011; Stolt et al., 2012), supporting the transition to locutionary communication. In sum, our results support the idea that the use of canonical *babblings* may be more subject to social shaping during development than the other prelinguistic oral sounds (Bloom & Beaumont, 1993; Bloom & Lo, 1990; Gros-Louis et al., 2014; Locke, 2006). Verbal content was virtually absent in our study, which explains why it does not adjust to the mother's attention. However, mouth sounds and non-CV vocalizations were used for cross-modal adjustment in 15-20-month-old infants onwards. We suggest that once babbling is used for cross-modal adjustment, infants extend this strategic use to other vocalizations and to mouth sounds. Further investigation is needed to explore the differential use of vocal signals and their integration in gesture-vocal coupling according to the audience's attention.

Conclusion

To our knowledge, this is the first study investigating attention-sensitive communication in preverbal infants based on an extended repertoire of signals. Unimodal adjustment (7 months onwards) preceded crossmodal adjustment (11 months onwards) and both were reached by infants through the use of distal deictic gestures (including canonical pointing gestures) and babbling, but also through the pragmatic use of other silent-visual and audible signals. Gaze-coordinated signals

were better adjusted to maternal attention for silent-visual signals, in a transitory way. In contrast, audible signals were preferred to address visually inattentive mothers independently of gaze, which remains consistent with their mother's (auditory) attention. Overall, these results indicate that attention-sensitive communication is relevant to the development of early pragmatic skills and that the illocutionary use of signals may be more reliably approximated by this capacity than by gaze-coordination with signals. Future studies should investigate the predictive potential of *unimodal* and *cross-modal* adjustments.

References

Aureli, T., Spinelli, M., Perucchini, P., & Odorico, L. D. (2017). *The Pointing – Vocal Coupling Progression in the First Half of the Second Year of Life*. *22*(6), 801–818. https://doi.org/10.1111/infa.12181

Bates, E., Camaioni, L., & Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill-Palmer Quarterly of Behavior and Development*, *21*(3), 205–226.

Beebe, B., Messinger, D., Bahrick, L. E., Margolis, A., Buck, K. A., & Chen, H. (2016). A Systems View of Mother–Infant Face-to-Face Communication. *Developmental Psychology*, *52*(4), 556–571. https://doi.org/10.1037/a0040085.supp

Beuker, K. T., Rommelse, N. N. J., Donders, R., & Buitelaar, J. K. (2013). Development of early communication skills in the first two years of life. *Infant Behavior and Development*, *36*(1), 71–83. https://doi.org/10.1016/j.infbeh.2012.11.001

Bloom, K., & Beaumont, S. (1993). Adults' attributions of intentionality to vocalizing infants. *The Grants Register 2022*, *13*, 235–247. https://doi.org/10.1057/978-1-349-96042-2_124111

Bloom, K., & Lo, E. (1990). Adult perceptions of vocalizing infants. *Infant Behavior and Development*, *13*(2), 209–219. https://doi.org/10.1016/0163-6383(90)90031-3

Bourjade, M., Cochet, H., Molesti, S., & Guidetti, M. (2020). Is Conceptual Diversity an Advantage for Scientific Inquiry ? A Case Study on the Concept of ' Gesture ' in Comparative Psychology. *Integrative Psychological and Behavioral Science*, *54*, 805–832.

Bourjade, M., Dafreville, M., Scola, C., & Jover, M. (n.d.). Six-month-old infant's communication in a comparative perspective: how do maternal attention and interaction matter? *(Submitted)*.

Byrne, R. W., Cartmill, E., Genty, E., Graham, K. E., Hobaiter, C., & Tanner, J. (2017). Great ape gestures: intentional communication with a rich set of innate signals. *Animal Cognition*, *20*(4), 755–769. https://doi.org/10.1007/s10071-017-1096-4

Camaioni, L. (1997). The Emergence of Intentional Communication in Ontogeny, Phylogeny, and Pathology. *European Psychologist*, *2*(3), 216–225. https://doi.org/10.1027/1016-9040.2.3.216

Camaioni, L., Perucchini, P., Bellagamba, F., & Colonnesi, C. (2004). The Role of Declarative Pointing in Developing a Theory of Mind. *Infancy*, *5*(3), 291–308. Retrieved from https://doi.org/10.1207/s15327078in0503_3

Cameron-Faulkner, T., Theakston, A., Lieven, E., & Tomasello, M. (2015). The relationship between infant holdout and gives, and pointing. *Infancy*, *20*(5), 576–586. https://doi.org/10.1111/infa.12085

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social Cognition , Joint Attention , and Communicative Competence from 9 to 15 Months of Age. In *Monographs of the Society for Research in Child Development* (Vol. 63, p. (Serial No.255)).

Choi, B., Wei, R., & Rowe, M. L. (2021). Show, Give, and Point Gestures Across Infancy Differentially Predict Language Development. *Developmental Psychology*, *57*(6), 851–862. https://doi.org/10.1037/dev0001195

Cochet, H., & Byrne, R. W. (2016). Infant Behavior and Development Full length article Communication in the second and third year of life : Relationships between nonverbal social skills and language. *Infant Behavior and Development*, *44*, 189–198. https://doi.org/10.1016/j.infbeh.2016.07.003

Colonnesi, C., Stams, G. J. J. M., Koster, I., & Noom, M. J. (2010). The relation between

pointing and language development: A meta-analysis. *Developmental Review*, *30*(4), 352–366. https://doi.org/10.1016/j.dr.2010.10.001

Csibra, G., & Volein, Á. (2008). Infants can infer the presence of hidden objects from referential gaze information. *British Journal of Developmental Psychology*, *26*(1), 1–11. https://doi.org/10.1348/026151007X185987

D'Odorico, L., Cassibba, R., & Salerni, N. (1997). Temporal Relationships between Gaze and Vocal Behavior in Prelinguistic and Linguistic Communication. *Journal of Psycholinguistic Research*, *26*(5), 539–556. https://doi.org/10.1023/A:1025027830861

Dafreville, M., Guidetti, M., & Bourjade, M. (n.d.). Attention-sensitive signaling by 7-to-20-month-old infants in a comparative perspective. *(Submitted)*.

Dafreville, M., Hobaiter, C., Guidetti, M., Sillam-Dussès, D., & Bourjade, M. (2021). Sensitivity to the communicative partner's attentional state: A developmental study on mother–infant dyads in wild chimpanzees (Pan troglodytes schweinfurthii). *American Journal of Primatology*, *83*(12). https://doi.org/10.1002/ajp.23339

Dennett, D. C. (1983). Intentional systems in cognitive ethology: The "Panglossian paradigm" defended. *Behavioral and Brain Sciences*, *6*(3), 343–355. https://doi.org/10.1017/S0140525X00016393

Devouche, E., & Gratier, M. (2001). Microanalyse du rythme dans les échanges vocaux et gestuels entre la mère et son bébé de 10 semaines. *Devenir*, *13*(2), 55–82. https://doi.org/10.3917/dev.012.0055

Donnellan, E., Bannard, C., Mcgillion, M. L., Slocombe, K. E., & Matthews, D. (2019). *Infants ' intentionally communicative vocalizations elicit responses from caregivers and are the best predictors of the transition to language : A longitudinal investigation of infants ' vocalizations , gestures and word production*. (March), 1–21. https://doi.org/10.1111/desc.12843

Esteve-Gibert, N., & Guellaï, B. (2018). Prosody in the auditory and visual domains: A developmental perspective. *Frontiers in Psychology*, *9*(March), 1–10. https://doi.org/10.3389/fpsyg.2018.00338

Falk, D. (2004). Prelinguistic evolution in early hominins: Whence motherese? *Behavioral and Brain Sciences*, *27*(04). https://doi.org/10.1017/S0140525X04000111

Franco, F., & Butterworth, G. (1996). Pointing and social awareness: declaring and requesting in the second year. *Journal of Child Language*, *23*(02), 307–336. https://doi.org/10.1017/s0305000900008813

Gros-Louis, J., West, M. J., & King, A. P. (2014). Maternal responsiveness and the development of directed vocalizing in social interactions. *Infancy*, *19*(4), 385–408. https://doi.org/10.1111/infa.12054

Gros-Louis, J., & Wu, Z. (2012). Twelve-month-olds' vocal production during pointing in naturalistic interactions: Sensitivity to parents' attention and responses. *Infant Behavior and Development*, *35*(4), 773–778. https://doi.org/10.1016/j.infbeh.2012.07.016

Guidetti, M. (2005). Yes or no ? How young French children combine gestures and speech to agree and refuse *. *Journal of Child Language*, *32*, 911–924. https://doi.org/10.1017/S0305000905007038

Guidetti, M., Fibigerova, K., & Colletta, J.-M. (2014). Gestures and multimodal development. In M. Seyfeddinipur & M. Gullberg (Eds.), *From Gesture in Conversation to Visible Action as Utterance* (pp. 351–370). https://doi.org/10.1075/z.188

Harold, M. P., & Barlow, S. M. (2013). Effects of environmental stimulation on infant vocalizations and orofacial dynamics at the onset of canonical babbling. *Infant Behavior and Development*, *36*(1), 84–93. https://doi.org/10.1016/j.infbeh.2012.10.001

Igualada, A., Bosch, L., & Prieto, P. (2015). Language development at 18 months is related to multimodal communicative strategies at 12 months. *Infant Behavior and Development*, *39*, 42–52. https://doi.org/10.1016/j.infbeh.2015.02.004

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science*, *16*(5), 367–371. https://doi.org/10.1111/j.0956-7976.2005.01542.x

Jensen, T. S., Bøggild-Andersen, B., Schmidt, J., Ankerhus, J., & Hansen, E. (1988). Perinatal Risk Factors and First-Year Vocalizations: Influence on Preschool Language and Motor Performance. *Developmental Medicine & Child Neurology*, *30*(2), 153–161. https://doi.org/10.1111/j.1469-8749.1988.tb04746.x

Kita, S. (2003). Pointing: A foundational building block of human communication. In *Pointing* (Psychology, pp. 9–16).

Lameira, A. R., & Call, J. (2018). Time-space-displaced responses in the orangutan vocal system. *Science Advances*, *4*(11), 1–8. https://doi.org/10.1126/sciadv.aau3401

Liszkowski, U. (2006). Infant pointing at 12 months: communicative goals, motives, and social-cognitive abilities. In *Roots of human sociality* (Routledge, pp. 153–178).

Liszkowski, U., Albrecht, K., Carpenter, M., & Tomasello, M. (2008). Infants ' visual and auditory communication when a partner is or is not visually attending. *Infant Behavior & Development*, *31*, 157–167. https://doi.org/10.1016/j.infbeh.2007.10.011

Liszkowski, U., Brown, P., Callaghan, T., Takada, A., & de Vos, C. (2012). A Prelinguistic Gestural Universal of Human Communication. *Cognitive Science*, *36*(4), 698–713. https://doi.org/10.1111/j.1551-6709.2011.01228.x

Liszkowski, U., Carpenter, M., Henning, A., Striano, T., & Tomasello, M. (2004). Twelve-month-olds point to share attention and interest. *Developmental Science*, *7*(3), 297–307. https://doi.org/10.1111/j.1467-7687.2004.00349.x

Liszkowski, U., Carpenter, M., Striano, T., & Tomasello, M. (2006). 12- and 18-Month-Olds Point To Provide Information for Others. *Journal of Cognition and Development*, *7*(2), 173–187. https://doi.org/10.1207/s15327647jcd0702_2

Liszkowski, U., & Rüther, J. (2021). Ontogenetic origins of infant pointing. *The Oxford Handbook of Human Symbolic Evolution*. https://doi.org/10.1093/oxfordhb/9780198813781.013.31

Locke, J. L. (2006). Parental selection of vocal behavior. *Human Nature*, *17*(2), 155–168. https://doi.org/10.1007/s12110-006-1015-x

Locke, J. L. (2017). Development of the Capacity for Spoken Language. In P. Fletcher & B. MacWhinney (Eds.), *The Handbook of Child Language*. https://doi.org/10.1111/b.9780631203124.1996.x

Locke, J. L., & Bogin, B. (2006). Language and life history: A new perspective on the development and evolution of human language. *Behavioral and Brain Sciences*, *29*(3), 259–280. https://doi.org/10.1017/S0140525X0600906X

Long, H. L., Bowman, D. D., Yoo, H., Burkhardt-Reed, M. M., Bene, E. R., & Kimbrough Oller, D. (2020). Social and endogenous infant vocalizations. *PLoS ONE*, *15*(8 August 2020), 1–18. https://doi.org/10.1371/journal.pone.0224956

Mastin, J. D., & Vogt, P. (2016). Infant engagement and early vocabulary development: A naturalistic observation study of Mozambican infants from 1;1 to 2;1. *Journal of Child Language*, *43*(2), 235–264. https://doi.org/10.1017/S0305000915000148

Mccune, L., & Zlatev, J. (2015). Dynamic systems in semiotic development: The transition toreference. *Cognitive Development*, *36*, 10. https://doi.org/10.1016/j.cogdev.2015.09.010

Mehr, S. A., & Krasnow, M. M. (2017). Parent-offspring conflict and the evolution of infant-directed song. *Evolution and Human Behavior*, *38*(5), 674–684.

https://doi.org/10.1016/j.evolhumbehav.2016.12.005
Murillo, E., & Casla, M. (2020). Multimodal representational gestures in the transition to multi-word productions. *Infancy*, *26*(1), 104–122. https://doi.org/10.1111/infa.12375
Murillo, E., Montero, I., & Casla, M. (2019). *Rhythmic movements with objects at 9 months are related to proximal deictic gestures at 12 months .*
Murillo, E., Montero, I., & Casla, M. (2021). On the Multimodal Path to Language: The Relationship Between Rhythmic Movements and Deictic Gestures at the End of the First Year. *Frontiers in Psychology*, *12*(February), 1–8. https://doi.org/10.3389/fpsyg.2021.616812
Nakata, T., & Trehub, S. E. (2004). Infants' responsiveness to maternal speech and singing. *Infant Behavior and Development*, *27*(4), 455–464. https://doi.org/10.1016/j.infbeh.2004.03.002
Northrup, J. B., & Iverson, J. M. (2020). *Multimodal coordination of vocal and gaze behavior in mother-infant dyads across the first year of life*. https://doi.org/10.1111/infa.12369
Oller, D. K. (2000). *The Emergence of the Speech Capacity*.
Oller, D. K., Buder, E. H., Ramsdell, H. L., Warlaumont, A. S., Chorna, L., & Bakeman, R. (2013). Functional flexibility of infant vocalization and the emergence of language. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(16), 6318–6323. https://doi.org/10.1073/pnas.1300337110
Oller, D. K., Eilers, R. E., Neal, A. R., & Cobo-Lewis, A. B. (1998). Late onset canonical babbling: A possible early marker of abnormal development. *American Journal on Mental Retardation*, Vol. 103, pp. 249–263. https://doi.org/10.1352/0895-8017(1998)103<0249:LOCBAP>2.0.CO;2
Oller, D. K., Eilers, R. E., Neal, A. R., & Schwartz, H. K. (1999). Precursors to speech in infancy. *Journal of Communication Disorders*, *32*(4), 223–245. https://doi.org/10.1016/s0021-9924(99)00013-1
Oller, D. K., Griebel, U., & Warlaumont, A. S. (2016). Vocal Development as a Guide to Modeling the Evolution of Language. *Topics in Cognitive Science*, *8*(2), 382–392. https://doi.org/10.1111/tops.12198
Oller, D. K., Ramsay, G., Bene, E., Long, H. L., & Griebel, U. (2021). Protophones, the precursors to speech, dominate the human infant vocal landscape. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1836). https://doi.org/10.1098/rstb.2020.0255
Pearce, N. (2016). Analysis of matched case-control studies. *BMJ (Online)*, *352*, 1–4. https://doi.org/10.1136/bmj.i969
Reynolds Losin, E. A., Russell, J. L., Freeman, H., Meguerditchian, A., & Hopkins, W. D. (2008). Left hemisphere specialization for oro-facial movements of learned vocal signals by captive chimpanzees. *PLoS ONE*, *3*(6), 1–7. https://doi.org/10.1371/journal.pone.0002529
Rodrigues, E. D., Marôco, J., & Frota, S. (2021). Communicative gestures in 7–12-month infants: A phylogenetic comparative approach. *Infant and Child Development*, *30*(5), 1–11. https://doi.org/10.1002/icd.2262
Rodríguez, C., Moreno-Núñez, A., Basilio, M., & Sosa, N. (2015). Ostensive gestures come first: Their role in the beginning of shared reference. *Cognitive Development*, *36*, 142–149. https://doi.org/10.1016/j.cogdev.2015.09.005
Rossano, F. (2013). Gaze in Conversation. In *The handbook of conversation analysis* (p. 308).
Rüther, J., & Liszkowski, U. (2020). Ontogenetic Emergence of Cognitive Reference Comprehension. *Cognitive Science*, *44*(7). https://doi.org/10.1111/cogs.12869
Salomo, D., & Liszkowski, U. (2013). *Sociocultural Settings In fl uence the Emergence of Prelinguistic Deictic Gestures*. *84*(4), 1296–1307. https://doi.org/10.1111/cdev.12026

Scola, C., Bourjade, M., & Jover, M. (2015). Social interaction is associated with changes in infants' motor activity. *Socioaffective Neuroscience & Psychology*, *5*(1), 28256. https://doi.org/10.3402/snp.v5.28256

Siegal, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences (2nd ed)* (McGraw Hil). Boston.

Stern, D. (1999). Vitality contours: The temporal contour of feelings as a basic unit for constructing the infant's social experience. *Early Social Cognition: Understanding Others in the First Months of Life*, 67–80.

Stern, D., Beebe, B., Jaffe, J., & Benett, S. (1977). The infant's stimulus world during social interaction: A study of caregiver behaviors with particular reference to repetition and timing. *Studies in Mother-Infant Interaction*, 177–202.

Stoel-Gammon, C. (2011). Relationships between lexical and phonological development in young children. *Journal of Child Language*, *38*(1), 1–34. https://doi.org/10.1017/S0305000910000425

Stolt, S., Lehtonen, L., Haataja, L., & Lapinleimu, H. (2012). Development and predictive value of early vocalizations in very-low-birth-weight children: A longitudinal study. *Clinical Linguistics and Phonetics*, *26*(5), 414–427. https://doi.org/10.3109/02699206.2011.648365

Tomasello, M., & Call, J. (2018). Thirty years of great ape gestures. *Animal Cognition*, (0123456789). https://doi.org/10.1007/s10071-018-1167-1

Wallez, C., Schaeffer, J., Meguerditchian, A., Vauclair, J., Schapiro, S. J., & Hopkins, W. D. (2012). Contrast of hemispheric lateralization for oro-facial movements between learned attention-getting sounds and species-typical vocalizations in chimpanzees: Extension in a second colony. *Brain and Language*, *123*(1), 75–79. https://doi.org/10.1016/j.bandl.2012.07.002

Wu, Z., & Gros-Louis, J. (2014). Infants' prelinguistic communicative acts and maternal responses: Relations to linguistic development. *First Language*, *34*(1), 72–90. https://doi.org/10.1177/0142723714521925

Wu, Z., & Gros-Louis, J. (2017). The Value of Vocalizing: 10-Month-Olds' Vocal Usage Relates to Language Outcomes at 15 Months. *Infancy*, *22*(1), 23–41. https://doi.org/10.1111/infa.12150

Supplementary material

Beyond pointing gestures: Attention-sensitive communication in 7-to-20-month-old infants develops pragmatic skills

Mawa Dafreville[1*], Michèle Guidetti [1], Marie Bourjade[1,2]

[1] CLLE, Université de Toulouse, CNRS, Toulouse, France
[2] Institut Universitaire de France

TABLE S1 Repertoire of human 7-to-20-month-old infants' multimodal communication.

Silent-visual signals					
Channel of production	**Category of communicative signals**	**Signal**	**Definition**	***Onset***	***Offset***
Gestural	Body movement	Circular movement of the limbs (VAM)	Repetitive movements of the arms and/or legs and/or hands and/or fingers performed two or more times. Included: the infant holds a silent object during it; when the infant swings his/her arms while walking. Excluded : the infant holds a noisy object during it (code AO, see 3.1.1)	It starts at the image on which a movement is visible.	A new occurrence is recorded after a break longer than 1 second.
		Synchronised Trunk/Members movement (VSTM)	The infant leans back against the chair or against his/her mother and raises the pelvis up while stretching the trunk and the legs. The arms can be raised up or not. Included: when the infant is sitting on his/her mother's lap in an embrace	It starts when the trunk is stretching.	A new occurrence is recorded after a break longer than 1 second.

		position for example and leans back against her.		
	Swagger (VS)	The infant moves his/her trunk along an axis (back and forth, from side to side, or up and down). An occurrence is recorded if at least a two-way movement is achieved (e.g. one movement back followed by one movement forth).	It starts when the movement begins.	When the infant stops moving.
Proximal deictic gestures	Giving (VGO)	The infant holds the object with at least one hand towards the partner or he/she makes the object accessible to the recipient.	It starts when the infant begins his/her movement towards the object.	It ends when the infant no longer has the object in his/her hand.
	Receiving (VRO)	The infant takes/touches, with at least one of his/her hands, the object that the partner has handed or made accessible to him/her. Included: when the infant holds food that is not intended for him/her (e.g. when the partner is eating). Excluded: when it is included in the feeding (for example when the mother gives some pasta) if the infant was not holding it before.	It starts when the infant touches the object that the partner is handing to him/her.	It ends 1 second after the first contact between the infant and the object.
	Showing (VSO)	The infant without outstretched arm, holds in his/her hands an object in the direction of the partner, looking at the partner for more than 1 second while keeping the posture. Included: pointing towards himself/herself or the partner's body (dyadic pattern). For example, when the mother asks him/her to show her a part of his/her body ("where's your nose?").	It starts when the child looks at his/her partner with the object in his/her hands.	It ends when the child relaxes his/her arms or changes posture.

	Distal deictic gestures	Whole-hand pointing gesture (VWHPO)	The infant points with the whole hand and fingers extended in the direction of the object or another individual.	It starts when the infant begins his/her movement towards the object/individual	It ends when the infant releases his/her arm (even a little) or changes the target
		Index-pointing gesture (VIPO)	The infant points with his/her index finger towards an external object or individual (triadic pattern).	It starts when the infant begins his/her movement towards the object/individual.	It ends when the infant releases his/her arm or changes the target.
		Reaching attempt (VRAO)	The infant tries to reach out for one (or more) object(s)/or another individual out of range, without pointing, but leaning the trunk towards the object in a time window of 1 second before or after stretching his/her arm with his/her hand open. The infant can repeatedly open and close his/her fingers in the direction of the object (e.g., grasping movements). Excluded: when the infant makes the same movement but with an object in his/her hand already.	It starts when the infant begins his/her movement towards the object/individual.	It ends when the infant releases his/her arm or changes the target.

Audible Signals

Channel of production	Category of communicative signals	Signal	Definition	*Onset*	*Offset*
Oral	Mouth sounds	Mouth sounds (VOMS)	Any non-voiced sound resulting from the air breathed in and out, without vibrating the vocal folds (e.g. sigh, tongue clicking, raspberry).	It starts when the sound is first audible.	It ends only if the mother talks to her baby while the sound is stopped or

		Included: loud breathing sound that lasts less than 5 seconds. Excluded: if the infant has food in his/her mouth while producing mouth sounds (e.g. while breastfeeding).		after 1 second silent or if another type of oral sound replaces it (no acoustic gestures) while the sound is stopped
Non-CV vocalisations	Vowel-like sounds (VOP)	Vowel-like sounds due to the vibration of the vocal folds, as well as squeals and growls. Excluded: vegetative sounds (e.g., sneezes or burps).	It starts when the sound is first audible.	It ends only if the mother talks to her baby while the sound is stopped or after 1 second silent or if another type of oral sound replaces it (no acoustic gestures) while the sound is stopped.
	Grunt (VOG)	The infant produces a vowel sound on breathing out with a rhythmic pattern where the impulse is on a strong beat and the resulting sound on a weak beat.	It starts when the sound is audible.	It ends only if the mother talks to her baby while the sound is stopped or after 1 second silent or if another type of oral sound replaces it (no acoustic gestures) while the sound is stopped.
Babbling	Babbling (VOB)	The infant emits short or continuous sounds, composed of vowels and consonants forming isolated or repeated syllables, from the vibration of the vocal folds and the opening and closing of the mouth.	It starts when the sound is audible.	It ends only if the mother talks to her baby while the sound is stopped or after 1 second silent or if another type of oral sound replaces

					it (no acoustic gestures) while the sound is stopped.
	Verbal content	Pseudo-word (VONW)	The infant does not clearly say a word but utters a word which sounds like it. We recorded it only when the parent repeated the word to confirm it (e.g. when the child said, "Caaaaag" and his/her parent needs to confirm by asking him/her, "what do you want? Do you want some cake?").	It starts when the infant pronounces the first syllable of the pseudo-word.	It ends at the end of the pseudo-word.
		Word (VOW)	The infant says a single word of one of his/her parent languages that the parent understands.	It starts when the infant pronounces the first syllable of the word.	It ends at the end of the word.
		The infant clearly says a clause comprising at least subject and an object that the parent understands.	It starts when the child pronounces the first syllable of the first word of the clause.	A new clause is recorded for each association of subject, verb and object.	The infant clearly says a clause comprising at least subject and an object that the parent understands.
Gestural	With an object	Sound made with an object (AO)	Any sound produced by an action in the environment with an object in the hands (e.g., hit, scratch, scrape). Included: if the infant pulls on a tree branch that makes a noise. Excluded: Oral sounds or if the sound results in a movement made by the mother (for example, she grasps her baby's arm in which he/she holds a maracas); Sounds made by an electronic toy.	It starts when the sound is audible.	It ends only if the mother talks to her baby while the sound is stopped or after 1 second silent or if another type of acoustic gesture replaces it (excluded : non-voiced sounds and vocalizations) while the sound is stopped.

Sound made without an object (AWO)	Any sound produced by an action on the environment without an object (e.g., hit, scratch, scrape) or by a self-directed action (e.g., clap hands) and the baby is not moving in the room. Included: noise stirring the water (i.e. playing in a stream) Excluded: oral sounds	It starts when the sound is audible.	It ends only if the mother talks to her baby while the sound is stopped or after 1 second silent or if another type of acoustic gesture replaces it (excluded : non-voiced sounds and vocalizations) while the sound is stopped.	Sound made without an object (AWO)

All our analyses were replicated on a **smaller data set consisting of gaze-coordinated signals only** so as to give an indication of the patterns of gaze-coordinated signal use. Here, we consider a *gaze-coordinated signal* as being a signal associated to a gaze towards the mother (at least once in a window from 1 second before the onset of the signal to 1 second after the offset).

As mentioned in the main article, because the sample size was medium (30 mother-infant dyads) and the data did not follow a normal distribution, we used non-parametric statistics to test defined hypotheses (Siegal & Castellan, 1988), including Bonferroni corrections in cases of multiple comparisons. One-sample permutation tests were performed to determine if the infants produced significantly fewer silent-visual signals when the mothers did not show visual attention than when they did, i.e. *unimodal adjustment* (as in Dafreville et al., 2021). Fisher's exact tests on paired data (Pearce, 2016) were performed to determine if the infants produced significantly more audible- or- contact signals than silent-visual gestures when the mothers did not show visual attention than when they did, i.e. *cross-modal adjustment* (as in Dafreville et al., 2021). All tests were conducted using R v3.6.1 software (http://cran.r-project.org) with p-values equal to or lower than 0.05 required for significance. All statistical tests were two- tailed.

The sample of gaze-coordinated signals consisted of 1576 audible signals, including 938 oral signals and 354 audible gestures, and 284 silent-visual signals.

Table S2. Gaze-coordinated signals given the mother's visual attention

Age Range	Modality			Attentive	Inattentive	Total
7-10				**460**	**111**	**571**
	AudibleS			**366**	**85**	**451**
		(Gestural)	Audible gestures *with* object	50	15	65
			Audible gestures *without* object	34	14	48
		(Oral)	Mouth sounds	45	11	56
			Non-CV vocalisations	215	45	260
			Babblings	22	0	22
			Verbal content	0	0	0
	VisualS			**94**	**26**	**120**
		(Gestural)	Body movements	68	25	93
			Deictic gestures	26	1	27
			• *Proximal*	*13*	*1*	*14*
			• *Distal*	*13*	*0*	*13*
11-14				**506**	**126**	**632**
	AudibleS			**403**	**111**	**514**
		(Gestural)	Audible gestures *with* object	98	31	129
			Audible gestures *without* object	30	5	35
		(Oral)	Mouth sounds	53	10	63
			Non-CV vocalisations	188	54	242
			Babblings	34	11	45
			Verbal content	0	0	0
	VisualS			**103**	**15**	**118**
		(Gestural)	Body movements	35	6	41
			Deictic gestures	68	9	77
			• *Proximal*	*46*	*7*	*53*
			• *Distal*	*22*	*2*	*24*
15-20				**336**	**37**	**373**
	AudibleS			**293**	**34**	**327**
		(Gestural)	Audible gestures *with* object	55	10	65
			Audible gestures *without* object	11	1	12
		(Oral)	Mouth sounds	25	2	27
			Non-CV vocalisations	135	17	152
			Babblings	43	3	46
			Verbal content	24	1	25
	VisualS			**43**	**3**	**46**
		(Gestural)	Body movements	11	2	13
			Deictic gestures	32	1	33
			• *Proximal*	*15*	*1*	*16*
			• *Distal*	*17*	*0*	*17*
			Total	**1302**	**274**	**1576**

Table S3. Summary of the permutation tests on gaze-coordinated silent-visual signals for *unimodal* adjustment. Significant p-values are indicated in bold and when the significance differs from the initial ones mentioned in the article, we have underlined them.

Permutation tests	**7-10** (N=11)	**11-14** (N=12)	**15-20** (N=7)
Body movements	**p= 0.016**	p= 0.113	p= 0.312
Deictic gestures	**p= 0.016**	**p= 0.002**	p= 0.062
• *Proximal*	p= 0.125	**p= 0.005**	p= 0.250
• *Distal*	p= 0.125	**p= 0.008**	p= 0.125

Table S4. Summary of p-values and odds ratio (OR) gaze-coordinated signals for *cross-modal* adjustment. Significant p-values are indicated in bold and when the significance differs from the initial ones mentioned in the article, we have underlined them.

Fisher's exact tests	**7-10** (N=11)	**11-14** (N=12)	**15-20** (N=7)
Audible gestures *with* object	OR = 0.922 p= 1	OR = 0.462 p= 0.196	OR = 0.387 p= 1
Audible gestures *without* object	OR = 0.673 p= 1	OR = 0.875 p= 1	OR = 0.771 p= 1
Mouth sounds	OR = 1.463 p= 1	OR = 0.773 p= 1	OR = 0.373 p= 0.922
Non-CV voc.	OR = 1.320 p= 1	OR = 0.508 p= 1	OR = 0.555 p= 1
Babblings	OR = inf p= 0.055	OR = 0.453 p= 0.130	OR = 1 p= 1
Verbal content	OR = 0 p= 1	OR = 0 p= 0.370	OR = 1.663 p= 1

5.
COMMUNICATION SENSIBLE A L'ATTENTION MATERNELLE ET PREDICTION LANGAGIERE

5.1. ARTICLE 4 : L'AJUSTEMENT DE LA COMMUNICATION A L'ATTENTION VISUELLE DE LA MERE PREDIT-IL LES COMPETENCES LANGAGIERES ULTERIEURES ?

Mawa Dafreville[1*], Michèle Guidetti [1], Marie Bourjade[1,2]

[1] CLLE, Université de Toulouse, CNRS, Toulouse, France

[2] Institut Universitaire de France

Synthèse de l'article 4

Contexte : Bien que la *communication sensible à l'attention du destinataire* constitue un marqueur d'intentionnalité conventionnellement étudié chez l'enfant préverbal, peu d'études ont testé son potentiel de prédiction langagière. Cette étude vise donc à tester l'effet prédicteur des ajustements *unimodal* et *intermodal* entre 7 à 20 mois sur les compétences de production langagière (i.e. lexicales, syntaxiques et pragmatiques) entre 18 à 42 mois.

Méthode : A la suite des études observationnelles réalisées sur les enfants préverbaux (voir articles 2 et 3, partie 4), un questionnaire en ligne permettant l'évaluation du Développement du Langage de Production en Français (DLPF; Bassano et al., 2005, 2020) chez l'enfant a été soumis à l'ensemble des parents des enfants ayant participé à l'étude. Vingt-et-un d'entre eux y ont répondu. Puis, a été testée la corrélation entre des scores de performances aux ajustements (généraux et de certains signaux) et les scores de performances langagières du DLPF. La latence entre la dernière observation et le DLPF était variable d'un individu à l'autre.

Résultats : Les ajustements *unimodaux* et *intermodaux* prédiraient davantage les performances lexicales et syntaxiques, que pragmatiques. Aussi, la latence entre la dernière observation et le DLPF influençait significativement les effets retrouvés. L'ajustement *unimodal* des mouvements corporels avec regard constituerait un prédicteur lexical chez les enfants ayant été observés plus de 16,5 mois avant le remplissage du questionnaire.

Conclusion : Certaines limites méthodologiques de l'étude (e.g. puissance statistique) appellent à poursuivre les investigations quant à l'effet de prédiction des ajustements. Toutefois, ces premiers résultats incitent à davantage considérer les signaux *non-canoniques* dans l'étude de la communication de l'enfant.

L'ajustement de la communication à l'attention visuelle de la mère prédit-il les compétences langagières ultérieures ?

Introduction

La *communication sensible à l'attention du destinataire* constitue un marqueur d'intentionnalité conventionnellement étudié chez des espèces non-humaines (*Papio anubis*: Bourjade et al., 2014; *Pan troglodytes*: Fröhlich et al., 2018; *Gorilla gorilla*: Genty et al., 2009; *Canis familiaris*: Horowitz, 2009; Kaplan, 2011; *Corvus corax*: Pika & Bugnyar, 2011) mais aussi chez l'enfant au stade préverbal (Gros-Louis & Wu, 2012; Igualada et al., 2015; Liszkowski et al., 2008; Rodrigues et al., 2021; Wu & Gros-Louis, 2014, 2017). Jusqu'alors, peu d'études ont montré que la *communication sensible à l'attention du destinataire* constituait un prédicteur langagier bien que les études ayant testé ce marqueur d'intentionnalité l'aient testé avec des comportements prédicteurs du langage (i.e. les marqueurs qu'on appellera ici *canoniques*, voir article 3, partie 4.2) tels que le geste de pointage (Cochet & Byrne, 2016; Colonnesi et al., 2010; Goldin-Meadow & Alibali, 2014; Iverson & Goldin-Meadow, 2005; Liszkowski et al., 2012; Mcgillion et al., 2017) ou des vocalisations impliquant le babillage (Donnellan et al., 2019; Jensen et al., 1988; Oller et al., 1998; Oller et al., 1999; Stoel-Gammon, 2011; Stolt et al., 2012). A notre connaissance, les résultats de deux études ayant investigué les capacités d'ajustement des signaux à l'attention visuelle suggèrent que l'ajustement des vocalisations d'enfants âgés de 10 à 13 mois à l'attention visuelle maternelle prédisait de meilleures performances de production et de compréhension langagière au MBCDI (*MacArthur-Bates Communicative Development Inventories ;* Fenson, 2007; Jackson-Maldonado, 2012) à l'âge de 15 mois (Wu & Gros-Louis, 2014, 2017). Toutefois comme la grande majorité des études ayant investigué la communication sensible à l'attention du destinataire sur les enfants préverbaux (Gros-Louis & Wu, 2012; Igualada et al., 2015; Liszkowski et al., 2008; Wu & Gros-Louis, 2014, 2017), ces deux études ne s'intéressaient qu'aux vocalisations (dans leur ensemble et sans distinction) et aux

gestes des pointage (Wu & Gros-Louis, 2014, 2017). De plus, les études investiguant la *communication sensible à l'attention du destinataire* montraient de nombreuses divergences méthodologiques et de résultats. C'est pourquoi, nos récentes investigations se sont attelées à construire deux descripteurs facilitant l'étude de la *communication sensible à l'attention du destinataire* et donc à l'investigation de son pouvoir de prédiction (Dafreville et al., 2021, *soumis*, 2022a, 2022b).

En effet, partant du principe du codage éthologique, à savoir catégoriser les signaux en fonction de la modalité sensorielle de perception prédominante, une première étude proposait d'investiguer la *communication sensible à l'attention du destinataire* à l'aide de ces deux descripteurs d'ajustement : *unimodal* et *intermodal* (Dafreville et al., 2021). Ces descripteurs ont été construits sur la base de la contrainte perceptive sur laquelle repose la *communication sensible à l'attention du destinataire* à savoir la non-perceptibilité des signaux silencieux-visuels quand le destinataire est visuellement inattentif. La construction de ces deux descripteurs résulte de la distinction de capacités sous-jacentes telles que celle d'éviter la production de signaux imperceptibles (*mismatching*) en inhibant les signaux silencieux-visuels face à un destinataire visuellement inattentif (i.e. ajustement *unimodal*) et celle de favoriser les signaux perceptibles (*matching*) par l'utilisation de signaux audibles-ou-tactiles dans cette même situation (i.e. ajustement intermodal) (Dafreville et al., 2021). Cette étude réalisée sur de jeunes chimpanzés enfants, juvéniles et adolescents, montrait que chez eux, l'acquisition des capacités d'ajustement *intermodal* précédait celle des capacités d'ajustement *unimodal* au cours du développement (Dafreville et al., 2021), soulignant ainsi l'intérêt d'utiliser ces descripteurs dans une perspective développementale.

Répliquée sur des enfants humains préverbaux âgés de 7 à 20 mois, cette étude avait testé l'effet du contact physique et du *Langage Adressé à l'Enfant* (LAE) sur ces capacités d'ajustement à l'attention visuelle maternelle, montrant ainsi un effet significatif du LAE sur ces ajustements (Article 2, partie 4.1; Dafreville et al.,

soumis, 2022a). Les enfants montraient une émergence des capacités d'ajustement *unimodal* précédant celle des capacités d'ajustement *intermodal* à partir des âges respectifs de 8 et 11 mois (Article 2, partie 4.1; Dafreville et al., *soumis,* 2022a). Par ailleurs, les enfants produisaient significativement plus de combinaisons multimodales distales (audio-visuelles) quand la mère était visuellement attentive que quand elle ne l'était pas (Article 2, partie 4.1; Dafreville et al., *soumis,* 2022a). Partant de là, il aura été question dans une seconde étude de confronter des marqueurs *canoniques* (i.e. des comportements prédicteurs du langage tels que le pointage ou le babillage) et *non-canoniques* (i.e. des comportements présentant un intérêt dans le développement communicationnel de l'enfant mais non considérés comme prédicteurs langagiers tels que les *bruits de bouches* ou les *gestes audibles avec un objet*) dans l'étude de ces différents ajustements (Article 2, partie 4.1; Dafreville et al., *soumis,* 2022b). Cette étude montrait que les marqueurs canoniques n'étaient pas les seuls à être ajustés (voir **Table 1**). Par exemple, les enfants de [11-14] mois ajustaient leurs gestes déictiques *proximaux* (i.e. gestes silencieux-visuels avec un objet tenu dans la main tels que le don–*giving*- ou le fait de montrer -*showing*-) et leurs gestes audibles *avec* objets (e.g. taper sur la table avec une cuillère ; voir **Table 1**). Les enfants de [15-20] mois ajustaient leurs *bruits de bouche* et leurs *vocalisations non-consonantiques* (e.g. *squeal, growl*, *grunt*; Oller, 2000). Par ailleurs, cette même étude montrait que l'ajustement de certains signaux était dépendant de la coordination du regard et de l'âge (**voir Table 1**). Ainsi les enfants de [7-10] mois ajustaient leurs *mouvements corporels* à condition d'y associer un regard vers le destinataire (voir **Table 1**).

Table1. Récapitulatif des principaux résultats retrouvés dans les études précédentes ayant investigué les ajustements *unimodal* et *intermodal*.

Etudes	Ajustement	Signaux ajustés	Signaux dépendant du regard
(Dafreville et al., 2021)	Création des descripteurs		
(Dafreville et al., *soumis,* 2022a)	**Unimodal**	Général	
	Intermodal Combinaisons	Général Audio-visuelles (i.e. *distale)*	
(Dafreville et al., *soumis,* 2022b)	**Unimodal**	Gestes déictiques *proximaux*	Mouvements corporels (7-10) Gestes déictiques *distaux* (11-14)
	Intermodal • *Gestes audibles*	Gestes audibles *avec* objet	
	• *Signaux oraux*	Babillages Bruits de bouche Vocalisations non-consonantiques	

Au regard de tout cela, il s'agit dans cette dernière étude de tester le potentiel de prédiction langagière de ces ajustements *unimodaux* et *intermodaux*. Pour cela, à la suite des études observationnelles réalisées récemment sur les enfants de 7 à 20 mois (articles 2 et 3; partie 4; Dafreville et al., *soumis*, 2022a, 2022b), un questionnaire en ligne permettant l'évaluation du *Développement du Langage de Production en Français* (*DLPF*; Bassano et al., 2005, 2020) a été soumis à l'ensemble des parents des enfants ayant participé à l'étude. Ce questionnaire permettait l'évaluation des compétences *lexicales* (i.e. vocabulaire), *syntaxiques* (i.e. phrases et grammaire) et *pragmatiques* (i.e. communication et utilisation du langage) d'enfants âgés de 18 à 42 mois. Le choix de ce questionnaire et non du MBCDI (*MacArthur-Bates Communicative Development Inventories ;* Fenson, 2007; Jackson-Maldonado, 2012) a été motivé par le fait de pouvoir tester des enfants observés précédemment (articles 2 et 3, partie 4; Dafreville et al., *soumis*, 2022a, 2022b) âgés de plus de 30 mois au moment du questionnaire. De plus, le DLPF avait été révisé récemment (Bassano et al., 2020).

Méthode

Participants

Les études observationnelles avaient été réalisées sur 30 dyades mère-enfants de Toulouse (France) et ses alentours (articles 2 et 3, partie 4; Dafreville et al., *soumis*, 2022a, 2022b). Le questionnaire de DLPF (Bassano et al., 2005, 2020) y faisant suite a été diffusé en ligne à l'ensemble des parents des enfants ayant participé à l'étude. Les parents de 25 enfants ont été contactés (l'adresse mail de ceux des 5 enfants restant n'étant plus valides) et ceux de 21 d'entre eux y ont répondu. Nous avons ainsi recueilli les réponses aux questionnaires de parents de 8, 9 et 4 enfants ayant respectivement fait partie des tranches d'âges des 7-10 mois, 11-14 mois et 15-20 mois lors des études observationnelles (articles 2 et 3, partie 4; Dafreville et al., *soumis*, 2022a, 2022b). Les réponses au questionnaire recouvraient des enfants de 18 à 34 mois.

Procédure

Les études observationnelles ont été réalisées par MD entre mars 2019 et mars 2020 (voir détails articles 2 et 3, partie 4; Dafreville et al., *soumis*, 2022a, 2022b) et le questionnaire du DLPF a été diffusé de novembre à décembre 2020. De ce fait, la période séparant la dernière observation d'un enfant et le remplissage du questionnaire par son/ses parents était variable d'un enfant à l'autre, présentant une moyenne de 15,56 mois (±2,97 mois). La mise en ligne du questionnaire du DLPF a été réalisée via l'utilisation du logiciel *Qualtrics* permettant ainsi de générer un lien menant au questionnaire qui avait été envoyé par e-mail aux parents.

Procédure concernant la protection des données personnelles

Avant de débuter le questionnaire, une première page présentant l'étude permettait aux parents de donner leur consentement sans lequel ils n'avaient pas accès au questionnaire. A la fin du questionnaire en ligne, les parents étaient soumis à un amendement au consentement nous autorisant à traiter les données. Toutes les données collectées à la suite du questionnaire ont été extraites de la plateforme

sans aucune sauvegarde en ligne et ont été anonymisées, traitées et sauvegardées sur un seul et même appareil.

Echantillonnage des réponses et mise en lien avec les données comportementales

Score des performances langagières du DLPF

Le DLPF est un questionnaire permettant l'évaluation de compétences *lexicales* (i.e. vocabulaire), *syntaxiques* (i.e. phrases et grammaire) et *pragmatiques* (i.e. communication et utilisation du langage) d'enfants âgés de 18 à 42 mois. Pour cela il génère plusieurs scores comme répertoriés dans la **Table 2** ci-dessous.

Table 2. Scores des compétences lexicales, syntaxiques et pragmatiques du questionnaire DLPF (Bassano et al., 2005, 2020).

Volet de compétences	Nature des compétences	Score	Compétences évaluées
Lexique	(lexicale)	**TOTVOC**	Production de mots produits
Syntaxe	Morphologie	**FORNOM**	Production de formes nominales
		FORVERB	Production de formes verbales
		FVPART	Production de formes verbales particulières
		FVERRO	Production de formes verbales erronées
	Phrase	**COMBNOT**(1)	Production de phrase (jamais, parfois, souvent)
	Complexité	**STRUCTPX**	Production de structures complexes
		PHRACPX	Production de phrases complexes
Pragmatique	Echanges langagiers	**ECHLANG**	Participation aux échanges langagiers
	Utilisation du langage	**UTILANG**	Utilisation du langage dans l'expression des actes de parole
	Organisation du langage	**ORGMESS**	Organisation discursive des messages

(1) La variable n'a pas été exploitées pour cette étude puisqu'elle était qualitative (réponse : jamais, parfois, souvent) et servait essentiellement à orienter directement la mère vers la rubrique pragmatique, si l'enfant ne produisait pas du tout de phrase.

Pour traiter les réponses au questionnaire obtenues, chaque *score obtenu* a été associé à un *score attendu* pour l'âge que l'enfant avait au moment où les parents remplissaient le questionnaire.

Scores de performance aux ajustements

Pour faire correspondre ces réponses aux résultats comportementaux des études observationnelles ayant investigué les ajustements *unimodal* et *intermodal* pour l'étude de la *communication sensible à l'attention du destinataire*, plusieurs scores ont été créés pour modéliser les ajustements et le pattern des combinaisons audio-visuelles. L'ajustement *unimodal* est modélisé par le score **UNIM**, l'ajustement *intermodal* par les scores **INTERM1** et **INTERM2** et le pattern des combinaisons audio-visuelles par **COMBIAV** (pour le détail, voir **Table 3**). Ces scores ont été calculés pour l'ensemble des signaux mais également pour des signaux spécifiques ayant montré un ajustement significatif dans les précédentes études (voir **Table 1** et article 3, Dafreville et al., *soumis,* 2022b). Le score **UNIM** a été calculé pour les signaux silencieux-visuels à savoir : les *gestes déictiques* avec et sans regard associé et les *mouvements corporels* avec regard. Les scores **INTERM1** et **INTERM2** ont été calculés pour les signaux audibles à savoir : les *gestes audibles avec objet*, les *babillages*, les *bruits de bouche* et les *vocalisations non-consonantiques*.

Contrôle temporel entre observations et DLPF

Du fait de la variabilité de la période séparant la dernière observation et le remplissage du questionnaire d'un individu à l'autre, une variable *latence-DLPF* a été considérée. Pour les représentations graphiques, les périodes *courtes* et périodes *longues* ont été distinguées. Pour délimiter ces périodes, le seuil considéré avait été la médiane des périodes entre dernière observation et DLPF qui était de 16,5 mois. Les périodes étaient ainsi considérées comme étant *courtes* et *longues* quand elles faisaient respectivement moins et plus de 16,5 mois (**voir Figure 1**).

Figure 1. Corrélation entre la période séparant la dernière observation et le remplissage du questionnaire et l'âge moyen des enfants aux observations. La médiane des périodes entre dernière observation et DLPF est représentée par la droite en pointillée grise. Cette médiane aura permis de délimiter les périodes dites *longues* et *courtes* de latence entre observation et DLPF, et ainsi catégoriser les enfants selon ces périodes.

Analyses statistiques

Pour tester l'effet des capacités d'ajustement sur les scores de performance langagières du DLPF (voir **Table 2**), une sélection de modèles sur des données de proportions suivant une loi binomiale et une fonction *logit* a été réalisée (*GLMM*; Bolker et al., 2009; Harrison, 2015). Une variable de proportion a été créée à partir des *scores obtenus* et *attendus* pour modéliser la comparaison entre les scores. C'est donc cette variable de comparaison qui, calculée pour chacun des scores du DLPF, constituait la variable dépendante. Les variables explicatives d'intérêt étaient les *scores d'ajustements* généraux et de certains signaux ainsi que celui de *combinaisons* audio-visuelles (voir **Tables 1 et 3**).

Table 3. Scores comportementaux d'ajustement des signaux et de combinaison audio-visuelle testés dans cette étude

Ajustement associé	Score	Définition	Calcul
Unimodal	**UNIM**	Nombre de signaux silencieux-visuels (sv) produits quand la mère est visuellement inattentive*	$\mathbf{SV_{inatt}}$
Intermodal	**INTERM 1**	Différence de production entre signaux audible-ou-tactiles (at) et silencieux-visuels (sv) quand la mère est visuellement inattentive	$\frac{(at - sv)_{inatt}}{(at + sv)_{inatt}}$
	INTERM 2	Comparaison des productions de signaux audible-ou-tactiles (at) et silencieux-visuels (sv) entre conditions d'attention et d'inattention visuelle maternelle	$\frac{(at - sv)_{inatt}}{(at + sv)_{inatt}} - \frac{(at -}{(at +}$
(autre)	**COMBIAV**	Différence de quantité de combinaisons audio-visuelles (combi) d'une condition visuelle à l'autre	$Combi_{att} - Combi_{inatt}$

* (*mismatching*; Dafreville et al., *soumis*, 2021a)

Ces variables explicatives ont toutes été testées seules et en interaction avec la variable *latence-DLPF* (i.e. période entre dernière observation et DLPF). Notons que dans cette sélection de modèles, les scores d'ajustements n'ont pas été testés avec l'âge moyen et l'âge de l'enfant au moment du remplissage du questionnaire pour les raisons suivantes. Premièrement, on reconnaît un effet de l'âge moyen sur les capacités d'ajustements (voir détails articles 2 et 3, partie 4; Dafreville et al., *soumis*, 2022a, 2022b) et deuxièmement, l'âge de l'enfant au moment du remplissage du questionnaire est de fait considéré dans la variable dépendante de score de DLPF permettant la comparaison entre scores obtenu et attendu. Aussi, des variables de contrôle ont été considérées permettant de justifier que c'est bien l'effet d'ajuster d'un signal et non juste la production de ce signal qui influence les scores langagiers (voir *Matériel Additionnel*, **Table S1**). Prenons l'exemple du babillage, imaginons que le modèle avec le score d'ajustement *intermodal* du babillage est le plus explicatif d'un score lexical. Or, on sait que dans la littérature, la production de babillage (indépendamment de son utilisation en contexte)

constitue un prédicteur langagier (Donnellan et al., 2019; Jensen et al., 1988; Oller et al., 1998; Oller et al., 1999; Stoel-Gammon, 2011; Stolt et al., 2012). De ce fait, pour s'assurer que c'est bien l'*ajustement* du babillage et pas le simple fait de produire plus de babillage (indépendamment de l'attention maternelle), il aura fallu mettre en compétition les deux variables dans la sélection de modèle, i.e. l'*ajustement* du babillage (i.e. INTERM1, INTERM2) et la production globale (e.g. occurrence) du babillage. Les modèles ont été implémentés en utilisant la fonction *glmer* du package R *lme4* version 1.1- 23 avec la fonction d'optimisation de *aictab* du package R *AICcmodavg* pour calculer l'*AICc*, qui est une transformation du Akaike Information Criterion (AIC) utilisé pour les petits échantillons. Pour la sélection du modèle, les modèles avec les valeurs AICc les plus basses sont considérés comme les modèles les plus explicatifs, mais les modèles dans un ΔAICc de 2 sont considérés comme équivalents (Burnham & Anderson, 2004). Dans certains cas, parmi les modèles équivalents les plus explicatifs, il y avait des scores d'ajustements équivalents à au moins un modèle n'incluant ni scores d'ajustement (**UNIM, INTERM1, INTERM2**) ni de combinaisons (**COMBIAV**). Dans ce cas, on considérait que l'effet d'ajustement potentiel n'étant pas prédominant ici, son pouvoir de prédiction n'était pas suffisamment fort pour être retenu dans cette étude. Une fois que le meilleur modèle a été sélectionné, des corrélations de Spearman ont été réalisées avec correction de Bonferroni. Ces tests de corrélation étaient testés entre une variable de *différence entre scores obtenu et attendu* (score obtenu - score attendu) et le score d'ajustement. Ainsi plus cette différence est élevée et plus le score de DLPF est bon. Tous les tests ont été effectués à l'aide du logiciel R v3.6.1 (http://cran.r-project.org) en considérant les p-values égales ou inférieures à 0,05 comme étant significatives. Par ailleurs, bien que les variables de notre échantillon ne suivent pas une loi normale, nous avons tout de même représenté la tendance linéaire des corrélations afin d'illustrer l'effet testé sur les représentations graphiques.

Résultats

- **Effet des scores d'ajustement sur les scores de performances lexicales, syntaxiques et pragmatiques du DLPF**

L'analyse GLMM visait à identifier les différents scores d'ajustement qui influençaient le score du DLPF testé. Concernant les compétences *lexicales*, le score TOTVOC a été affecté par l'ajustement *unimodal* des *mouvements corporels* avec regard produits quand la mère est visuellement inattentive (**UNIM**) et la *latence-DLPF* comme en témoigne le modèle d'interaction unique le plus explicatif (AICc = 246.87 ; comparaison avec modèle nul, $p < 0{,}001$; wi = 0.82 ; voir **Table S1**). Concernant les compétences *syntaxiques*, seuls un score de complexité est significativement influencé par les scores d'ajustements. Le score de **STRUCTPX** a été affecté par l'ajustement *intermodal* du babillage (**INTERM2**), des gestes audibles avec objets (**INTERM1**) et la *latence-DLPF* comme en témoignent les modèles les plus explicatifs (babillage : AICc = 125.68 ; comparaison avec modèle nul, $p < 0{,}001$; wi = 0.35 ; gestes audibles avec objet : AICc = 126.90 ; comparaison avec modèle nul, $p = 0{,}001$; wi = 0.19 ; voir **Table S1**). Concernant les compétences *pragmatiques*, aucun score n'a été significativement influencé par les scores d'ajustements. L'effet des scores d'ajustement sur les scores de performances lexicales et syntaxiques vont donc être précisés dans les analyses ci-dessous.

- **Effet de prédiction lexicale : production de mots (TOTVOC)**

Pour tester l'effet du nombre de *mouvements corporels* avec regard produits quand la mère est visuellement inattentive (**UNIM**) et de la *latence* sur le score TOTVOC, des tests de corrélation de Spearman ont été réalisés sur deux sous-échantillons d'enfants selon que le temps écoulé entre la dernière observation et DLPF ait été *courte* ou *longue* (voir **Figure 1**). Les enfants dont la *latence-DLPF* était considérée comme *longue* et produisant moins de *mouvements corporels* avec regard quand la mère était visuellement inattentive (**UNIM**), présentaient de meilleures

scores lexicaux (**TOTVOC ;** coefficient de corrélation de Spearman, r = −0.811, p = 0. 009, N=10 ; **voir Figure 2**) alors que ce n'est pas le cas pour les enfants dont la *latence-DLPF* était considérée comme *courte* (coefficient de corrélation de Spearman, r = −0.463, p = 0. 304, N=11). En somme, moins les enfants produisaient de *mouvements corporels* avec regard quand la mère était visuellement inattentive (**UNIM**), plus l'enfant possédait de mots dans son vocabulaire **(voir Figure 2**). Par ailleurs, graphiquement, en rapportant cette corrélation à chaque tranche d'âge, on soupçonne peu d'effet de celle-ci sur la tendance observée (**voir Figure 3**).

Figure 2. Corrélation entre le nombre de *mouvements corporels* avec regard quand la mère est visuellement inattentive (*mismatchings)* **et la performance lexicale TOTVOC** *(score obtenu – score attendu)* avec droite linéaire et intervalle de confiance de 'tendance'. Plus la différence de score est élevée, plus le score obtenu est bon (p-value : *<0.05, **<0.01, ***<0.001).

Figure 2. Corrélation entre le nombre de *mouvements corporels* avec regard quand la mère est visuellement inattentive (*mismatchings*) **et la performance lexicale de TOTVOC** (*score obtenu – score attendu*) avec droite linéaire et intervalle de confiance de 'tendance'. Plus la différence de score est élevée, plus le score obtenu est bon (p-value : *<0.05, **<0.01, ***<0.001). La corrélation par tranche d'âge n'a pas été testée par manque de puissance statistique des sous-échantillons.

- **Effet de prédiction syntaxique : production de structures complexes (STRUCTPX)**

Pour tester l'effet de l'ajustement *intermodal* du babillage (**INTERM2**), des gestes audibles avec objets (**INTERM1**) et la *latence-DLPF* sur le score **STRUCTPX**, des tests de corrélation de Spearman ont été réalisés sur deux sous-échantillons d'enfants selon que la *latence-DLPF* ait été *courte* ou *longue* (voir **Figure 1**). Il n'y a pas de corrélation significative entre l'ajustement *intermodal* du babillage (**INTERM2**) des enfants et leur score STRUCTPX, quelle que soit la longueur de la *latence-DLPF* (*courte* : coefficient de corrélation de Spearman, $r = -0.156$, $p = 1$, N=11 ; longue : coefficient de corrélation de Spearman, $r = 0.332$, $p = 0.697$,

N=10). Les enfants ne montrent pas non plus de corrélation significative entre l'ajustement *intermodal* des gestes audibles avec objets (**INTERM1**) et le score STRUCTPX, quelle que soit la longueur de *latence-DLPF* (*courte* : coefficient de corrélation de Spearman, r = 0.067, p = 1, N=11 ; longue : coefficient de corrélation de Spearman, r = 0.606, p = 0. 126, N=10 ; voir **Figure 4**).

Figure 4. Corrélation entre l'ajustement intermodal des gestes audibles avec objets (INTERM1) et la performance syntaxique STRUCTPX *(score obtenu – score attendu)* avec droite linéaire et intervalle de confiance de 'tendance'. Plus la différence de score est élevée, plus le score obtenu est bon. La droite en pointillé indique le seuil à dépasser pour que les signaux audibles soient produits en plus grande quantité que les signaux silencieux-visuels quand la mère est visuellement inattentive.

Discussion

Cette étude avait pour but de tester l'effet de différents ajustements *unimodaux* et *intermodaux* sur les performances lexicales, syntaxiques et

pragmatiques ultérieures. Ces ajustements prédiraient davantage les performances lexicales et syntaxiques, que pragmatiques. Aussi, la *latence-DLPF* influençait significativement les effets retrouvés. En effet, les tendances à la prédiction observées sont plus accentuées chez les enfants ayant été observés plus de 16,5 mois avant le remplissage du questionnaire (i.e. période *longue*). Chez ces derniers, l'ajustement *unimodal* des *mouvements corporels* avec regard constituerait un prédicteur lexical. Chez ces mêmes enfants, l'ajustement *intermodal* du babillage et des gestes audibles avec objets n'influence finalement pas la production de structures complexes (i.e. utilisation de déterminant ou de pronom). Toutefois, les limites méthodologiques de l'étude (notamment de puissance statistique) ne permettent pas de rejeter pleinement son implication dans les performances syntaxiques.

Les ajustements *unimodaux* et *intermodaux* ne prédisent pas de compétence pragmatique et la *latence-DLPF* influençait significativement les effets retrouvés. Pourtant, le principe de la *communication sensible à l'attention du destinataire* a été initialement définie sur la base des actes de langage (Austin, 1962) et donc de la pragmatique (Bates et al., 1975; Bruner, 1975, 1981). Le fait que ces ajustements ne prédisent pas de compétences pragmatiques relève probablement du fait qu'ils soient déjà la manifestation d'un niveau d'acte de langage relativement avancé et donc moins sujet au développement que les autres pans de la communication (i.e. lexique et syntaxe). Par ailleurs, toujours dans cette perspective des actes de langage, cet effet de la *latence-DLPF* souligne l'importance de l'expérience de l'enfant via notamment l'effet perlocutoire (i.e. effet sur le destinataire) dans l'acquisition des fonctions illocutoires (i.e. communication intentionnelle) nécessaire à l'acquisition de fonctions locutoires (i.e. langage). On peut donc imaginer que si les performances locutoires sont fonction de l'expérience de l'enfant, elles sont aussi fonction de la durée de pratique d'une même capacité d'ajustement. C'est pourquoi, l'influence de la *latence-DLPF* sur les compétences de production langagières (propre au stade locutoire) prend tout son sens dans le cadre de cette étude.

L'ajustement *unimodal* des *mouvements corporels* avec regard, plus que n'importe quel autre ajustement d'autres signaux testés (voir **Table 1**) constituerait un prédicteur lexical. Ce qu'il y a d'intéressant c'est que l'effet de l'ajustement des *mouvements corporels* avait été mis en compétition avec l'ajustement de gestes déictiques avec et sans regards nettement plus étudiés et reconnus comme prédicteurs langagiers par de nombreuses études comme pour le geste de *pointage* (Cochet & Byrne, 2016; Colonnesi et al., 2010; Goldin-Meadow & Alibali, 2014; Iverson & Goldin-Meadow, 2005; Liszkowski et al., 2012; Mcgillion et al., 2017) mais aussi par des études récentes défendant l'importance des gestes proximaux de *giving* et *showing* dans le développement communicationnel (e.g. Choi et al., 2021). Une explication possible pourrait être liée à une insuffisante puissance statistique comme évoqué par Dafreville et al. (*soumis*, 2022b). Toutefois, dans cette même étude, il avait été montré que les gestes déictiques pris dans leur ensemble s'ajustaient de manière plus franche et pérenne que les mouvements corporels (Dafreville et al., *soumis*, 2022b), mais l'effet de prédiction langagière de leur ajustement n'a pas surpassé celui des mouvements corporels avec regard. Une autre autre explication possible serait liée à la nécessité d'ajuster les mouvements corporels au préalable pour ajuster l'ensemble des signaux silencieux-visuels. Les mouvements corporels avec regard font partie des premiers signaux silencieux-visuels, autres que le sourire, à émerger bien avant l'âge de 6 mois (Iverson & Fagan, 2004; Nakata & Trehub, 2004; Scola et al., 2015) et à s'ajuster, avec les gestes déictiques, à partir de 7 mois au cours du développement (Dafreville et al., *sumitted*-b). Ils pourraient donc marquer l'entrée dans l'ajustement *unimodal* et leur ajustement constituerait donc une *amorce* nécessaire à l'ajustement des autres signaux. Si l'on regarde les résultats des analyses relatives à la sélection de modèle (voir **Table S1**), on peut constater que le modèle le plus explicatif suivant est celui de l'ajustement unimodal *général*. Ceci suggère que si on enlevait les modèles avec ajustements des mouvements corporels avec regard, l'ajustement unimodal général serait le plus explicatif pour prédire des compétences langagières plusieurs mois après son émergence.

Pourquoi l'ajustement *unimodal* des mouvements corporels serait donc un prédicteur lexical alors que, contrairement aux gestes déictiques, ils ne sont pas systématiquement associés à des compétences de communication référentielle qui constituent un moteur de curiosité et donc d'apprentissage ? Une explication pourrait être qu'à travers l'activité motrice, l'enfant délimiterait de façon précoce les tours de parole (Nakata & Trehub, 2004; Scola et al., 2015). En effet, Scola et al. (2015) ont montré que les enfants âgés de 5 à 9 mois faisaient moins de mouvements corporels quand la mère parlait ou chantait. Si l'activité motrice (corps et membres) est fonction du Langage Adressé à l'Enfant (LAE), elle est de fait fonction de nombreux épisodes verbaux au travers desquels il commencerait à ajuster ses signaux. Par ailleurs, Dafreville et al. (*soumis*, 2022a) ont montré que le LAE avait un effet *catalyseur* sur l'émergence des capacités d'ajustement *unimodal* puisque les enfants ajustaient leurs signaux silencieux-visuels plus tôt dans les situations où la mère parle que quand elle ne parle pas. Ainsi, si l'ajustement *unimodal* des mouvements corporels est optimisé par le LAE alors la quantité d'inputs verbaux pourrait expliquer l'exposition à plus de vocabulaire dans ces situations et donc montrer un pouvoir de prédiction lexicale. Néanmoins des analyses complémentaires seraient nécessaires pour tester cette hypothèse.

Concernant l'effet de l'ajustement *intermodal*, il n'y aurait pas d'effet de prédiction mais il n'est pas à exclure qu'il y ait un effet d'interaction entre la nature du signal audible ajusté et l'âge aux observation sur un potentiel effet de prédiction syntaxique (voir dans *Matériel Additionnel* **Figures S1 et S2**). L'étude présente plusieurs limites méthodologiques telles que la petite taille de l'échantillon ayant répondu au questionnaire (rapporté au découpage par tranche d'âge étudiées) et la variabilité temporelle entre les observations et le remplissage du questionnaire d'un individu à l'autre. De plus les réponses au questionnaire des enfants ayant été observés à 15-20 mois (N=4) étaient deux fois moins représentées que celles relatives aux enfants observés à 11-14 mois (N=9) et 7-10 mois (N=8). Toutefois, malgré ces limites, ces premiers résultats suggèrent plusieurs points intéressants. Chez les enfants observés à 7-10 mois, on remarque qu'entre les ajustements du

babillage et gestes audibles avec objet, les individus ont des scores d'ajustement relativement stables (i.e. soit les enfants ajustent leur babillage et leur gestes audibles, soit ils n'ajustent aucun des deux). Toutefois deux d'entre eux, AP26 et AD3 qui ajustent bien leurs gestes audibles, n'ajustent pas leur babillage. Ce sont par ailleurs ceux qui présentent de meilleurs scores de structures complexes (**STRUCTPX**). Chez les enfants observés à 11-14 mois, les scores sont plus largement dispersés autour du score attendu avec globalement de meilleurs scores d'ajustement des gestes audibles que du babillage excepté pour AX15, LA16 et LO6 qui ajusteront mieux leur babillage et qui font tous partie des scores allant au-delà du seuil attendu. Notons tout de même que GA3 qui a le meilleur score chez les enfants de la catégorie *période courte*, a le meilleur score d'ajustement de gestes audibles (mais n'ajuste pas du tout le babillage) et CO15 qui a le meilleur score chez les enfants de la catégorie période *longue*, ajuste aussi bien son babillage que ses gestes audibles. Chez les enfants observés à 15-20 mois, les scores obtenus tournent davantage autour du score attendu excepté pour AU30 qui est un peu en dessous mais qui est le seul des 4 enfants à être né prématurément. Par ailleurs on constate que parmi les trois enfants présentant un score autour du score attendu, AR13 (exposé à trois langues dans son quotidien) est le seul à ajuster son babillage alors qu'il est le seul des enfants de cette tranche d'âge à ne pas ajuster ses gestes audibles, comme pour les enfants observés à 7-10 mois.

Conclusion

Il s'agissait dans cette dernière étude de tester le pouvoir de prédiction langagière des ajustements *unimodaux* et *intermodaux*. Il en résulte que ces ajustements prédiraient davantage les performances lexicales et syntaxiques, que pragmatiques et que ce pouvoir de prédiction semble être fonction de l'expérience de l'enfant (i.e. sa durée de pratique de l'ajustement depuis l'émergence des ajustements). De plus, l'ajustement de certains signaux spécifiques, plus que l'ajustement des signaux en général semble être déterminant pour l'acquisition de

compétences langagières futures. L'ajustement *unimodal* prédirait les compétences lexicales. De plus, les signaux spécifiques ajustés vont au-delà des signaux *canoniques* conventionnels, ce qui souligne l'importance d'inclure davantage de signaux à émergence précoce (e.g. les *mouvements corporels*, et les gestes audibles avec objet) dans l'étude de la communication de l'enfant préverbal. Toutefois, il semblerait y avoir un effet de l'âge sur ce potentiel de prédiction langagière qu'il nous faudrait renseigner dans de futures investigations.

Références

Austin, J. L. (1962). Speech acts. In *Choice Reviews Online* (Vol. 37). https://doi.org/10.5860/choice.37-0017

Bassano, D., Labrell, F., & Bonnet, P. (2020). *Évaluer Les Débuts Du Langage Avec Le DLPF (révision)*.

Bassano, D., Labrell, F., Champaud, C., Lemétayer, F., & Bonnet, P. (2005). Le DLPF : un nouvel outil pour l'évaluation du développement du langage de production en français. In *Enfance* (Vol. 57). https://doi.org/10.3917/enf.572.0171

Bates, E., Camaioni, L., & Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill-Palmer Quarterly of Behavior and Development*, *21*(3), 205–226.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology and Evolution*, *24*(3), 127–135. https://doi.org/10.1016/j.tree.2008.10.008

Bourjade, M., Meguerditchian, A., Maille, A., Gaunet, F., & Vauclair, J. (2014). Olive baboons, Papio anubis, adjust their visual and auditory intentional gestures to the visual attention of others. *Animal Behaviour*, *87*, 121–128. https://doi.org/10.1016/j.anbehav.2013.10.019

Bruner, J. (1975). The ontogenesis of speech acts. *Journal of Child Language*, *2*(1), 1–19. https://doi.org/10.1017/S0305000900000866

Bruner, J. (1981). *Intention in the structure of action and interaction* (Lipsett L). Norwood: Ablex, Advances in infancy research.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods and Research*, *33*(2), 261–304. https://doi.org/10.1177/0049124104268644

Choi, B., Wei, R., & Rowe, M. L. (2021). Show, Give, and Point Gestures Across Infancy Differentially Predict Language Development. *Developmental Psychology*, *57*(6), 851–862. https://doi.org/10.1037/dev0001195

Cochet, H., & Byrne, R. W. (2016). Communication in the second and third year of life: Relationships between nonverbal social skills and language. *Infant Behavior and Development*, *44*, 189–198. https://doi.org/10.1016/j.infbeh.2016.07.003

Colonnesi, C., Stams, G. J. J. M., Koster, I., & Noom, M. J. (2010). The relation between pointing and language development: A meta-analysis. *Developmental Review*, *30*(4), 352–366. https://doi.org/10.1016/j.dr.2010.10.001

Dafreville, M., Guidetti, M., & Bourjade, M. (n.d.-a). Attention-sensitive signaling by 7-to-20-month-old infants in a comparative perspective. *(Submitted)*.

Dafreville, M., Guidetti, M., & Bourjade, M. (n.d.-b). Beyond pointing gestures: Attention-sensitive communication in 7-to-20-month-old infants develops pragmatic skills. *Child Development*.

Dafreville, M., Hobaiter, C., Guidetti, M., Sillam-Dussès, D., & Bourjade, M. (2021). Sensitivity to the communicative partner's attentional state: A developmental study on mother–infant dyads in wild chimpanzees (Pan troglodytes schweinfurthii). *American Journal of Primatology*, *83*(12). https://doi.org/10.1002/ajp.23339

Donnellan, E., Bannard, C., Mcgillion, M. L., Slocombe, K. E., & Matthews, D. (2019). *Infants ' intentionally communicative vocalizations elicit responses from caregivers and are the best predictors of the transition to language : A longitudinal investigation of infants ' vocalizations , gestures and word production*. (March), 1–21. https://doi.org/10.1111/desc.12843

Fenson, L. (2007). *MacArthur-Bates communicative development inventories*. Baltimore: MD: Paul H. Brookes Publishing Company.

Fröhlich, M., Wittig, R. M., & Pika, S. (2018). The ontogeny of intentional communication in chimpanzees in the wild. *Developmental Science*, *22*(1), 1–14. https://doi.org/10.1111/desc.12716

Genty, E., Breuer, T., Hobaiter, C., & Byrne, R. W. (2009). Gestural communication of the gorilla (Gorilla gorilla): repertoire , intentionality and possible origins. *Animal Cognition*, *12*, 527–546. https://doi.org/10.1007/s10071-009-0213-4

Goldin-Meadow, S., & Alibali, M. W. (2014). Gesture's role in speaking, learning, and creating language Susan. *Bone*, *23*(1), 1–7. https://doi.org/10.1146/annurev-psych-113011-143802.Gesture

Gros-Louis, J., & Wu, Z. (2012). Twelve-month-olds' vocal production during pointing in naturalistic interactions: Sensitivity to parents' attention and responses. *Infant Behavior and Development*, *35*(4), 773–778. https://doi.org/10.1016/j.infbeh.2012.07.016

Harrison, X. A. (2015). A comparison of observation-level randomeffect and Beta-Binomial models for modelling overdispersion in Binomial data in ecology & evolution. *PeerJ*, *2015*(7). https://doi.org/10.7717/peerj.1114

Horowitz, A. (2009). Attention to attention in domestic dog (Canis familiaris) dyadic play. *Animal Cognition*, *12*(1), 107–118. https://doi.org/10.1007/s10071-008-0175-y

Igualada, A., Bosch, L., & Prieto, P. (2015). Language development at 18 months is related to multimodal communicative strategies at 12 months. *Infant Behavior and Development*, *39*, 42–52. https://doi.org/10.1016/j.infbeh.2015.02.004

Iverson, J. M., & Fagan, M. K. (2004). *Infant Vocal – Motor Coordination : Precursor to the Gesture – Speech System ? 75*(4), 1053–1066.

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science*, *16*(5), 367–371. https://doi.org/10.1111/j.0956-7976.2005.01542.x

Jackson-Maldonado, D. (2012). MacArthur-Bates Communicative Development Inventories. *The Encyclopedia of Applied Linguistics*. https://doi.org/10.1002/9781405198431.wbeal0745

Jensen, T. S., Bøggild-Andersen, B., Schmidt, J., Ankerhus, J., & Hansen, E. (1988). Perinatal Risk Factors and First-Year Vocalizations: Influence on Preschool Language and Motor Performance. *Developmental Medicine & Child Neurology*, *30*(2), 153–161. https://doi.org/10.1111/j.1469-8749.1988.tb04746.x

Kaplan, G. (2011). Pointing gesture in a bird- merely instrumental or a cognitively complex behavior? *Current Zoology*, *57*(4), 453–467. https://doi.org/10.1093/czoolo/57.4.453

Liszkowski, U., Albrecht, K., Carpenter, M., & Tomasello, M. (2008). Infants ' visual and auditory communication when a partner is or is not visually attending. *Infant Behavior & Development*, *31*, 157–167. https://doi.org/10.1016/j.infbeh.2007.10.011

Liszkowski, U., Brown, P., Callaghan, T., & Takada, A. (2012). *A Prelinguistic Gestural Universal of Human Communication*. *36*, 698–713. https://doi.org/10.1111/j.1551-6709.2011.01228.x

Mcgillion, M. L., Herbert, J. S., Pine, J., Vihman, M., DePaolis, R., Keren-Portnoy, T., & Matthews, D. (2017). What Paves the Way to Conventional Language? The Predictive Value of Babble, Pointing, and Socioeconomic Status. *Child Development Perspectives*, 2013–2015.

Murillo, E., Montero, I., & Casla, M. (2019). *Rhythmic movements with objects at 9 months are related to proximal deictic gestures at 12 months* .

Nakata, T., & Trehub, S. E. (2004). Infants' responsiveness to maternal speech and singing.

Infant Behavior and Development, *27*(4), 455–464. https://doi.org/10.1016/j.infbeh.2004.03.002

Oller, D. K. (2000). *The Emergence of the Speech Capacity*.

Oller, D. K., Eilers, R. E., Neal, A. R., & Cobo-Lewis, A. B. (1998). Late onset canonical babbling: A possible early marker of abnormal development. *American Journal on Mental Retardation*, Vol. 103, pp. 249–263. https://doi.org/10.1352/0895-8017(1998)103<0249:LOCBAP>2.0.CO;2

Oller, D. K., Eilers, R. E., Neal, A. R., & Schwartz, H. K. (1999). Precursors to speech in infancy. *Journal of Communication Disorders*, *32*(4), 223–245. https://doi.org/10.1016/s0021-9924(99)00013-1

Pika, S., & Bugnyar, T. (2011). The use of referential gestures in ravens (Corvus corax) in the wild. *Nature Communications*, *2*(1), 560–565. https://doi.org/10.1038/ncomms1567

Rodrigues, E. D., Marôco, J., & Frota, S. (2021). Communicative gestures in 7–12-month infants: A phylogenetic comparative approach. *Infant and Child Development*, *30*(5), 1–11. https://doi.org/10.1002/icd.2262

Scola, C., Bourjade, M., & Jover, M. (2015). Social interaction is associated with changes in infants' motor activity. *Socioaffective Neuroscience & Psychology*, *5*(1), 28256. https://doi.org/10.3402/snp.v5.28256

Stoel-Gammon, C. (2011). Relationships between lexical and phonological development in young children. *Journal of Child Language*, *38*(1), 1–34. https://doi.org/10.1017/S0305000910000425

Stolt, S., Lehtonen, L., Haataja, L., & Lapinleimu, H. (2012). Development and predictive value of early vocalizations in very-low-birth-weight children: A longitudinal study. *Clinical Linguistics and Phonetics*, *26*(5), 414–427. https://doi.org/10.3109/02699206.2011.648365

Wu, Z., & Gros-Louis, J. (2014). Infants' prelinguistic communicative acts and maternal responses: Relations to linguistic development. *First Language*, *34*(1), 72–90. https://doi.org/10.1177/0142723714521925

Wu, Z., & Gros-Louis, J. (2017). The Value of Vocalizing: 10-Month-Olds' Vocal Usage Relates to Language Outcomes at 15 Months. *Infancy*, *22*(1), 23–41. https://doi.org/10.1111/infa.12150

Matériel additionnel

L'ajustement de la communication à l'attention visuelle de la mère prédit-il les compétences langagières ultérieures ?

Mawa Dafreville[1*], Michèle Guidetti [1], Marie Bourjade[1,2]

[1] CLLE, Université de Toulouse, CNRS, Toulouse, France
[2] Institut Universitaire de France

Table S1. Sorties GLMM après sélection de modèles pour les scores de compétences lexicales (TOTVOC) et syntaxiques (STRUCTPX)

			TOTVOC			STRUCTPX		
Num	**Modèles**	**K**	**AICc**	**ΔAICc**	**AICcWt**	**AICc**	**ΔAICc**	**AICcWt**
1	Individu	2	271,83	24,96	0	133,52	7,84	0,01
2	Tranche d'âge	4	276,79	29,91	0	139,17	13,49	0
3	Sexe	3	272,91	26,04	0	131,57	5,89	0,02
4	Age moyen	3	273,37	26,49	0	136,2	10,52	0
5	PDOQ(1)	3	272,99	26,12	0	131,33	5,66	0,02
6	$UNIM_{général}$	3	259,64	12,76	0	133,84	8,16	0,01
7	$INTERM1_{général}$	3	267,44	20,57	0	129,48	3,8	0,05
8	$INTERM2_{général}$	3	274,39	27,51	0	135,73	10,05	0
9	Total de GD(2)	3	273,63	26,76	0	135,96	10,28	0
10	$UNIM_{GD}$	3	273,79	26,91	0	136,23	10,55	0
11	Total de GD distaux	3	274,03	27,16	0	135,59	9,91	0
12	$UNIMD_{GDdistaux}$	3	273,93	27,06	0	136,26	10,59	0
13	Total de GD proximaux	3	274	27,13	0	136,26	10,59	0
14	$UNIMD_{GDproximaux}$	3	274,33	27,45	0	136,07	10,4	0
15	Proportion de GAO(3)	3	274,5	27,62	0	136,24	10,56	0

16	Total de GAO	3	272,7	25,83	0	136,1	10,43	0	
17	INTERM1$_{GAO}$	3	273,57	26,7	0	136,18	10,5	0	
18	INTERM2$_{GAO}$	3	273,47	26,59	0	136,02	10,35	0	
19	Proportion de babillage	3	272,66	25,78	0	135,71	10,04	0	
20	Total de babillage	3	270,91	24,03	0	134,49	8,81	0	
21	INTERM1$_{babillage}$	3	269,68	22,8	0	134,46	8,79	0	
22	INTERM2$_{babillage}$	3	274,23	27,36	0	136,26	10,58	0	
23	Proportion de BB[(4)]	3	274,09	27,21	0	136,18	10,51	0	
24	Total de BB	3	274,15	27,28	0	135,51	9,83	0	
25	INTERM1$_{BB}$	3	272,9	26,03	0	133,53	7,85	0,01	
26	INTERM2$_{BB}$	3	274,57	27,7	0	136,14	10,46	0	
27	Proportion de NonCV[(5)]	3	272,27	25,4	0	136,15	10,48	0	
28	Total de NonCV	3	273,59	26,71	0	133,94	8,26	0,01	
29	INTERM1$_{NonCV}$	3	270,01	23,13	0	131,54	5,86	0,02	
30	INTERM2$_{NonCV}$	3	274,2	27,32	0	134,95	9,28	0	
31	UNIM$_{général}$: PDOQ	5	257,63	10,75	0	134,22	8,55	0	
32	INTERM1$_{général}$: PDOQ	5	263,58	16,7	0	129,36	3,69	0,06	
33	INTERM2$_{général}$: PDOQ	5	278,44	31,56	0	136,13	10,45	0	
34	UNIM$_{GD}$:PDOQ	5	278,1	31,23	0	136,34	10,67	0	
35	UNIMD$_{GDdistaux}$:PDOQ	5	279,05	32,17	0	136,86	11,19	0	
36	UNIMD$_{GDproximaux}$:PDOQ	5	273,71	26,84	0	136,63	10,95	0	
37	INTERM1$_{GAO}$:PDOQ	5	277,96	31,08	0	**127,12**	**1,44**	**0,17**	p = 0,001
38	INTERM2$_{GAO}$:PDOQ	5	276,22	29,35	0	134,62	8,94	0	
39	INTERM1$_{babillage}$:PDOQ	5	270,4	23,52	0	132,38	6,7	0,01	
40	INTERM2$_{babillage}$:PDOQ	5	277,7	30,82	0	**125,68**	**0**	**0,35**	p < 0,001
41	INTERM1$_{BB}$:PDOQ	5	275,3	28,42	0	134,39	8,71	0	
42	INTERM2$_{BB}$:PDOQ	5	279,53	32,66	0	137,84	12,17	0	
43	INTERM1$_{NonCV}$:PDOQ	5	263,95	17,08	0	133,28	7,61	0,01	
44	INTERM2$_{NonCV}$:PDOQ	5	279,1	32,23	0	135,76	10,08	0	

45	COMBIAV	3	274,53	27,65	0		135,57	9,9	0
46	COMBIAV :PDOQ	5	278,52	31,64	0		134,31	8,64	0
47	Total de GD + Regard	3	274,05	27,17	0		135,71	10,03	0
48	$UNIM_{GD + Regard}$	3	274,32	27,44	0		135,99	10,32	0
49	Total de GD distaux + Regard	3	274,52	27,65	0		134,29	8,61	0
50	$UNIMD_{GDdistaux + Regard}$	3	274,57	27,69	0		136,15	10,48	0
51	Total de GD proximaux + Regard	3	274,09	27,21	0		136,26	10,59	0
52	$UNIMD_{GDproximaux + Regard}$	3	274,57	27,69	0		136,15	10,48	0
53	Total de MC(6) + Regard	3	261,58	14,7	0		129,13	3,45	0,06
54	$UNIM_{MC + Regard}$	3	250,01	3,13	0,17		130,86	5,18	0,03
55	$UNIM_{GD + Regard}$:PDOQ	5	279,33	32,46	0		137,51	11,83	0
56	$UNIMD_{GDdistaux + Regard}$:PDOQ	5	279,41	32,53	0		137,75	12,08	0
57	$UNIMD_{GDproximaux + Regard}$:PDOQ	5	279,41	32,53	0		137,75	12,08	0
58	**$UNIM_{MC + Regard}$:PDOQ**	**5**	**246,87**	**0**	**0,82**	$p < 0,001$	128,59	2,92	0,08

(1) PDOQ : *Période entre dernière observation et questionnaire DLPF*
(2) GD : Gestes déictiques
(3) GAO : Gestes audibles avec objets
(4) BB : Bruits de bouche
(5) NonCV : Vocalisation non-consonantique

Figure S1. Corrélation entre l'ajustement intermodal du babillage (INTERM2) et la performance syntaxique STRUCTPX *(score obtenu – score attendu)* avec droites linéaires de 'tendance' par tranche d'âge. Plus la différence de score est élevée, plus le score obtenu est bon (p-value : *<0.05, **<0.01, ***<0.001). La corrélation par tranche d'âge n'a pas été testée par manque de puissance statistique des sous-échantillons.

Figure S2. Corrélation entre l'ajustement intermodal du geste audible avec objet (INTERM1) et la performance syntaxique STRUCTPX *(score obtenu – score attendu)* avec droites linéaires de 'tendance' par tranche d'âge. Plus la différence de score est élevée, plus le score obtenu est bon (p-value : *<0.05, **<0.01, ***<0.001). La corrélation par tranche d'âge n'a pas été testée par manque de puissance statistique des sous-échantillons.

6.
DISCUSSION GENERALE

6.1. RAPPEL DES PRINCIPAUX RESULTATS

Ce travail de thèse avait pour but d'étudier le développement de la *communication sensible à l'attention du destinataire* dans une perspective comparative. A travers une approche multimodale, ce travail devait permettre d'alimenter les modèles théoriques sur le remaniement de la signalisation multimodale menant à l'émergence de la parole (essentiellement audible et vocale) (i) chez l'enfant en période préverbale de 7 à 20 mois et (ii) à l'échelle évolutive. Faisant état de la divergence méthodologique quant aux travaux existants sur la *communication sensible à l'attention du destinataire* (voir partie 1.3), il aura été question dans un premier temps d'opérationnaliser son étude. Pour cela, nous avons introduit deux descripteurs d'ajustement *unimodal* et *intermodal* que nous avons pu tester sur le chimpanzé immature (voir partie 3) et l'enfant humain âgé de 7 à 20 mois (voir partie 4). L'ajustement *unimodal* relève de la capacité à éviter la production de signaux imperceptibles (*mismatching*) en inhibant les signaux silencieux-visuels face à un destinataire visuellement inattentif. L'ajustement *intermodal* relève de la capacité à favoriser les signaux perceptibles (*matching*) par l'utilisation de plus de signaux audibles-ou-tactiles que de signaux silencieux-visuels quand le destinataire est visuellement inattentif (Dafreville et al., 2021). Quelle que soit l'espèce étudiée et la question traitée dans nos investigations, l'étude de ces deux descripteurs aura permis de construire une histoire développementale de la *communication sensible à l'attention du destinataire* chez le primate immature humain et non-humain.

Nos études ont montré que les jeunes primates chimpanzés et humains observés en condition écologique ajustaient leurs signaux à l'attention visuelle de leur mère et que ces capacités d'ajustement *unimodal* et *intermodal* émergeaient à différents stades de leur développement (voir parties 3 et 4). En s'intéressant à la nature des signaux impliqués dans l'émergence de ces capacités, l'ajustement des signaux à l'attention visuelle va bien au-delà de l'ajustement des signaux que nous

qualifions ici de *canoniques* (i.e. des comportements reconnus comme étant des prédicteurs fondamentaux du langage par le champ d'étude, voir Article 2, partie 4.2). Ce qui nous a donc conduit à nous questionner quant à l'importance de ces signaux dans le développement communicationnel et donc l'effet de prédiction langagière de leur ajustement chez l'enfant.

La capacité d'ajustement *unimodal* pourrait constituer un prédicteur *lexical* du langage chez l'enfant préverbal, cependant, nos observations ne semblent pas montrer d'effet de prédiction de la capacité d'ajustement *intermodal* (voir partie 5). Par ailleurs, notons que l'ajustement de certains signaux pris isolément est plus prédicteur que les capacités d'ajustement de signaux pris dans leur ensemble (voir partie 5). Toutefois, sachant que ces capacités d'ajustement sont retrouvées chez les deux espèces, on peut se demander pourquoi celles retrouvées chez l'espèce humaine mènent au langage alors que celles retrouvées chez le chimpanzé n'y mènent pas. Ceci nous amène donc à comparer les trajectoires développementales de ces capacités (construites à partir de nos échantillons) et discuter les points convergents et divergents entre ces deux espèces.

Sur la base de ces investigations, il s'agira ici de (i) construire une histoire développementale des différents ajustements pour chacune des espèces étudiées et de (ii) discuter des perspectives méthodologiques et théoriques que cette étude suggère concernant l'effet des prédictions langagières. Puis, on s'attèlera à (iii) intégrer nos résultats dans une perspective évolutionnaire en réalisant *une* comparaison inter-espèce de la *communication sensible à l'attention du destinataire* entre chimpanzés et humains. Pour finir, (iv) seront discutées les limites de cette investigation et les perspectives qui en découlent.

6.2. HISTOIRES DEVELOPPEMENTALES DE LA COMMUNICATION SENSIBLE A L'ATTENTION DU DESTINATAIRE CHEZ LE PRIMATE IMMATURE

Sur les Figures 1 et 2 ci-dessous, des *cartes développementales* modélisant le développement des capacités d'ajustement *unimodal* et *intermodal* ont été construites pour chacune des espèces étudiées à partir des résultats obtenus lors des études menées sur les chimpanzés (article 1, partie 3) et les enfants préverbaux (articles 2 et 3, partie 4). Ces cartes permettent de représenter la trajectoire développementale de la *communication sensible à l'attention du destinataire* à partir des signaux par lesquels émergent les deux types d'ajustement (sous la forme de « ramifications »).

Sur ces cartes, les signaux ajustés en fonction de l'âge des individus testés ont donc été représentés : silencieux-visuels pour l'ajustement *unimodal* et audibles et tactiles pour l'ajustement *intermodal*. Les signaux qui, avant émergence de l'ajustement testé, montraient une forme partielle ou inachevée de cet ajustement était qualifiés de *proto-ajustés*. La méthode de reconnaissance du *proto-ajustement* des signaux est renseignée ci-dessous pour chacune des espèces (voir parties 6.2.1 et 6.2.2)

6.2.1. Chez le chimpanzé : une histoire de **négociation et de manipulation d'objet**

Figure 1, pour identifier les signaux *proto-ajustés* chez le chimpanzé, la méthode était descriptive, elle se basait sur le tableau en Annexe 7. Pour l'ajustement *unimodal*, les signaux silencieux-visuels était qualifiés de proto-ajustés s'ils étaient moins produits en condition d'inattention visuelle qu'en condition d'attention visuelle. Pour *l'intermodal*, les signaux proto-ajustés étaient les signaux

audibles-ou-tactiles qui n'étaient pas plus produits en condition d'inattention visuelle qu'en condition d'attention visuelle.

Figure 1. Carte développementale de l'ajustement *unimodal* et *intermodal* des gestes du chimpanzé de l'enfance à l'adolescence. Les gestes *proto-ajustés* sont des signaux qui, avant émergence, montrent une forme partielle ou inachevée d'ajustement.

Comme illustré sur la Figure 1, chez le jeune chimpanzé la capacité d'ajustement *intermodal* des signaux émergeait au stade juvénile. La capacité d'ajustement *unimodal* quant à elle semblait émerger au stade adolescent (voir Article 1, partie 3). En s'intéressant à la nature des signaux *proto-ajustés* au cours du développement, on constate qu'avant l'émergence de la capacité d'ajustement associée (i.e. *unimodal* pour les signaux silencieux-visuels et *intermodal* pour les signaux audible-ou-tactiles), il y a toujours plus de signaux *proto-ajustés* (i.e. montrant une forme partielle ou inachevée d'ajustement) que de signaux ajustés au moment

de l'émergence (voir Figure 1 et annexe 7, Table S1). Concernant l'ajustement *intermodal*, celui-ci est préférentiellement tactile au stade juvénile puis devient préférentiellement audible à l'adolescence (voir Article 1, partie 3). Notons par ailleurs que les gestes audibles qui s'ajustent sont exclusivement produits avec un objet. Ces premiers constats suggèrent qu'il y aurait comme une « élection » de signaux (parmi les signaux proto-ajustés « candidats ») au moment de l'émergence des capacités d'ajustement que nous chercherons à discuter au regard (i) du *sens* associé au signal (Hobaiter & Byrne, 2014) et (ii) de l'intégration d'un objet lors de l'ajustement.

En effet, la nature des gestes *tactiles* et *silencieux-visuels* via lesquels émergent la capacité d'ajustement à l'attention de la mère pourrait être dépendante de l'enjeu de la négociation associé à ceux-ci au moment de l'interaction. Hobaiter et Byrne (2014) avaient, sur cette même communauté de chimpanzés (de tout âge), cherché à rapporter le sens le plus probable associé à chacun des gestes du répertoire. Pour cela, ils s'étaient intéressés aux ASOs (*apparently satisfactory outcome*; Hobaiter & Byrne, 2014), c'est-à-dire le sens qu'on peut attribuer à un signal en se basant sur la réponse comportementale du destinataire ayant entraîné la cessation de communication du signaleur (en apparence satisfait). Ils ont ainsi pu mettre en place plusieurs ASOs parmi lesquels on distingue ceux de *négation* (i.e. « Arrête ça », « va-t'en ») de ceux de *négociation* qui initie une interaction sociale (e.g. « épouille-moi ») ou la développe (e.g. « rapproche-toi », « donne-moi cet objet », « repositionne-toi plutôt comme ça »). Comme on peut le voir **Table 1**, les seuls gestes à être proto-ajustés chez l'enfant chimpanzé sont tactiles et trois d'entre eux sur quatre (i.e. *grab* , *hit other*, *push, stomp-2-feet other*) ont un premier sens de *négation* alors que le seul à s'apparenter à de la *négociation* est le seul à s'ajuster concrètement au stade juvénile (i.e. *push(directed)*). Ce constat suggère que les contextes de négociation entre l'enfance et le stade juvénile seraient davantage moteurs de l'émergence de la capacité d'ajustement *intermodal*. Le *push(directed)* (i.e. geste via lequel émerge l'ajustement intermodal) est un geste où le signaleur encercle de son bras son destinataire pour le rapprocher de lui. Ce geste a pour but de repositionner

le corps du destinataire lors de l'épouillage mais aussi de l'inciter à se rapprocher de lui. Ce qui aurait du sens au regard du sevrage qui marque le passage du stade enfant à juvénile. De plus, avant l'émergence de l'ajustement *unimodal* au stade adolescent, on constate que parmi les trois gestes silencieux-visuels proto-ajustés au stade juvénile, le seul à ne pas concrètement s'ajuster à l'adolescence (i.e. *raise arm*) est le seul des signaux à ne pas être principalement de *négociation* (en considérant le sens premier et secondaire, voir Table1). Ces résultats soulignent donc l'importance de la négociation dans l'émergence de ces capacités d'ajustement du jeune chimpanzé; l'enjeu inhérent au principe de *négociation* semblant ainsi constituer une pression d'apprentissage efficace sur le développement de ces capacités d'ajustements. En effet, il semble que les comportements relatifs à la négociation (en comparaison à ceux de la négation) soient plus sujets à une diversité de réponses possibles (plus ou moins satisfaisantes pour le signaleur) de la part du destinataire. Les gestes apparentés à la négociation sont essentiellement des gestes de repositionnement (e.g. *Present (directed)*, *Push (directed)*) et il existe donc une multitude de repositionnements possibles. De ce fait, l'optimisation de la signalisation s'en voit davantage nécessaire et c'est en cela que ce contexte exerce une pression d'apprentissage efficace sur le jeune chimpanzé.

Table 1. Mise en correspondance des gestes ajustés (ou proto-ajustés) avec le sens le plus probable associé à partir de Hobaiter et Byrne (2014) par tranche **d'âge.**

Modalité sensorielle perçue	Signaux *proto-ajustés* et *ajustés*	Sens primaire	Sens secondaire	Négation	Négociation	Enfant	Juvénile	Adolescent
Tactile	• *Grab*	« Arrête ça »	*na*	+		|		
	• *Hit other*	« Arrête ça »	« Va-t'en »	+		|		
	• *Push*	« Arrête ça »	« Va-t'en »	+		|		
	• *Stomp 2-feet other*	« Arrête ça »	« Suis-moi »	+	(+)	|		
	• *Push (directed)*	« Repositionne-toi »	« Rapproche-toi »		+	|	+	
Audible	• *Hit object/ground*	« Va-t'en »	« Arrête ça »	+				+
	• *Stomping 2 feet object ground*	« Arrête ça »	« Suis-moi »	+	(+)			+
Silencieux-visuels	• *Raise arm*	« Je veux cet objet »	« Va-t'en »	(+)	+		|	
	• *Reach*	« Je veux cet objet »	« Viens »		+		|	+
	• *Present (directed)*	« Epouille-moi plutôt par ici »	« Epouille-moi »		+		|	+

Note : Les gestes proto-ajustés sont indiqués par « \| » et les gestes ajustés par « + » .

Concernant l'ajustement des gestes *audibles*, ils semblent davantage être facilités par l'intégration d'un objet dans la communication. Comme on peut le voir **Figure 1**, la préférence initialement tactile devient audible dans le cadre de l'ajustement *intermodal* chez les chimpanzés adolescents par l'ajustement de deux gestes produits avec objet (i.e. *hit object/ground*, *stomping-2-feet object/ground*). Toutefois, des analyses complémentaires ont montré que les adolescents produisaient davantage de gestes audibles sans objet (i.e. *big loud scratching*, n=22/30 gestes audibles) qu'avec objet (voir Annexe 7, Table S2). Ces analyses montrent également que l'intégration de l'objet dans les ajustements se faisaient déjà avant chez les juvéniles alors qu'à ce stade seul un geste tactile sans objet (i.e. *push(directed)*) était ajusté et que la majorité des gestes produits avec objet était audibles et non-ajustés (voir **Figure 1**). Ces résultats suggèrent ainsi que l'intégration d'un objet dans l'interaction pourrait avoir été nécessaire à l'ajustement des gestes audibles

des jeunes chimpanzés à l'état d'attention visuel maternel. Ces résultats pourraient être rapprochés de certaines théories émergentes concernant l'émergence de la communication intentionnelle chez l'enfant humain (occidental). Ces théories suggèrent que l'émergence de la communication intentionnelle soit dépendante de l'intégration physique de l'objet dans la communication gestuelle (Liszkowski & Rüther, 2021; Rodríguez et al., 2015). Ces théories avaient toutefois été construites sur la base des gestes déictiques *proximaux* (i.e. gestes pour lesquels l'objet de référence est manuellement tenu par l'enfant, par exemple le *showing* ou le *giving*) silencieux-visuels et ignoraient ainsi la dimension audible que ceux-ci peuvent avoir (ce point sera discuté dans les parties 6.2.2.2, 6.4.1.2 et 6.5.2.2).

Cette préférence pour l'ajustement *intermodal* des signaux *audibles* ne serait pourtant que transitoire et/ou propre à l'interaction avec la mère. En effet, les études observationnelles réalisées sur des chimpanzés majoritairement adultes (voir partie 1.3.1.2) ne montrent pas d'ajustement des gestes audibles à l'attention visuelle du destinataire conspécifique (Hobaiter & Byrne, 2011; Pika & Liebal, 2006; Tomasello et al., 1994, 1997). En revanche, certaines d'entre elles montrent un ajustement des gestes tactiles (Hobaiter & Byrne, 2011; Tomasello et al., 1994). Ces résultats suggèrent que la préférence audible pour l'ajustement *intermodal* chez l'adolescent serait transitoire. Elle pourrait montrer un intérêt développemental propre à cette tranche d'âge. En effet, la période adolescente est une période pendant laquelle le jeune chimpanzé va drastiquement s'émanciper de sa mère par une augmentation de la fréquence des échanges avec d'autres partenaires que lui (Pusey, 1990; Stumpf et al., 2009; Watts & Pussey, 1993). Dans notre échantillon, le sous-échantillon de chimpanzés adolescents est majoritairement représenté par des femelles (N=3/4) et le seul mâle adolescent est celui qui produit le moins de signaux audibles du groupe. Par ailleurs, on sait que dans les communautés de chimpanzés, les femelles adolescentes émigrent entre 12 et 13 ans (Stumpf et al., 2009). Peu d'études n'ont véritablement élucidé les raisons et/ou les modalités poussant la femelle à émigrer mais il semble que plus que pour des raisons biologiques (e.g. âge des premières chaleurs, risque de consanguinité, niveau de

stress), ce phénomène d'émigration soit prédit par les interactions sociales la précédant (Stumpf et al., 2009). La femelle adolescente aura ainsi tendance à interagir de plus en plus avec les mâles adolescents et adultes et de moins en moins avec sa mère entre 1 et 2 ans avant son départ du groupe (Stumpf et al., 2009). Ces résultats pourraient expliquer les nôtres. En effet, durant cette période développementale, la femelle chimpanzé adolescente s'investissant donc moins dans les échanges avec sa mère qu'avec des partenaires non-maternels (qui représenteraient des enjeux sociaux importants avant l'émigration), pourrait avoir tendance à utiliser des signaux moins coûteux (e.g. ne nécessitant pas de déplacement) tels que les gestes audibles (contrairement aux gestes tactiles).

6.2.2. **Chez l'enfant préverbal** : une histoire **d'ajustement transitoire et d**e signaux-leviers

6.2.2.1. ***L'histoire des ajustements à l'attention visuelle***

Figure 2, pour identifier les signaux *proto-ajustés* chez l'enfant préverbal, la méthode était descriptive à partir des graphiques et des valeurs de probabilités (partie 4.2). Pour l'ajustement *unimodal*, les signaux silencieux-visuels proto-ajustés étaient moins produits en condition d'inattention qu'en condition d'attention visuelle. Pour l'ajustement *intermodal*, les signaux proto-ajustés étaient les signaux audibles qui en proportion (par rapport aux signaux silencieux-visuels) étaient plus produits condition d'inattention visuelle qu'en condition d'attention visuelle. Etaient aussi considérés comme proto-ajustés les signaux dont la répartition était significativement différente d'une condition visuelle à l'autre (e.g. discrimination du babillage dans la partie 4.2) mais avec le pattern inverse (i.e. qui en proportion -par rapport aux signaux silencieux-visuels- étaient plus produits condition d'attention visuelle qu'en condition d'inattention visuelle).

Figure 2. Carte développementale de l'ajustement *unimodal* et *intermodal* des signaux de l'enfant préverbal âgé de 7 à 20 mois. Les gestes *proto-ajustés* sont des signaux qui avant émergence montrent une forme partielle ou inachevée d'ajustement. *(1) le Langage Adressé à l'Enfant (LAE) ; (2) les signaux ajustés à condition d'être associé à un regard vers le destinataire.*

Comme illustré sur la Figure 2, chez les enfants humains préverbaux la capacité d'ajustement *unimodal* des signaux silencieux-visuels émergerait dès l'âge de 7-8 mois alors que celle d'ajustement *intermodal* émergerait à l'âge de 11-14 mois (voir Article 2, partie 4.1). En s'intéressant à la nature des signaux ajustés au cours du développement, on constate qu'autour de l'émergence des capacités d'ajustement, l'évolution de ces signaux est différente selon l'ajustement considéré (voir Figure 2). Pour l'ajustement *unimodal*, on observe un pattern de développement plutôt « sélectif », c'est-à-dire que parmi les signaux par lesquels cet ajustement émerge

(i.e. *mouvements corporels* et *gestes déictiques*), seuls certains continueront à être ajustés au cours du temps (i.e. plus que les *gestes déictiques* ; voir Figure 2 et partie 4.2). Pour l'ajustement *intermodal*, on observe un pattern de développement plutôt « additif », c'est-à-dire que les signaux par lesquels cet ajustement émerge (i.e. *gestes audibles avec objet* et *babillage*) continueront à être ajustés au cours du temps et les autres signaux auparavant non-ajustés s'ajusteront par-dessus (i.e. les *vocalisations non-consonantiques* et les *bruits de bouche* ; voir Figure 2 et partie 4.2). Par ailleurs, que ce soit pour les signaux silencieux-visuels ou audibles, l'ajustement des signaux produits avec objet pris à part (i.e. gestes déictiques *proximaux* et *gestes audibles avec objet*) s'expriment tous deux à partir de l'âge de 11-14 mois, c'est-à-dire au moment où l'ajustement *intermodal* émerge. Ces premiers constats suggèrent que, bien que (i) ces capacités d'ajustement suivent des trajectoires développementales qui leurs sont propres, (ii) elles pourraient s'influencer l'une l'autre par l'intégration de l'objet lors de ces ajustements.

Concernant l'ajustement *unimodal,* les *mouvements corporels* associés à un regard qui étaient ajustés à 7-10 mois cessent de l'être à partir de 11-14 mois contrairement aux gestes déictiques. Soulignons que ces *mouvements corporels* ne s'ajustent qu'à condition d'être associés à un regard vers le destinataire. Par ailleurs, ces signaux émergent relativement tôt dans le développement, avant 5 mois (Nakata & Trehub, 2004; Scola et al., 2015) bien avant l'émergence des gestes déictiques à 8-9 mois (Bates et al., 1975; Cameron-Faulkner et al., 2015; Choi et al., 2021). De plus, ils sont plus produits que les gestes déictiques par les enfants de 7-10 mois dans notre échantillon (n=181/225 soit 80,44%) alors que la tendance est très vite inversée plus tard (**11-14 :** n= 65/203 soit 32,02% ; **15-20 :** n=28/78 soit 35,90% ; voir article 3, partie 4.2). Du fait de son émergence relativement précoce et de sa fréquence d'utilisation avant 11 mois, ils pourraient ainsi avoir servi d'amorce à l'ajustement des signaux silencieux-visuels. Toutefois, dès lors que les gestes déictiques sont plus fréquemment utilisés que ces mouvements corporels, ces signaux d'amorce ne s'ajustent plus. Une explication pourrait être que tout l'investissement nécessaire à l'optimisation de la signalisation (notamment par

l'ajustement) est concentré sur les gestes déictiques dont l'effet perlocutoire surpasserait celui des mouvements corporels (voir partie 4.2). Plusieurs études montrent que les gestes déictiques suscitent une responsivité parentale vocale vers 11-12 mois dans la mesure où ces gestes seraient plus souvent interprétés par l'adulte comme étant de la communication (e.g. Donnellan et al., 2019; Wu & Gros-Louis, 2015). Certains auteurs montrent que les parents répondent davantage au geste déictique (combiné à des vocalisations ou non) qu'aux vocalisations seules à l'âge de 12 mois (Wu & Gros-Louis, 2015), mais que globalement ils répondront nettement plus à des signaux (vocaux et/ou gestuels) associés à un regard vers le destinataire à 11 mois (Donnellan et al., 2019). A notre connaissance, aucune étude n'a encore comparé l'effet perlocutoire (i.e. quantifier la responsivité du parent vocale et/ou autre) des mouvements corporels à ceux des gestes déictiques pour voir s'il existait une corrélation entre la fréquence d'utilisation du signal et la responsivité parentale. Si c'était le cas, il serait intéressant de voir s'il existait une corrélation entre la responsivité parentale et la qualité de l'ajustement de ce signal. Dans la continuité de cela, parce que l'ajustement des mouvements corporels était dépendant du regard, le regard pourrait avoir joué un rôle amplificateur de cet effet perlocutoire (Donnellan et al., 2019) motivant l'ajustement.

Concernant l'ajustement *intermodal*, l'ajustement des gestes *audibles avec objet* et le *babillage* pourrait avoir servi d'amorce à l'ajustement des autres signaux audibles. Comme discuté partie 4.2, le babillage est la vocalisation prélinguistique la plus sophistiquée en comparaison aux autres vocalisations (Bloom & Beaumont, 1993; Bloom & Lo, 1990; Gros-Louis et al., 2014; Locke, 2006). Entraînant plus de responsivité parental, le babillage serait de fait soumis à un effet perlocutoire facilitant l'accès à l'illocution (i.e. communication intentionnelle) puis à la locution (i.e. communication verbale) de par son pouvoir de prédiction langagière (Donnellan et al., 2019; Jensen et al., 1988; Oller et al., 1998; Oller et al., 1999; Stoel-Gammon, 2011; Stolt et al., 2012). Une fois ajustée, le babillage pourrait avoir joué un rôle de *signal-levier* nécessaire à l'ajustement des autres sons oraux à savoir les *vocalisations non-consonantiques* et les *bruits de bouche* sujets à moins de responsivité

parentale. Des analyses complémentaires seraient nécessaires pour vérifier cette hypothèse.

Par ailleurs, les signaux audibles ajustés ont tous été proto-ajustés (i.e. ont montré une forme partielle ou inachevée d'ajustement.) au préalable, excepté les *bruits de bouche* alors qu'ils sont produits en proportion relativement constante et importante au cours du développement (entre 16 et 21% sans LAE ; voir Annexe 7, Table S2). Ce qui distingue les bruits de bouche des autres signaux oraux est qu'ils ne *respectent* pas les règles infra-phonologiques des autres protophones (ni son vocalique, ni son consonantique). Les nourrissons utilisent leurs lèvres (ou une combinaison de la langue et de lèvres) pour produire des sons de friction bruyants (i.e. *raspberries*; Oller, 2000). Ils apparaissent au stade d'extension du répertoire vocal durant lequel les enfants sont clairement dans une phase d'exploration du contrôle qu'ils peuvent avoir de leurs lèvres et de leur voix. Pourtant, peu d'études se sont attelées à explorer leur utilisation alors qu'ils sont porteurs de nouvelles caractéristiques de friction (Oller, 2000) et que l'enfant humain de 6 mois semble lui attribuer une valeur sociale (Bourjade et al., *soumis*). Bien qu'à notre connaissance ils aient été peu étudiés dans l'étude de la transition vers le stade linguistique chez l'enfant, il semble toutefois que certaines études sur le primate non-humain captif suggèrent qu'ils puissent constituer un précurseur potentiel du langage humain (Hopkins et al., 2007; Lameira et al., 2013; Meguerditchian et al., 2014; Taglialatela et al., 2008). C'est pourquoi il serait intéressant de faire des analyses complémentaires quant aux fonctions qu'ils servent en comparaison aux autres signaux oraux. Ce manque d'intérêt pour l'étude des bruits de bouche chez l'enfant est également visible pour les gestes audibles impliquant notamment un objet que nous discutons ci-après.

L'ajustement unimodal et *intermodal* des signaux intégrant un objet émergent en même temps, suggérant ainsi que l'action instrumentale puisse constituer un lien entre la communication visuelle et audible voire vocale. Certains auteurs suggèrent que l'intégration d'un objet dans l'interaction via l'utilisation de gestes déictiques *proximaux* (i.e. gestes pour lesquels l'objet de référence est tenu par le signaleur tels

que le *showing* ou le *giving/receiving*) soit nécessaire à l'utilisation intentionnelle du pointage distal (Liszkowski & Rüther, 2021; Rodríguez et al., 2015). Sur cette base, certaines études récentes ont montré que ces gestes déictiques *proximaux* produits avant l'âge de 14 mois constituaient de bons prédicteurs de production de gestes de pointage (Cameron-Faulkner et al., 2015) et de compréhension et de production langagières (Choi et al., 2021). Toutefois, comme discuté dans la partie 4.2, ces auteurs n'appréhendaient les gestes avec objet qu'à travers les gestes déictiques *proximaux* (i.e. *giving, showing*) donc silencieux-visuel, laissant inexplorée la dimension acoustique du geste dans leurs investigations. Au regard de nos résultats et des questionnements quant à la transition modale entre le geste déictique (impliquant de fait un objet de référence) à la parole chez l'enfant humain, il serait possible que le geste audible *avec* objet, de par son canal de production gestuel et sa perceptibilité audible, pourrait constituer *un* signal « pivot » moteur de la transition du geste à la vocalisation.

Les travaux actuels défendent toutefois l'idée que cette transition du geste à la vocalisation soit davantage facilitée par la cooccurrence de production du geste (souvent déictique) et de la vocalisation (Aureli et al., 2017; Fasolo & D'Odorico, 2012; Iverson & Fagan, 2004; Iverson & Goldin-meadow, 2005; McCune et al., 2020; Murillo & Casla, 2020; Murillo et al., 2018). Toutefois, aucune de ces études ne considérait (i) le geste audible dans leurs analyses et (ii) son rôle potentiel dans cette transition des signaux gestuels aux signaux vocaux. En effet, peu d'études sur le développement communicationnel de l'enfant humain considéraient les gestes audibles comme un signal de communication à part entière (Bourjade et al., *soumise*; Kersken et al., 2018; Rodrigues et al., 2021) et d'autres études intégraient des gestes audibles dans leurs analyses sans considérer leur dimension audible (*rythmic movements with/without object*; Murillo et al., 2019). De façon intéressante, cette dernière étude montrait que les mouvements rythmiques étaient préférentiellement produits seuls que combinés à des vocalisations chez les enfants de 9 à 12 mois et qu'ils étaient préférentiellement produits avec objet que sans (Murillo et al., 2019). Cette étude montrait également une corrélation positive entre la production de ces

mouvements rythmiques *avec* objet combinés à des vocalisations et la production de gestes déictiques proximaux (i.e. gestes silencieux-visuels produits avec objet) à 9 mois (Murillo et al., 2019). Ces résultats combinés aux nôtres supportent l'idée que l'étude de la dimension multimodale de l'action instrumentale pourrait davantage nous renseigner sur la transition du geste à la vocalisation.

Cette transition modale pourrait alors se faire selon plusieurs étapes dont l'enchaînement serait fonction de la modalité du geste combiné aux vocalisations et des capacités d'ajustement. Nos résultats ont montré que seule la production de combinaisons *distales* (i.e. sans signaux tactiles, donc uniquement composées de signaux audibles et silencieux-visuels) était fonction de l'attention visuelle maternelle (voir partie 4.1). Quelle que soit la tranche d'âge, les enfants produisaient plus de combinaisons audio-visuelles quand la mère était visuellement attentive que quand elle ne l'était pas. Toutefois, les combinaisons audio-visuelles considérées pouvaient aussi bien être composées de gestes audibles que de signaux oraux (en plus des signaux silencieux-visuels ; voir partie 4.1). De plus, du fait d'avoir centré nos investigations sur la contrainte perceptive des ajustements à l'attention visuelle maternelle, les combinaisons de plusieurs signaux ayant une seule modalité sensorielle de signal (e.g. geste audible combiné à des vocalisations) n'ont pas été considérées dans les analyses. C'est pourquoi des analyses complémentaires sur la nature des signaux impliqués dans ces combinaisons seraient nécessaires. Elles nous permettraient de tester l'effet de l'âge sur la nature des signaux combinés aux vocalisations et retracer évolution au cours du développement.

6.2.2.2. *Effet du Langage Adressé à l'Enfant (LAE) sur les ajustements*

Du fait de la multi-sensorialité de l'attention maternelle perçue par l'enfant (Esteve-Gibert & Guellaï, 2018b; Guellaï et al., 2019; Guellai & Streri, 2011), il s'agissait de s'interroger sur l'influence de variables attentionnelles telles que le contact physique ou le LAE sur les capacités d'ajustement des enfants humains

(voir partie 4.1). Il en résultait donc que le LAE et non le contact physique influençait significativement ces capacités d'ajustement à l'attention visuelle maternelle (voir article 2, partie 4.1).

Sur la Figure 3 ci-dessous, les trajectoires développementales des capacités d'ajustement *unimodal* et *intermodal* de l'enfant préverbal ont été représentées en fonction de la présence ou non de LAE maternel. L'ajustement est décliné selon quatre niveaux d'acquisition : « non-acquis » (i.e. le descripteur n'est pas statistiquement validé), « en cours d'acquisition » (i.e. le pattern attendu commence à être visible mais le descripteur n'est toujours pas statistiquement validé), « acquis » (i.e. le descripteur est statistiquement validé) et « maîtrisé » (i.e. le descripteur est statistiquement validé et il est significativement meilleur qu'en *simple* niveau « acquis »).

Pour pourvoir représenter l'effet du LAE, il aura fallu s'appuyer sur les résultats des analyses inférentielles réalisées dans l'article 2 comparant les ajustements en condition avec et sans LAE à chaque tranche d'âge (voir partie 4.1).

Figure 3. Trajectoires développementales des capacités d'ajustement *unimodal* et *intermodal* en fonction du LAE maternel chez l'enfant préverbal. Le niveau « acquis » est le niveau minimum pour lequel le descripteur a été validé.

Comme illustré sur la Figure 3, l'ajustement *unimodal* et l'ajustement *intermodal* émergeraient (i.e. les courbes dépassent le niveau « acquis ») aux tranches d'âge respectives de 7-10 mois et 11-14 mois, que la mère s'adresse à l'enfant (courbe verte) ou non (courbe noire). Quand la mère s'adressait à son enfant, l'ajustement *unimodal* était plus marqué que quand elle était silencieuse jusqu'à l'âge de 15-20 mois. En revanche, pour l'ajustement *intermodal*, ce pattern était inversé puisque l'ajustement était plus marqué quand la mère était silencieuse que quand elle s'adressait à son enfant. Toutefois, durant cet ajustement *intermodal*, quand leur

mère était visuellement inattentive, les enfants préféraient utiliser des gestes audibles que des signaux oraux quand leur mère était silencieuse, alors qu'ils préféraient les signaux oraux aux gestes audibles quand leur mère s'adressait à eux. En somme, on pourrait dire que le LAE (i.e. le fait que la mère s'adresse à son enfant) influence quantitativement (i.e. plus de signaux ajustés en proportion) l'ajustement *unimodal* et qualitativement (i.e. préférence orale donc de forme plus proche de la forme finale du langage) l'ajustement *intermodal* (voir partie 4.1). De plus, des analyses complémentaires quant à la nature des signaux via lesquels émergent ces capacités d'ajustements quand la mère s'adresse à son enfant ont été réalisées (voir annexe 7, Table S3 et Figure S1). Ces analyses complémentaires montrent que le LAE améliorerait l'ajustement *unimodal* sur l'ensemble des signaux silencieux-visuels pris isolément. *A contrario*, bien qu'en situation d'inattention visuelle le LAE inciterait l'enfant à produire des vocalisations plus que des gestes audibles, aucune de ces vocalisations ne s'ajustent à l'attention visuelle.

Une explication possible à ce pattern relèverait de la fonction de régulation émotionnelle portée par le LAE qui en condition d'inattention suggère que l'enfant ne ressente pas le besoin de maintenir ou regagner l'attention maternelle. En effet, en admettant que l'inattention visuelle constitue une rupture attentionnelle qui pourrait produire chez l'enfant de l'agitation, la présence de LAE pourrait apaiser ou maintenir l'attention de l'enfant au point qu'il inhibe la production de l'ensemble des signaux. Cette inhibition de la signalisation aurait donc pour conséquence d'accentuer le pattern d'ajustement *unimodal* (qui se caractérise de fait par une inhibition des signaux silencieux-visuels quand la mère est visuellement inattentive). En revanche l'inhibition des signaux audibles aurait pour conséquence de s'éloigner du pattern comportemental attendu pour l'ajustement *intermodal* (i.e. pour laquelle la situation d'inattention visuelle devrait encourager la production de gestes audibles en comparaison aux signaux silencieux-visuels). Ces résultats seront discutés au regard de l'hypothèse de Falk (2004), partie 6.4.2.1.

L'effet du LAE sur l'ajustement *intermodal* questionne quant à la pertinence du descripteur en contexte de LAE. En effet, aucune des vocalisations ne s'ajuste

véritablement à l'attention visuelle, elles sont pour la plupart moins produites que les signaux silencieux-visuels et il existe une grande variabilité interindividuelle quel que soit le groupe d'âge considéré (voir Annexe 7, Figure S3). Or, on sait qu'en condition sans LAE les signaux audibles (gestuels ou oraux) s'ajustent, donc ces mêmes enfants sont capables d'ajustement *intermodal*. De ce fait, la présence de LAE pourrait rendre le descripteur caduc. Au regard du réagencement modal, on pourrait imaginer que son rôle serait avant tout d'inciter l'enfant à produire des vocalisations. Cependant, du fait qu'il ait un effet sur l'ajustement de signaux autres que vocaux -à savoir silencieux-visuels- son rôle serait plausiblement plus large que seulement vocal. Il serait donc intéressant, dans la continuité du principe d'ajustement *intermodal,* de comprendre comment le LAE aide à l'articulation des différentes modalités de signaux produites lors de futures investigations.

6.3. L'AJUSTEMENT DE LA COMMUNICATION A L'ATTENTION VISUELLE DE LA MERE PREDIRAIT DES COMPETENCES LANGAGIERES ULTERIEURES

Comme développé dans la partie 5, bien que la *communication sensible à l'attention du destinataire* constitue un marqueur d'intentionnalité conventionnellement étudié chez l'enfant humain préverbal (Gros-Louis & Wu, 2012; Igualada et al., 2015; Liszkowski et al., 2008; Rodrigues et al., 2021; Wu & Gros-Louis, 2014, 2017), peu d'études ont montré voire testé son potentiel de prédiction langagière (Wu & Gros-Louis, 2014, 2017). C'est pourquoi, il s'agissait dans le cadre de ce travail, à partir des descripteurs d'ajustement *unimodal* et *intermodal* de tester leur potentiel de prédiction. Il en résulte que l'ajustement *unimodal* prédirait des compétences lexicales mais l'ajustement *intermodal* ne prédirait pas de compétence langagières (discuté article 4, partie 5). Malgré de notables limites méthodologiques, cette première investigation de l'effet de

prédiction des ajustements suggère un certain nombre de points méthodologiques et théoriques à considérer pour de futures analyses.

6.3.1. Apports méthodologiques

L'élaboration de cette dernière étude (partie 5) sur l'effet de prédiction des ajustements propose des outils qui pourront servir l'approche développementale dans de futures investigations. Rappelons que jusqu'alors les études ayant investigué la *communication sensible à l'attention du destinataire* montraient de nombreuses divergences méthodologiques limitant la comparabilité des résultats existants. A cela, nous avions proposé les deux descripteurs d'ajustement *unimodal* et *intermodal*. Dans les parties 3 et 4, nous avons montré que la distinction de ces capacités sous-jacentes à la *communication sensible à l'attention du destinataire* était pertinente pour l'approche développementale et permettait de construire une histoire développementale (voir partie 6.2). Dans la partie 5, les résultats sur l'effet de prédiction suggéraient une fois de plus que ces capacités servaient des fonctions bien distinctes au regard de l'émergence du langage chez l'enfant préverbal. Toutefois, il semble important de souligner que cette dernière étude présente de nombreuses limites méthodologiques qui limitent la puissance statistique (e.g. petitesse de l'échantillon, latence entre observation et questionnaire variable). De ce fait, nos résultats sont à considérer avec une certaine nuance mais le *process* méthodologique nous ayant permis d'y arriver aura permis la création d'outils pouvant servir l'approche développementale. Par exemple, les scores d'ajustements (i.e. UNIM et INTERM1/INTERM2) élaborés pour cette étude pourraient constituer de bons outils de modélisation pour les études longitudinales. En effet, ceux-ci permettaient de quantifier et de graduer la performance, ce qui au-delà des tests de prédiction langagière pourrait faciliter le suivi longitudinal d'un individu en développement (humain ou non-humain) ou soumis à un entraînement comportemental (e.g. thérapeutique chez l'enfant humain).

6.3.2. Apports et questionnements théoriques

Par ailleurs, les résultats de cette première étude soulèvent divers points de discussion quant à l'appréhension des futures études développementales. D'une part, on remarque que les variables influençant le plus les scores de prédictions lexicales et syntaxiques sont les variables d'ajustement de signaux *non-canoniques* (i.e. des comportements présentant un intérêt dans le développement communicationnel de l'enfant mais non considérés comme prédicteurs langagiers tels que les *mouvements corporels* ou les *gestes audibles avec un objet*). Parfois même, ces modèles surpassent ceux composés de variables relatives aux signaux *canoniques* (i.e. des comportements prédicteurs du langage tel que le geste de pointage distal). Nos résultats incitent donc les futures investigations à considérer davantage de signaux qui émergeraient de façon plus précoce (e.g. les mouvements corporels, les bruits de bouche) que d'autres beaucoup plus étudiés (e.g. babillage et gestes déictiques, voir détail parties 1.4.2.1.1). D'autre part, la mise en corrélation des performances des ajustements (i.e. d'ordre pragmatique) et des scores de performance langagière d'ordre lexical voire syntaxique (discuté article 5) soutient fortement l'idée de co-dépendance de ces trois aspects du langage dans le développement de l'enfant.

Si on défend l'idée d'une transition modale conduisant à un usage vocal (donc audible) dominant par la parole alors pourquoi le seul ajustement qui va dans le sens d'une prédiction langagière est celui qui implique exclusivement des signaux silencieux-visuels ? Une explication possible est que ce résultat pourrait appuyer l'idée que le statut d'*amorce* à l'ajustement, plus que la nature même du signal ajusté, confère à celui-ci une importance capitale dans le développement communicationnel. En effet, nos résultats (voir partie 4 et illustré par la Figure 2), montrent que finalement le seul ajustement le plus plausiblement prédicteur langagier est un des premiers de la « chaîne » des ajustements, i.e. celui des mouvements corporels associés au regard émergeant à 7-10 mois. De plus, c'est un des seuls ajustements à exclusivement dépendre de l'association du regard. Or, comme discuté partie 4.2, l'association du regard aux signaux n'améliore pas nécessairement les performances aux ajustements chez les enfants préverbaux plus

âgés. Ces résultats suggéraient que la nécessité de coordonner son regard au signal produit pour l'ajuster à l'attention du destinataire soit une stratégie d'optimisation de l'échange propre à des enfants très jeunes. L'utilisation du regard par la suite serait moins nécessaire à l'optimisation de l'échange et donc moins systématiquement associé à la signalisation de l'enfant. Comme à l'image d'un *programme* que l'on démarre, on pourrait imaginer que c'est le statut d'*amorce* du comportement, plus que la nature du comportement en soi, qui lui confère un pouvoir de prédiction. Une autre explication pourrait être le manque de puissance statistique de l'échantillon (i.e. grande variabilité des données à l'intérieur de l'échantillon manipulé et de sa petitesse ; voir partie 5). Ainsi, avec davantage d'individus et un plus grand contrôle des différences de temporalité de collecte, il n'est pas impossible qu'un ajustement *intermodal* (arrivant après l'ajustement *unimodal*) puisse révéler un potentiel de prédiction langagière.

Ces résultats peuvent également être rapportés à l'approche comparative. En effet, l'approche *evo-devo* défend un apport mutuel de la psychologie du développement et la psychologie évolutionnaire (Bjorklund & Pellegrini, 2000). De ce fait, investiguer les prédicteurs du langage chez l'enfant observé en condition écologique continue de nous renseigner quant aux précurseurs potentiels du langage humain. Construits sur la base des études sur la communication intentionnelle chez l'enfant humain (Bates et al., 1975; Beuker et al., 2013; Bruner, 1975, 1981; Liszkowski et al., 2004) et des espèces non-humaines (Fröhlich et al., 2018; Genty et al., 2009; Leavens et al., 2005; Townsend et al., 2017) les descripteurs de prédiction ici testés sont opérationnels sur l'étude d'espèces humaines et non-humaines en condition écologique. De ce fait, si les ajustements (nos descripteurs) sont retrouvés chez les deux espèces étudiées ici, alors pourquoi l'une mène au langage et l'autre non ? Cela fera l'objet des points discutés ci-après.

6.4. MISE EN PERSPECTIVE COMPARATIVE ET CONFRONTATION DES RESULTATS AUX THEORIES DU CONFLIT ATTENTIONNEL

6.4.1. Confrontation des résultats en inter-espèce

Dans cette thèse il s'agissait de construire une histoire développementale de la *communication sensible à l'attention du destinataire* propre à chaque espèce de primate étudiée ici (i.e. chimpanzé ou humain) à partir de l'étude des capacités d'ajustement à l'attention maternelle (i.e. *unimodal* et *intermodal* ; voir Figure 1 et 2). Bien que la comparabilité de ces deux populations ne soit pas facilitée du fait de certaines divergences méthodologiques (voir parties 2,3 et 4), à partir d'une première comparaison nous discuterons de (i) l'ordre d'émergence de ces ajustements à l'attention du destinataire et de (ii) l'importance de l'objet dans le développement de ces capacités ajustements.

6.4.1.1. *Ordre différentiel d'émergence des ajustements d'une espèce à l'autre*

Chez le jeune chimpanzé, l'ajustement *intermodal* émerge avant l'ajustement *unimodal* alors que chez l'enfant préverbal, l'ordre d'émergence est inversé (i.e. ajustement *unimodal* puis *intermodal*). S'interroger sur ce point pourrait revenir à se demander pourquoi le jeune chimpanzé ajustait ses gestes tactiles *avant* ses gestes silencieux-visuels à l'attention visuelle alors que l'enfant ajustait ses gestes silencieux-visuels *avant* ses signaux audibles (gestuels et oraux). Une explication possible relèverait du souci d'*efficacité* du signal qui va de pair avec celui de la responsivité maternelle.

En effet, le jeune chimpanzé ajustait ses gestes tactiles *avant* ses gestes silencieux-visuels à l'attention visuelle (voir Figure 1). Certaines études ont montré que chez le chimpanzé immature, les premiers signaux à émerger dans le répertoire comportemental étaient tactiles et silencieux-visuels (autour de 10 mois) avant l'émergence des gestes audibles (autour de 18 mois ; Schneider et al., 2012). Puis,

les jeunes chimpanzés utilisaient de moins en moins de gestes tactiles et de plus en plus de communication distale (i.e. gestes silencieux-visuels et audibles ; Schneider et al., 2012) avec l'âge mais préféraient davantage utiliser des gestes tactiles avec leur mère (Fröhlich et al., 2016a). Comme discuté précédemment (voir partie 6.2.1), si le premier ajustement émerge à travers l'enjeu de la *négociation*, le jeune chimpanzé préfèrerait utiliser un signal qu'il maîtrise depuis plus longtemps (donc *a priori* mieux) pour optimiser ses chances d'atteindre son but de communication (i.e. convaincre sa mère). Toutefois, cela pourrait aussi provenir de la mère qui serait plus répondante au signaux tactiles qu'aux autres signaux. Dans ce cas cette préférence pour l'ajustement tactile résulterait de l'expérience de la responsivité maternelle. Des analyses complémentaires ont montré qu'en effet, parmi les gestes tactiles produits par le chimpanzé juvénile, le *push(directed)* (i.e. le geste par lequel émerge l'ajustement intermodal, voir Figure 1), faisait partie des deux gestes tactiles (avec le *Touch* qui est aussi un geste de négociation) pour lesquels la mère donnait le plus haut taux de réponses satisfaisantes (*push(directed)* : 17% ; *Touch* : 19%). Ce résultat semblerait donc plausible avec l'hypothèse générale de souci d'efficacité du signal. Cependant, il ne nous renseigne pas quant à la raison pour laquelle il serait efficace (mobilisation d'une meilleure stratégie de communication chez l'enfant ou plus de de responsivité maternelle). Des analyses complémentaires seront donc nécessaires pour tester ces hypothèses.

Par ailleurs, l'enfant préverbal ajustait ses gestes silencieux-visuels *avant* ses signaux audibles (gestuels et oraux ; voir Figure 2). Comme discuté précédemment (voir partie 6.2.2), il semble probable que la responsivité parentale en soit la cause. Des études ont montré que les gestes déictiques combinés ou non avec des vocalisations, plus que les vocalisations produites seules suscitaient plus de réponse parentale (Wu & Gros-Louis, 2015). De plus, de manière générale, la coordination du regard vers le destinataire à un signal produit suscitait également plus de réponse parentale (Donnellan et al., 2019; Northrup & Iverson, 2020b). Ces deux résultats semblent pouvoir expliquer pourquoi les gestes silencieux-visuels testés seraient ajustés *avant* les signaux audibles. Par ailleurs, il est possible que cette responsivité

aux gestes déictiques notamment, en comparaison aux vocalisations notamment, soit causée par le fait que le geste déictique soit plus facilement interprétable (moins de possibilité de sens à attribuer) par le parent qu'un son oral ou un geste audible soumis à davantage de flexibilité de sens. En effet, parmi les gestes déictiques (i.e. *giving*, *receiving*, *showing*, *reaching* et gestes de pointage), l'effet perlocutoire est probablement plus facilité que pour les gestes audibles ou les vocalisations que l'adulte pourrait plus facilement qualifier d'exploratoires que communicatives. A la manière de l'étude réalisée sur la communication gestuelle du chimpanzé avec le système ASO (*apparently satisfied outcome*; Hobaiter & Byrne, 2014), il aurait probablement été facile de tester cette hypothèse.

6.4.1.2. *Intégration de l'objet dans le geste : moteur de l'ajustement intermodal*

Comme discuté précédemment (voir partie 6.2.2.1), certains auteurs suggèrent que l'intégration de l'objet dans l'interaction pourrait avoir contribué à l'émergence de l'utilisation intentionnelle du geste de pointage, à travers l'utilisation des gestes déictiques *proximaux* (Liszkowski & Rüther, 2021; Rodríguez et al., 2015). On aurait donc pu s'attendre à ce que les gestes déictiques proximaux s'ajustent avant ceux distaux mais aucun de nos résultats sur l'ajustement *unimodal* ne montre des résultats allant véritablement dans ce sens.

Toutefois, en élargissant la définition du geste avec objet (*proximal*) aux gestes audibles, nos résultats montrent que l'ajustement du geste avec objet (i.e. tactile et audible pour le chimpanzé, audible et silencieux-visuel pour l'enfant humain) marque l'émergence des capacités d'ajustement *intermodal* pour chacune des espèces. De façon intéressante, ces *gestes audibles avec objet* codés chez l'enfant recouvrent en grande partie les gestes audibles recueillis chez le chimpanzé (voire Table 2). On remarque par ailleurs que les gestes relevés dans les analyses sur les enfants préverbaux (voir partie 4) sont très comparables à ceux retrouvés chez les

chimpanzés étudiés ici (voir partie 3 et Table 2) et qu'ils recouvrent les trois tranches d'âge.

Table 2. Gestes audibles-ou-tactiles produits avec objet observés pour chaque espèce

		Chimpanzés			**Humains**
	Période développementale	**Enfant**	**Juvénile**	**Adolescent**	**7-20 mois**
Gestes audibles	*Big loud scratch*		+		
	Hit object/ground			+	+
	Object move	+	+		+
	Object shake	+			+
	Stomp 2-feet object/ground	+			
	Stomping 2-feet object/ground			+	
Gestes tactiles	*Hit with object*	+			

Note : le Drumming a été retiré du tableau par ce qu'il n'inclut pas la préhension de l'objet et que dans le codage humain il correspondrait aux gestes audibles sans objet (qui de surcroît ne s'ajustent jamais).

A notre connaissance aucune étude n'a véritablement exploré, en contexte *écologique,* le potentiel communicationnel du geste audible dans une perspective comparative (Dufour et al., 2015; Hattori & Tomonaga, 2020; Hattori et al., 2013), encore moins à travers l'approche développementale. Notons tout de même que certains auteurs ont récemment mené des investigations sur la rythmicité du geste de *Drumming* (i.e. l'action de taper de manière répétée avec force sur un tronc d'arbre) souvent accompagné de vocalisations de *pant hoots* chez le chimpanzé mâle adulte (Babiszewska et al., 2015; Eleuteri et al., 2022). Ces études ont montré qu'il existait une *signature rythmique* (du geste et de sa coordination avec le *pant hoot*) propre à l'identité de l'individu, et que cette signature participait grandement à la dynamique sociale du groupe (Babiszewska et al., 2015; Eleuteri et al., 2022). Dans notre étude, le *drumming* serait qualifié de geste audible *sans* objet (puisque n'impliquait pas la préhension manuelle d'un objet pour produire un son), qui de

surcroît ne s'ajustait pas chez l'enfant préverbal (voir partie 4). Soulignons néanmoins la comparabilité des signaux dont il est question (i.e. gestes audibles et vocalisations). De plus, du fait de leur possible implication dans le réagencement modal de la signalisation de l'enfant préverbal (Murillo et al., 2019; articles 2 et 3, partie 4), notre étude suggèrerait un terrain d'investigation potentiel sur lequel la comparabilité des espèces serait facilitée pouvant servir l'approche *evo-devo*. Ce point sera étayé dans la partie 6.5.2.2.

6.4.2. Théories du conflit : effet du LAE et ajustement du babillage

Comme développé partie 1.1.3, cette thèse s'inscrit dans une approche *evo-devo* permettant d'étudier les théories du conflit attentionnel (voir partie 1.1.2 ; Falk, 2004; Locke, 2006; Mehr & Krasnow, 2017). Il s'agit ici de discuter certains de nos résultats à l'aune de ces théories concernant (i) l'effet du Langage Adressé à l'Enfant (LAE) sur la communication sensible à l'attention maternelle et (ii) l'ajustement du babillage.

6.4.2.1. *Effet du LAE et hypothèse de Falk (2004)*

Rappelons que **Falk (2004)** est à l'origine de l'hypothèse du *Putting-the-baby-down* selon laquelle la parole aurait émergé chez les premiers hominidés pour compenser la diminution de contacts physiques entre la mère et son enfant due à la généralisation de la bipédie. L'éloignement physique de plus en plus fréquent entre la mère et son bébé aurait ainsi poussé cette dernière à vocaliser (sous la forme d'un LAE « primitif ») pour compenser la distance physique et apaiser sa progéniture, aidant ainsi à l'émergence des substrats vocaux du protolangage chez l'humain (Falk, 2004 ; voir partie 1.1.2).

Dans notre échantillon d'enfants français (voir partie 4.1), quand la mère s'adresse à son enfant, l'ajustement *intermodal* est moins marqué que quand elle est silencieuse. Comme discuté dans la partie 6.2.2.2, il semble toutefois que le LAE

améliore qualitativement l'ajustement des signaux oraux quand la mère est visuellement inattentive et inciterait tendanciellement l'enfant à produire davantage de combinaisons *distales* (i.e. exclusivement composées de signaux audibles et silencieux-visuels) quand elle est visuellement attentive (voir partie 4.1). De plus, des études empiriques montrent que la réponse vocale de l'enfant au LAE (Donnelly & Kidd, 2021) et sa capacité à produire des combinaisons *distales* (notamment geste déictiques et vocalisation ; Fasolo & D'Odorico, 2012; Iverson & Goldin-meadow, 2005; Murillo et al., 2018) constituent des prédicteurs langagiers. Ainsi, au regard de l'hypothèse de Falk (2004), lorsque la mère parle, l'effet d'inattention (visuelle) serait moins ressenti par son enfant puisque la mère fournirait une forme d'attention compensatoire distale (que constitue le LAE). L'effet produit sur la signalisation de l'enfant (réponse vocale et production de combinaisons distales) incluant des prédicteurs langagiers chez l'enfant (contemporain), pourrait plausiblement avoir conduit à l'émergence du langage à l'échelle de l'espèce (comme suggéré par Falk, 2004).

Toutefois, Falk (2004) suggérait également que l'absence de contact physique qui aurait été compensée par un LAE « primitif » pourrait avoir influencé la communication de l'enfant. Or notre étude ne montre pas d'effet de la distance physique sur les signaux audibles et silencieux-visuels (voir article 2, partie 4.2) mais un effet sur les signaux tactiles qui ne représentent que 2% des signaux collectés chez l'enfant français (192 sur 9245). Toutefois, notre échantillon n'est aucunement représentatif de l'espèce humaine dans sa globalité (WEIRD; Henrich, et al., 2010). Selon la population humaine considérée, les modalités d'interaction entre parent et enfant sont variables notamment en termes d'importance du contact physique (Keller et al., 2005, 2009), ce point sera discuté dans la partie 6.5.2.2. De ce fait, l'effet de la distance physique maternelle sur la signalisation de l'enfant pourrait donc être retrouvé sur d'autres populations humaines.

6.4.2.2. *Ajustement du babillage et hypothèse de Locke (2006)*

Rappelons que **Locke (2006)** est à l'origine de l'hypothèse de *la sélection parentale* selon laquelle l'utilisation du *babillage* par l'enfant constituerait un indicateur de valeur adaptative (fitness), incitant le parent à poursuivre ses soins parentaux envers son nourrisson donné. Ainsi, les enfants produisant plus de babillage que les autres pourraient bénéficier de plus de réponses parentales, ce qui constitue un investissement parental supérieur. Cette hypothèse souligne l'importance du babillage dans l'interaction parent-enfant. L'hypothèse de Locke est soutenue par un grand nombre de travaux montrant que le babillage suscite plus de responsivité du parent. Par ailleurs, l'attention parentale délivrée soutiendrait le développement vocal de l'enfant nécessaire à l'émergence du langage (Esteve-Gibert & Guellaï, 2018a; Oller et al., 1999; Stoel-Gammon, 2011).

Nos résultats (partie 4.2) ont montré que le babillage se distinguait des autres signaux oraux. Selon la théorie des actes de langage chez l'enfant (Austin, 1962; Bates et al., 1975; Bruner, 1975), l'acquisition des fonctions illocutoires (i.e. ajustement à l'attention maternelle) et locutoires (i.e. langage) dépendent initialement de l'effet perlocutoire (i.e. la responsivité parentale). Dans notre étude, le babillage est le premier son oral à montrer des fonctions illocutoires (i.e. d'ajustement à l'attention visuelle maternelle) et étant, prédicteur du langage, ce babillage va de fait servir des fonctions locutoires (Donnellan et al., 2019; Jensen et al., 1988; Oller et al., 1998; Oller et al., 1999; Stoel-Gammon, 2011; Stolt et al., 2012). En effet, à 7-10 mois, l'ajustement *intermodal* du babillage n'émerge pas encore mais l'attention maternelle visuelle influence d'ores et déjà son utilisation. En admettant que le babillage, plus que les autres sons oraux, suscite plus de responsivité parental (effet *perlocutoire*), l'enfant va dans un premier temps chercher à maximiser cet effet perlocutoire en produisant plus de babillage face au parent visuellement attentif. A cet âge-ci la proportion de babillage (par rapport aux signaux silencieux-visuels) était plus importante quand la mère était visuellement attentive qu'inattentive. A 11-14 mois, cette vocalisation est la première à montrer

de l'ajustement *intermodal*, et donc à servir des fonctions *illocutoires* (i.e. de l'ordre de la communication intentionnelle). Puis à 15-20 mois, par-dessus ce babillage intentionnel, les autres vocalisations s'ajusteront juste avant de servir des fonctions *locutoires* (i.e. verbales).

Au regard de l'hypothèse de Locke (2006), nos résultats suggèrent donc que le babillage en particulier, étant vecteur de fonctions illocutoires (i.e. d'ajustement *intermodal* ici) en théorie dépendant de la responsivité parentale, ait plausiblement conduit à l'émergence du langage. Nos résultats suggèrent même que cette vocalisation conduirait à la maîtrise vocale plus générale (i.e. ajustement du babillage avant les autres vocalisations) qui est de fait nécessaire à l'acquisition du langage. Toutefois, des études complémentaires sont nécessaires à l'étayage de cette hypothèse.

6.5. LIMITES ET PERSPECTIVES

6.5.1. Dans une perspective développementale : des histoires à compléter

Les histoires développementales des ajustements (partie 6.2) construites à partir de nos résultats (parties 3 et 4) montrent un certain nombre de discontinuités qui appellent à être discutées et appréhendées dans de futures investigations.

6.5.1.1. *Dans l'étude des ajustements chez le jeune chimpanzé : vocalisations et profil du destinataire*

Il semble important de souligner deux principales limites de notre étude menée sur les jeunes chimpanzés lors de nos investigations. On discutera donc ici du fait que (i) les vocalisations n'aient pas été considérées dans nos analyses et (ii) les résultats des échanges avec la mère n'aient pas été comparés à ceux avec des partenaires non-maternels.

En effet, les vocalisations isolées ou combinées n'ont pas été prises en compte dans cette étude alors que celles-ci émergent dès la naissance et que leur utilisation suivrait une histoire développementale. Les chimpanzés produisent plus de gestes et vocalisations de façon isolée que de combinaisons des deux (Hobaiter et al., 2017). A partir du stade juvénile, les individus produiraient de moins en moins de gestes et de plus en plus de vocalisations jusqu'à l'âge adulte où ils augmenteraient leur utilisation de gestes (Hobaiter et al., 2017). Par ailleurs, des études rapportent qu'entre l'enfance et le stade juvénile, le chimpanzé utiliserait ses *grunt* de façon de plus en plus flexible (Dezecache et al., 2020) et ses cris d'alarme de façon intentionnelle selon les auteurs (i.e. associés à des alternances de regard; Dezecache et al., 2019). Ces études et nos résultats suggèrent ainsi que si les vocalisations avaient été considérées, l'ajustement *intermodal* avec des signaux audibles vocaux aurait pu émerger en même temps que celui du geste tactile au stade juvénile, voire avant. Cela n'aurait pas remis en question l'ordre d'émergence des ajustements mais plutôt la nature du signal par lequel l'ajustement émergerait et donc l'interprétation qu'on peut en faire.

De plus, les résultats des ajustements envers la mère n'ont pas été comparés à des ajustements envers les autres partenaires. Pourtant, certaines études soulignent un effet du lien avec le destinataire (maternel *vs* non-maternel) sur la signalisation du chimpanzé immature. En effet, les chimpanzés enfants et juvéniles déploieraient une communication intentionnelle plus marquée envers les partenaires non-maternels qu'envers la mère (Fröhlich et al., 2018) et plus de signaux silencieux-visuels vers les partenaires non-maternels que vers la mère lors des invitations au jeu (Fröhlich et al., 2016). Dans notre échantillon, les signaux sont essentiellement produits lors des contextes de jeu (34%) et de nourrissage (34% ;Dafreville et al., 2021). Ces résultats suggèrent donc que l'ajustement des signaux silencieux-visuels (i.e. l'ajustement *unimodal*) pourrait émerger avant l'adolescence avec des partenaires non-maternels. C'est pourquoi des analyses complémentaires seront nécessaires à la complétion de cette histoire développementale.

6.5.1.2. ***Dans l'étude des ajustements chez l'enfant*** *préverbal : responsivité maternelle et élaboration de la signalisation*

Les résultats de nos investigations nous ont permis de construire *une* histoire développementale des ajustements tenant compte des gestes et des vocalisations. Toutefois, l'ensemble des points discutés tout au long de cette histoire soulèvent plusieurs limites à cette investigation à savoir (i) le codage de la responsivité et (ii) l'analyse de l'articulation des différents signaux entre eux (i.e. séquences et combinaisons).

En effet, la responsivité n'a pas été finement relevée dans cette étude. Certains éléments concernant la façon dont l'effet perlocutoire (i.e. l'effet produit sur le destinataire) soutient l'émergence des fonctions illocutoires (i.e. l'ajustement des signaux à l'attention visuelle maternelle ici) manquent à nos interprétations. Comme il a été discuté précédemment (voir partie 6.2.2), certains signaux par lesquels émergent les ajustements seraient sujets à une forte responsivité de l'adulte à savoir les gestes déictiques (Wu & Gros-Louis, 2015) et le babillage (Bloom & Beaumont, 1993; Bloom & Lo, 1990; Gros-Louis et al., 2014; Locke, 2006). Néanmoins, d'autres signaux dont la responsivité n'a, à notre connaissance, pas encore été testée s'ajustent également (e.g gestes audible *avec* objet, bruits de bouche). Il aurait donc été intéressant d'inclure cette variable de responsivité à nos analyses. Par ailleurs, soulignons que la responsivité maternelle n'est pas exclusivement vocale et donc pourrait être aussi visuelle (i.e. maintien ou regain d'attention visuelle maternelle suite à une situation d'inattention visuelle). De ce fait, elle nous aurait permis d'appréhender un autre principal marqueur de communication intentionnelle à savoir la persistance et l'élaboration (voir partie 1.2.2.2.1) ; ce qui nous conduit au point de discussion suivant.

Il aurait été intéressant dans le cadre de cette investigation d'approfondir nos analyses quant à l'articulation combinatoire (i.e. de cooccurrence) et séquentielle des signaux entre eux au cours de ces ajustements. Dans nos analyses, nous nous sommes essentiellement intéressés à l'effet de l'attention visuelle sur les

combinaisons multimodales (i.e. de signaux de différentes modalités sensorielles de perception) sans renseigner la nature des signaux impliqués. Pourtant, comme il a été mentionné précédemment (partie 1.4.2.1.2), l'intégration des différents signaux dans des combinaisons multimodales suivrait elle aussi une histoire développementale (Aureli et al., 2017; Murillo et al., 2018). De plus, nos résultats soulignent l'importance du geste audible dans les ajustements et suggèrent sa possible implication dans le réagencement modal de la signalisation vers le langage essentiellement vocal (discuté partie 6.2.2.1; Murillo et al., 2019). Il semble donc pertinent de poursuivre nos investigations en ne nous intéressant pas seulement aux combinaisons multimodales mais aussi à celles impliquant les gestes audibles et vocalisations (Murillo et al., 2019). De plus, dans cette étude, la séquentialité des signaux intra-individuels et inter-individuels n'a pas été analysée. Cela nous permettrait d'explorer les potentielles stratégies de regain d'attention élaborées par l'enfant en situation d'inattention visuelle maternelle voire de non-responsivité maternelle (comme discuté plus haut). Ce dernier point souligne la complexité de l'ajustement *intermodal* et la nécessité d'affiner ce descripteur en vue de tester son effet de prédiction langagière (notamment syntaxique, voir partie 5 et 6.3.2).

6.5.2. Dans une perspective comparative : un manque de représentativité à considérer et une rythmicité à explorer

Une première comparaison des différentes histoires développementales montre d'ores et déjà des différences d'ajustements inter-espèces notables mais dont l'interprétation est à nuancer. Cependant, cette comparaison suggère également un terrain d'investigation potentiel sur lequel la comparabilité des espèces pourrait être facilitée et pourrait s'inscrire dans l'approche *evo-devo*.

6.5.2.1. ***Représentativité de l'échantillon humain*** *: des variations culturelles à considérer*

Les théories historico-culturelles de l'acquisition du langage ont mis en avant l'importance de la culture et de l'interaction sociale comme supports médiateurs de l'apprentissage du langage notamment par l'étayage parental (Chomsky et al., 1979; Vygotsky, 1978). Les comportements parentaux pourraient en effet être issus de modèles socio-culturels propres à certaines populations (Greenfield, 2009; Keller et al., 2009). Sur la base des travaux de Kagitcibasi (1996), Keller et al. (2009) définissent trois *contextes écologiques d'apprentissage* typiques : urbain occidental, rural non occidental, urbain non occidental. Keller et al. (2009) suggèrent que les parents des deux contextes les plus opposés (i.e. urbain occidental et rural non occidental) adoptent des stratégies parentales distinctes. En effet, la parentalité *distale* (i.e. composée d'interactions en *face-to-face* et de stimulations de l'enfant via des objets) serait observable en contexte urbain occidental. A l'opposé, la parentalité *proximale* (i.e. basée sur le contact corporel et les stimulations corporelles) serait quant à elle observable en contexte rural non occidental. Enfin, une parentalité « intermédiaire », composée à la fois de comportements proximaux et distaux serait observable en contexte urbain non occidental. Ainsi, Keller et al. (2009) mettent en avant l'importance du milieu écologique et de la culture dans la variation des comportements d'étayage maternel, tout particulièrement concernant le temps passé en contact physique et la communication distale au sein de la dyade mère-enfant.

Notre échantillon s'identifie clairement au contexte urbain occidental (WEIRD; Henrich, et al., 2010) et nos résultats seraient cohérents avec le modèle parental *distal*, du fait du peu de gestes tactiles produits par l'enfant (évoqué partie 6.4.1.1) et de l'importance de l'objet dans l'interaction (voir parties 6.2.2.1 et 6.4.1.2). En effet, admettons que le parent soumette son enfant à plus de distanciation physique (e.g. utilisation de la poussette plutôt que d'une écharpe de portage) et intègre davantage d'objets dans son environnement (e.g. mise en place d'un parc à jeux). Cet étayage parental susciterait alors moins de communication

tactile de la part de l'enfant et plus de manipulation de l'objet dans la dynamique d'interaction. Toutefois, dans le cadre de cette investigation, nous n'avons pas encore contrôlé le profil parental des mères de l'échantillon et testé véritablement l'effet du type d'étayage parental sur la signalisation de l'enfant. Dans le cadre du projet de recherche dans lequel s'intègre ce travail de thèse (*MODAL*, mené par Marie Bourjade), il s'agira prochainement d'explorer l'effet de l'étayage parental (voire alloparental) sur les capacités d'ajustements de l'enfant sur diverses populations humaines (Bader & Fouts, 2018; Mastin & Vogt, 2016).

Par ailleurs, rappelons que l'hypothèse de Falk (2004) repose sur un constat comparatif entre primates humains (i.e. l'enfant n'a pas la maturité motrice suffisante pour s'accrocher à sa mère pendant les déplacements) et non-humains (i.e. l'enfant va pouvoir le faire assez rapidement après la naissance). De ce fait, dans des études futures, il s'agirait également de contrôler cette hypothèse en répliquant la question sur d'autres populations de primates non-humains. Par exemple, l'idée serait de vérifier que la mère primate non-humaine n'utilise pas de stratégie d'attention compensatoire vocale (ou du moins pas assimilable à celle de la mère humaine) quand elle s'éloigne de son enfant et que l'enfant non-humain ne montre pas de préférence d'utilisation orale lors de l'ajustement *intermodal*.

6.5.2.2. *Rythmicité des échanges : du geste audible à la vocalisation ?*

Sur la base de nos investigations concernant le développement de la *communication sensible à l'attention du destinataire*, il semble que le geste avec objet jouerait un rôle dans le développement des capacités d'ajustement *intermodal* (discuté partie 6.4.1.2). En effet, quelle que soit l'espèce, le premier signal audible à être ajusté (voir les Figures 1 et 2) est produit avec objet.

L'étude du geste audible pourrait jouer un rôle dans le réagencement modal de la signalisation de l'enfant humain (discuté partie 6.2.2.1). Jusqu'alors, de nombreuses études ont montré l'importance de la combinaison entre gestes déictiques et vocalisation dans l'acquisition du langage (Aureli et al., 2017; Fasolo

& D'Odorico, 2012; Iverson & Fagan, 2004; Iverson & Goldin-meadow, 2005; McCune et al., 2020; Murillo & Casla, 2020; Murillo et al., 2018). Toutefois, de récents travaux suggèrent que des gestes audibles avec objet (*mouvements rythmique avec objet* ; Murillo et al., 2019) pourraient eux aussi servir le développement communicationnel de l'enfant préverbal en cours d'acquisition du langage (Murillo et al., 2019). De plus, nos résultats montrent que c'est par les gestes audibles *avec* objet (et non *sans*), qu'émerge la capacité d'ajustement *intermodal*, i.e. l'utilisation intentionnelle du signal audible. L'ensemble de ces résultats suggèrent que le geste audible avec objet, de par son canal de production gestuel, sa composante audible (comme celle du vocal) et son articulation avec les vocalisations, pourrait participer au réagencement modal de la signalisation de l'enfant préverbal. Cependant, d'autres investigations resteraient à faire pour étayer cette hypothèse.

Par ailleurs, le geste audible est comparable d'une espèce à l'autre (similitude de répertoire ; voir Table 2) et pourrait avoir servi des fonctions adaptatives chez le primate non-humain, comme souligné partie 6.4.1.2. De plus, des études sur chimpanzés ont montré que le geste audible de *drumming* (même si sans objet, selon notre définition) est principalement accompagné de vocalisations (Babiszewska et al., 2015; Crockford & Boesch, 2005; Eleuteri et al., 2022). Ce geste de *drumming* montrant une signature rythmique (du geste et de son articulation avec les vocalisations), participerait à la « mouvance » des individus, sous-jacente à la dynamique de fission-fusion (Aureli et al., 2008) de la communauté de chimpanzé (Babiszewska et al., 2015; Eleuteri et al., 2022). Ce geste aurait donc véritablement servi une fonction sociale voire adaptative à l'échelle du groupe. Il serait intéressant, dans de prochaines investigations, de s'intéresser à la rythmicité d'autres gestes notamment avec objet et de leur articulation avec d'autres signaux (telles que les vocalisations) en évaluant leurs fonctions sociales et potentiellement adaptatives.

Dans une perspective comparative, l'étude du geste audible pourrait ainsi s'inscrire dans l'approche *evo-devo*. En effet, comme il a été développé ci-dessus, l'étude du geste audible pourrait aider à construire un modèle de réagencement

modal de la signalisation de l'enfant préverbal. De plus, ce geste audible employé par l'enfant est largement comparable à celui employé par le chimpanzé. D'autant plus que quelle que soit l'espèce étudiée ici, le geste audible s'articule fréquemment avec des vocalisations selon une certaine « cohérence rythmique » (i.e. une coordination séquentielle non-aléatoire des différents signaux dans le temps; *Pan troglodytes*: Eleuteri et al., 2022; *Homo sapiens*: Murillo et al., 2019). On pourrait donc, à l'échelle du primate immature et/ou de la dyade mère-enfant, explorer l'articulation du geste (pas uniquement silencieux-visuel) avec les vocalisations (comme il a déjà pu être investigué sur des enfants plus jeunes dans des contextes de *face-to-face* ; e.g. Devouche & Gratier, 2001) au regard de la transition modale de la signalisation de l'enfant. Ainsi par l'approche *evo-devo*, il serait intéressant de savoir dans quelle mesure cette polyrythmie entre différents signaux pourrait nous renseigner quant au réagencement modal par lequel le langage aurait pu émerger. Ceci constituerait ainsi une extension de théories évolutionnaires (Gibson, 1991) et développementales (Liszkowski & Rüther, 2021; Rodríguez et al., 2015) selon lesquelles l'émergence et l'acquisition du langage serait dépendante de la manipulation d'objet (ce qui devrait toutefois être discuté au regard des variations interculturelles ; Keller et al., 2009).

6.6. CONCLUSION

Pour conclure, les présentes études observationnelles conduites sur les chimpanzés immatures et les enfants préverbaux français mettent en évidence le développement de la *communication sensible à l'attention du destinataire*.

A partir des descripteurs d'ajustement *unimodal* et *intermodal,* nous avons montré que les jeunes primates chimpanzés et humains observés en condition écologique, ajustaient leurs signaux à l'attention visuelle de leur mère. En effet, ces capacités d'ajustement émergeaient selon une certaine chronologie et l'étude

approfondie des signaux impliqués dans ce *processus* développemental nous aura permis de proposer des « esquisses » d'histoires développementales. De nombreux questionnements persistent toutefois quant à l'articulation des différents signaux dans la transition modale du geste au vocal pour l'enfant humain. En testant l'effet de prédiction de ces ajustements, nos résultats ont montré que la capacité d'ajustement *unimodal* constituait un prédicteur *lexical* du langage chez l'enfant préverbal mais ce n'était pas le cas concernant la capacité d'ajustement *intermodal*. Au-delà de ces résultats, les outils ayant été construits pour mener à bien ce travail pourraient constituer de bons outils de modélisation susceptibles de servir l'approche développementale dans de prochaines investigations sur diverses populations. Malgré une comparabilité limitée et un manque de représentativité de nos échantillons, une première confrontation des résultats obtenus concernant le développement des capacités d'ajustement permet de souligner la nécessité d'alimenter des modèles développementaux sur la communication gestuelle et vocale du jeune primate humain et non-humains et donc la poursuite de nos investigations. A travers l'approche *evo-devo*, ce travail de thèse suggère de nouveaux terrains d'investigation (e.g. l'étude comparative du geste audible) qui sembleraient pertinents pour l'étude du réagencement modal menant à l'émergence et l'acquisition du langage.

Les présents résultats pourront être complétés par des analyses ultérieures sur davantage de populations humaines et non-humaines dans le but d'en construire une « fresque multimodale » à la fois développementale et évolutionnaire.

7.
BIBLIOGRAPHIE

Altmann, J. (1974). Observational study of behaviour: sampling methods. *Behaviour*, *49*, 227–265.

Arbib, M. A. (2012). *How the brain got language : the mirror system hypothesis*. Oxford: Oxford University Press.

Arbib, M. A., Liebal, K., & Pika, S. (2008). Primate vocalization, gesture, and the evolution of human language. *Current Anthropology*, *49*(6), 1053–1076. https://doi.org/10.1086/593015

Armstrong, D., & Wilcox, S. (2007). *The Gestural Origin of Language* (Oxford Uni). Retrieved from https://doi.org/10.1093/acprof:oso/9780195163483.001.0001

Aureli, F., Schaffner, C. M., Boesch, C., Bearder, S. K., Call, J., Chapman, C. A., … Van Schaik, C. P. (2008). Fission-fusion dynamics new research frameworks. *Current Anthropology*, *49*(4), 627–654. https://doi.org/10.1086/586708

Aureli, T., Spinelli, M., Perucchini, P., & Odorico, L. D. (2017). *The Pointing – Vocal Coupling Progression in the First Half of the Second Year of Life*. *22*(6), 801–818. https://doi.org/10.1111/infa.12181

Austin, J. L. (1962). Speech acts. In *Choice Reviews Online* (Vol. 37). https://doi.org/10.5860/choice.37-0017

Babiszewska, M., Schel, A. M., Wilke, C., & Slocombe, K. E. (2015). *Social , Contextual , and Individual Factors Affecting the Occurrence and Acoustic Structure of Drumming Bouts in Wild Chimpanzees (Pan troglodytes)*. *134*(March 2014), 125–134. https://doi.org/10.1002/ajpa.22634

Bader, L., & Fouts, H. (2018). Cultural Models of Infant Emotions and Needs among the Gamo People of Southern Ethiopia. *Infant Mental Health Journal*, *39*(5), 497–510. https://doi.org/10.1002/imhj.21734.Cultural

Bard, K. (1994). Evolutionary roots of intuitive parenting: Maternal competence in chimpanzees. *Early Development and Parenting*, *3*(1), 19–28. https://doi.org/10.1002/edp.2430030104

Bard, K. (2019). Parenting in Nonhuman Primates. In *Handbook of Parenting*. https://doi.org/10.4324/9780429401459-3

Bard, K., Maguire-Herring, V., Tomonaga, M., & Matsuzawa, T. (2019). The gesture 'Touch': Does meaning-making develop in chimpanzees' use of a very flexible gesture? *Animal Cognition*, *22*(4), 535–550. https://doi.org/10.1007/s10071-017-1136-0

Bassano, D., Labrell, F., & Bonnet, P. (2020). *Évaluer Les Débuts Du Langage Avec Le DLPF (révision)*.

Bassano, D., Labrell, F., Champaud, C., Lemétayer, F., & Bonnet, P. (2005). Le DLPF : un nouvel outil pour l'évaluation du développement du langage de production en français. In *Enfance* (Vol. 57). https://doi.org/10.3917/enf.572.0171

Bates, E. (1971). *The development of conversational skill in 2, 3 and 4 year olds*.

Bates, E., Camaioni, L., & Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill-Palmer Quarterly of Behavior and Development*, *21*(3), 205–

226.

Beebe, B., Messinger, D., Bahrick, L. E., Margolis, A., Buck, K. A., & Chen, H. (2016). A Systems View of Mother–Infant Face-to-Face Communication. *Developmental Psychology*, *52*(4), 556–571. https://doi.org/10.1037/a0040085.supp

Bernicot, J., Comeau, J., Feider, H., European, S., Bernicot, J., Comeau, J., & Feider, H. (1993). Speech Acts by Mother and Child : Determining their Nature and Form. *European Journal of Psychology of Education*, *8*(1), 35–50.

Bernicot, J., & Marcos, H. (1993). La demande prélinguistique et linguistique: qu'apporte le langage? *Journal International de Psychologie*, *28*(3), 291–306.

Beuker, K. T., Rommelse, N. N. J., Donders, R., & Buitelaar, J. K. (2013). Development of early communication skills in the first two years of life. *Infant Behavior and Development*, *36*(1), 71–83. https://doi.org/10.1016/j.infbeh.2012.11.001

Bickerton, D. (1990). Language & species. In *Language & species*. Chicago: The University of Chicago Press.

Bjorklund, D. F., & Pellegrini, A. D. (2000). Child development and evolutionary psychology. *Child Development*, *71*(6), 1687–1708. https://doi.org/10.1111/1467-8624.00258

Bloom, K., & Beaumont, S. (1993). Adults' attributions of intentionality to vocalizing infants. *The Grants Register 2022*, *13*, 235–247. https://doi.org/10.1057/978-1-349-96042-2_124111

Bloom, K., & Lo, E. (1990). Adult perceptions of vocalizing infants. *Infant Behavior and Development*, *13*(2), 209–219. https://doi.org/10.1016/0163-6383(90)90031-3

Boesch, C., Bombjaková, D., Meier, A., & Mundry, R. (2019). Learning curves and teaching when acquiring nut-cracking in humans and chimpanzees. *Scientific Reports*, *9*(1), 1–14. https://doi.org/10.1038/s41598-018-38392-8

Bogin, B. (1997). Evolutionary Hypotheses for Human Childhood. *Yearbook of Physical Anthropology*, *40*, 63–89. https://doi.org/10.1002/(sici)1096-8644(1997)25+<63::aid-ajpa3>3.0.co;2-8

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology and Evolution*, *24*(3), 127–135. https://doi.org/10.1016/j.tree.2008.10.008

Botting, J., & Bastian, M. (2019). Orangutans (Pongo pygmaeus and Hybrid) and Gorillas (Gorilla gorilla gorilla) Modify Their Visual, but Not Auditory, Communicative Behaviors, Depending on the Attentional State of a Human Experimenter. *International Journal of Primatology*, *40*(2), 244–262. https://doi.org/10.1007/s10764-019-00083-0

Bourjade, M. (2017). Social Attention. *The International Encyclopedia of Primatology*, 1–2. https://doi.org/10.1002/9781119179313.wbprim0356

Bourjade, M., Cochet, H., Molesti, S., & Guidetti, M. (2020). Is Conceptual Diversity an Advantage for Scientific Inquiry ? A Case Study on the Concept of ' Gesture ' in Comparative Psychology. *Integrative Psychological and Behavioral*

Science, *54*, 805–832.

Bourjade, M., Dafreville, M., Scola, C., & Jover, M. (n.d.). Six-month-old infant's communication in a comparative perspective: how do maternal attention and interaction matter? *(Submitted)*.

Bourjade, M., Meguerditchian, A., Maille, A., Gaunet, F., & Vauclair, J. (2014). Olive baboons, Papio anubis, adjust their visual and auditory intentional gestures to the visual attention of others. *Animal Behaviour*, *87*, 121–128. https://doi.org/10.1016/j.anbehav.2013.10.019

Bronfenbrenner, U. (1992). *Ecological systems theory*. Jessica Kingsley Publisher.

Bronfenbrenner, U. (2005). *Ecological systems theory (1992)*.

Bründl, A. C., Tkaczynski, P. J., Nohon Kohou, G., Boesch, C., Wittig, R. M., & Crockford, C. (2021). Systematic mapping of developmental milestones in wild chimpanzees. *Developmental Science*, *24*(1), 1–13. https://doi.org/10.1111/desc.12988

Bruner, J. (1975). The ontogenesis of speech acts. *Journal of Child Language*, *2*(1), 1–19. https://doi.org/10.1017/S0305000900000866

Bruner, J. (1981). *Intention in the structure of action and interaction* (Lipsett L). Norwood: Ablex, Advances in infancy research.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods and Research*, *33*(2), 261–304. https://doi.org/10.1177/0049124104268644

Byrne, R. W., Cartmill, E., Genty, E., Graham, K. E., Hobaiter, C., & Tanner, J. (2017). Great ape gestures: intentional communication with a rich set of innate signals. *Animal Cognition*, *20*(4), 755–769. https://doi.org/10.1007/s10071-017-1096-4

Camaioni, L. (1997). The Emergence of Intentional Communication in Ontogeny, Phylogeny, and Pathology. *European Psychologist*, *2*(3), 216–225. https://doi.org/10.1027/1016-9040.2.3.216

Camaioni, L., Perucchini, P., Bellagamba, F., & Colonnesi, C. (2004). The Role of Declarative Pointing in Developing a Theory of Mind. *Infancy*, *5*(3), 291–308. Retrieved from https://doi.org/10.1207/s15327078in0503_3

Cameron-Faulkner, T., Theakston, A., Lieven, E., & Tomasello, M. (2015). The relationship between infant holdout and gives, and pointing. *Infancy*, *20*(5), 576–586. https://doi.org/10.1111/infa.12085

Canteloup, C., & Bovet, D. (2014). *Do Tonkean macaques (Macaca tonkeana) tailor their gestural and visual signals to fit the attentional states of a human partner ?* (Hewes 1973). https://doi.org/10.1007/s10071-014-0814-4

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social Cognition , Joint Attention , and Communicative Competence from 9 to 15 Months of Age. In *Monographs of the Society for Research in Child Development* (Vol. 63, p. (Serial No.255)).

Caselli, M. C., Rinaldi, P., & Volterra, V. (2019). *Early Action and Gesture " Vocabulary " and Its Relation With Word Comprehension and Production Author (s): Maria Cristina Caselli , Pasquale Rinaldi , Silvia Stefanini and Virginia Volterra Published by : Wiley on behalf of the Society for Research in . 83*(2), 526–542.

Choi, B., Wei, R., & Rowe, M. L. (2021). Show, Give, and Point Gestures Across Infancy Differentially Predict Language Development. *Developmental Psychology*, *57*(6), 851–862. https://doi.org/10.1037/dev0001195

Chomsky, N., Piaget, J., & Atran, S. (1979). *Théories du langage, théories de l'apprentissage: le débat entre Jean Piaget et Noam Chomsky*.

Clark, C. . (1977). *A preliminary report on weaning among chimpanzees of the Gombe National Park, Tanzania* (S. Chevalier-Skolnikoff & F. E. Poirier, eds.).

Cochet, H., & Byrne, R. W. (2016a). Communication in the second and third year of life: Relationships between nonverbal social skills and language. *Infant Behavior and Development*, *44*, 189–198. https://doi.org/10.1016/j.infbeh.2016.07.003

Cochet, H., & Byrne, R. W. (2016b). Infant Behavior and Development Full length article Communication in the second and third year of life : Relationships between nonverbal social skills and language. *Infant Behavior and Development*, *44*, 189–198. https://doi.org/10.1016/j.infbeh.2016.07.003

Cochet, H., & Vauclair, J. (2010). Pointing gestures produced by toddlers from 15 to 30 months: Different functions, hand shapes and laterality patterns. *Infant Behavior and Development*, *33*(4), 431–441. https://doi.org/10.1016/j.infbeh.2010.04.009

Cohen, L., & Billard, A. (2018). Social babbling: The emergence of symbolic gestures and words. *Neural Networks*, *106*, 194–204. https://doi.org/10.1016/j.neunet.2018.06.016

Colonnesi, C., Bonne, J. H. Z., Van der Zande, A., & Bögels, S. M. (2012). Coordination of gaze, facial expressions and vocalizations of early infant communication with mother and father. *Infant Behavior and Development*, *35*, 523–532.

Colonnesi, C., Stams, G. J. J. M., Koster, I., & Noom, M. J. (2010). The relation between pointing and language development: A meta-analysis. *Developmental Review*, *30*(4), 352–366. https://doi.org/10.1016/j.dr.2010.10.001

Corballis, M. . (2002). *From Hand to Mouth: The Origins of Language* (P. U. Press, ed.). Retrieved from http://hdl.handle.net/10822/951670

Cristià, A. (2011). Fine-grained variation in caregivers' /s/ predicts their infants' /s/ category. *The Journal of the Acoustical Society of America*, *129*(5), 3271–3280. https://doi.org/10.1121/1.3562562

Crockford, C., & Boesch, C. (2005). Call combinations in wild chimpanzees. *Behaviour*, *142*(4), 397–421. https://doi.org/10.1163/1568539054012047

Csibra, G. (2010). *Recognizing Communicative Intentions in Infancy Gergely Csibra*. *25*(2), 141–168.

Csibra, G., & Volein, Á. (2008). Infants can infer the presence of hidden objects from referential gaze information. *British Journal of Developmental Psychology*, *26*(1), 1–11. https://doi.org/10.1348/026151007X185987

D'Odorico, L., Cassibba, R., & Salerni, N. (1997). Temporal Relationships between Gaze and Vocal Behavior in Prelinguistic and Linguistic Communication. *Journal of Psycholinguistic Research*, *26*(5), 539–556. https://doi.org/10.1023/A:1025027830861

D'odorico, L., & Franco, F. (1991). *Selective production of vocalization types in different communication contexts*. *18*(1991), 475–499. https://doi.org/10.1017/S0305000900011211

Dafreville, M., Guidetti, M., & Bourjade, M. (n.d.-a). Attention-sensitive signaling by 7-to-20-month-old infants in a comparative perspective. *(Submitted)*.

Dafreville, M., Guidetti, M., & Bourjade, M. (n.d.-b). Beyond pointing gestures: Attention-sensitive communication in 7-to-20-month-old infants develops pragmatic skills. *Child Development*.

Dafreville, M., Hobaiter, C., Guidetti, M., Sillam-Dussès, D., & Bourjade, M. (2021). Sensitivity to the communicative partner's attentional state: A developmental study on mother–infant dyads in wild chimpanzees (Pan troglodytes schweinfurthii). *American Journal of Primatology*, *83*(12). https://doi.org/10.1002/ajp.23339

Darwin, C. (1958). *The origin of species: By means of natural selection of the preservation of favoured races in the struggle for life* (New York:).

De Onis, M. (2006). WHO Motor Development Study: Windows of achievement for six gross motor development milestones. *Acta Paediatrica, International Journal of Paediatrics*, *95*(SUPPL. 450), 86–95. https://doi.org/10.1080/08035320500495563

Delgado, C. F., Messinger, D. S., & Yale, M. E. (2002). *Infant responses to direction of parental gaze: a comparison of two still-face conditions*. *25*, 311–318.

Dennett, D. C. (1983). Intentional systems in cognitive ethology: The "Panglossian paradigm" defended. *Behavioral and Brain Sciences*, *6*(3), 343–355. https://doi.org/10.1017/S0140525X00016393

Deshpande, A., Gupta, S., & Sinha, A. (2018). Intentional communication between wild bonnet macaques and humans. *Scientific Reports*, *8*(1), 1–12. https://doi.org/10.1038/s41598-018-22928-z

Devouche, E., & Gratier, M. (2001). Microanalyse du rythme dans les échanges vocaux et gestuels entre la mère et son bébé de 10 semaines. *Devenir*, *13*(2), 55–82. https://doi.org/10.3917/dev.012.0055

Dezecache, G., Crockford, C., & Zuberbühler, K. (2019). The development of communication in alarm contexts in wild chimpanzees. *Behavioral Ecology and Sociobiology*, *73*(8). https://doi.org/10.1007/s00265-019-2716-6

Dezecache, G., Dahl, C. D., Zuberbühler, K., & Davila-Ross, M. (2020). A machine learning approach to infant distress calls and maternal behaviour of wild chimpanzees. *Animal Cognition*. https://doi.org/https://doi.org/10.1007/s10071-020-01437-5

Dezecache, G., Zuberbühler, K., Davila-Ross, M., & Dahl, C. D. (2020). *Flexibility in wild infant chimpanzee vocal behavior*. 9639. https://doi.org/https://doi.org/10.1093/jole/lzaa009

Donnellan, E., Bannard, C., Mcgillion, M. L., Slocombe, K. E., & Matthews, D. (2019). *Infants ' intentionally communicative vocalizations elicit responses from caregivers and are the best predictors of the transition to language : A longitudinal investigation of infants ' vocalizations , gestures and word production*. (March), 1–21. https://doi.org/10.1111/desc.12843

Donnelly, S., & Kidd, E. (2021). The Longitudinal Relationship Between Conversational Turn-Taking and Vocabulary Growth in Early Language Development. *Child Development*, *92*(2), 609–625. https://doi.org/10.1111/cdev.13511

Dufour, V., Poulin, N., Curé, C., & Sterck, E. H. M. (2015). Chimpanzee drumming : a spontaneous performance with characteristics of human musical drumming. *Nature Publishing Group*, (November 2014), 8–13. https://doi.org/10.1038/srep11320

Dunbar, R. (2004). *Gossip in Evolutionary Perspective*. https://doi.org/10.1037/1089-2680.8.2.100

Dunbar, R., & Turkel, W. (2006). Grooming, Gossip, and the Evolution of Language. *Language*. https://doi.org/10.2307/417520

Eggeling, W. . (1947). Observations on the Ecology of the Budongo Rain Forest, Uganda. *Journal of Ecology*, *34*(1), 20–87. https://doi.org/10.1017/CBO9781107415324.004

Eleuteri, V., Henderson, M., Soldati, A., Badihi, G., Zuberbühler, K., & Hobaiter, C. (2022). The form and function of chimpanzee buttress drumming. *Animal Behaviour*, *192*, 189–205. https://doi.org/10.1016/j.anbehav.2022.07.013

Emery, N. J. (2015). *The Eyes Have It : The Neuroethology , Function and Evolution of Social Gaze*. *7634*(September). https://doi.org/10.1016/S0149-7634(00)00025-7

Emmorey, K., Mehta, S., & Grabowski, T. J. (2007). The neural correlates of sign versus word production. *NeuroImage*, *36*(1), 202–208. https://doi.org/10.1016/j.neuroimage.2007.02.040

Esteve-Gibert, N., & Guellaï, B. (2018a). Prosody in the auditory and visual domains: A developmental perspective. *Frontiers in Psychology*, *9*(March), 1–10. https://doi.org/10.3389/fpsyg.2018.00338

Esteve-Gibert, N., & Guellaï, B. (2018b). Prosody in the auditory and visual domains: A developmental perspective. *Frontiers in Psychology*, *9*(March), 1–10. https://doi.org/10.3389/fpsyg.2018.00338

Falk, D. (2004). Prelinguistic evolution in early hominins: Whence motherese? *Behavioral and Brain Sciences*, *27*(04). https://doi.org/10.1017/S0140525X04000111

Fasolo, M., & D'Odorico, L. (2012). Gesture-plus-word combinations, transitional forms, and language development. *Gesture*, *12*(1), 1–15. https://doi.org/10.1075/gest.12.1.01fas

Fenson, L. (2007). *MacArthur-Bates communicative development inventories*. Baltimore: MD: Paul H. Brookes Publishing Company.

Fitch, T. (2010). *The Evolution of Language*. Cambridge University Press.

Fitch, W. T. (2000). *The evolution of speech : a comparative review*. *6613*(00), 258–267.

Franco, F., & Butterworth, G. (1996). Pointing and social awareness: declaring and requesting in the second year. *Journal of Child Language*, *23*(02), 307–336. https://doi.org/10.1017/s0305000900008813

Frankenhuis, W. E., & Tiokhin, L. (2018). Evolutionary Developmental Psychology: 2017 Redux. *Child Development*, *89*(6), 2282–2287.

https://doi.org/10.1111/cdev.13018

Franklin, B., Warlaumont, A. S., Messinger, D., Bene, E., Nathani Iyer, S., Lee, C. C., … Oller, D. K. (2014). Effects of Parental Interaction on Infant Vocalization Rate, Variability and Vocal Type. *Language Learning and Development*, *10*(3), 279–296. https://doi.org/10.1080/15475441.2013.849176

Fröhlich, M., & Hobaiter, C. (2018). The development of gestural communication in great apes. *Behavioral Ecology and Sociobiology*, *72*(12). https://doi.org/10.1007/s00265-018-2619-y

Fröhlich, M., Sievers, C., Townsend, S. W., Gruber, T., & van Schaik, C. P. (2019). Multimodal communication and language origins: integrating gestures and vocalizations. *Biological Reviews*, *1820*, 1809–1829. https://doi.org/10.1111/brv.12535

Fröhlich, M., & van Schaik, C. P. (2020). Must all signals be evolved? A proposal for a new classification of communicative acts. *Wiley Interdisciplinary Reviews: Cognitive Science*, (February), 1–13. https://doi.org/10.1002/wcs.1527

Fröhlich, M., Wittig, R. M., & Pika, S. (2016). Play-solicitation gestures in chimpanzees in the wild: Flexible adjustment to social circumstances and individual matrices. *Royal Society Open Science*, *3*(8). https://doi.org/10.1098/rsos.160278

Fröhlich, M., Wittig, R. M., & Pika, S. (2018). The ontogeny of intentional communication in chimpanzees in the wild. *Developmental Science*, *22*(1), 1–14. https://doi.org/10.1111/desc.12716

Gaunet, F., & Deputte, B. L. (2011). Functionally referential and intentional communication in the domestic dog: Effects of spatial and social contexts. *Animal Cognition*, *14*(6), 849–860. https://doi.org/10.1007/s10071-011-0418-1

Geary, D. C., & Bjorklund, D. F. (2000). Evolutionary Developmental Psychology. *Child Development*, *71*(1), 420–429. https://doi.org/10.1016/B978-0-08-097086-8.81018-1

Genty, E., Breuer, T., Hobaiter, C., & Byrne, R. W. (2009). Gestural communication of the gorilla (Gorilla gorilla): repertoire , intentionality and possible origins. *Animal Cognition*, *12*, 527–546. https://doi.org/10.1007/s10071-009-0213-4

Genty, E., & Byrne, R. W. (2010). Why do gorillas make sequences of gestures? *Animal Cognition*, *13*(2), 287–301. https://doi.org/10.1007/s10071-009-0266-4

Gibson, K. (1991). Tools , Language and Intelligence : Evolutionary Implications. *Royal Anthropological Institute of Great Britain and Ireland*, *26*(2), 255–264.

Goldin-Meadow, S., & Alibali, M. W. (2014). Gesture's role in speaking, learning, and creating language Susan. *Bone*, *23*(1), 1–7. https://doi.org/10.1146/annurev-psych-113011-143802.Gesture

Goldstein, M. H., Shwade, J. A., & Bornstein, M. H. (2009). The value of vocalizing: Five-month-old infants associate their own noncry vocalizations with responses from caregivers Michael. *Child Development*, *80*(3), 636–644. https://doi.org/10.1111/j.1467-8624.2009.01287.x.The

Goldstein, M., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, *19*(5), 515–523. https://doi.org/10.1111/j.1467-9280.2008.02117.x

Golinkoff, R. M., Can, D. D., Soderstrom, M., & Hirsh-Pasek, K. (2015). (Baby)Talk to Me: The Social Context of Infant-Directed Speech and Its Effects on Early Language Acquisition. *Current Directions in Psychological Science*, *24*(5), 339–344. https://doi.org/10.1177/0963721415595345

Gottlieb, G. (2000). Environmental and behavioral influences on gene activity. *Current Directions in Psychological Science*, *9*(3), 93–97. https://doi.org/10.1111/1467-8721.00068

Graham, K. E., Wilke, C., Lahiff, N. J., & Slocombe, K. E. (2019). Scratching beneath the surface: Intentionality in great ape signal production. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1789). https://doi.org/10.1098/rstb.2018.0403

Gratier, M. (2003). Expressive timing and interactional synchrony between mothers and infants: Cultural similarities, cultural differences, and the immigration exprerience. *Cognitive Development*, *18*(4), 533–554. https://doi.org/10.1016/j.cogdev.2003.09.009

Greenfield, P. M. (2009). Linking Social Change and Developmental Change: Shifting Pathways of Human Development. *Developmental Psychology*, *45*(2), 401–418. https://doi.org/10.1037/a0014726

Grice, H. P. (1969). Utterer's meaning and intentions. *The Philosophical Review*, *78*(2), 147–177.

Gros-Louis, J., West, M. J., & King, A. P. (2014). Maternal responsiveness and the development of directed vocalizing in social interactions. *Infancy*, *19*(4), 385–408. https://doi.org/10.1111/infa.12054

Gros-Louis, J., & Wu, Z. (2012). Twelve-month-olds' vocal production during pointing in naturalistic interactions: Sensitivity to parents' attention and responses. *Infant Behavior and Development*, *35*(4), 773–778. https://doi.org/10.1016/j.infbeh.2012.07.016

Grünloh, T., & Liszkowski, U. (2015). Prelinguistic vocalizations distinguish pointing acts. *Journal of Child Language*, *42*(06), 1312–1336. https://doi.org/10.1017/s0305000914000816

Guellaï, B., Callin, A., Bevilacqua, F., Schwarz, D., Pitti, A., Boucenna, S., & Gratier, M. (2019). Sensus communis: Some perspectives on the origins of non-synchronous cross-sensory associations. *Frontiers in Psychology*, *10*(MAR). https://doi.org/10.3389/fpsyg.2019.00523

Guellai, B., & Streri, A. (2011). *Cues for Early Social Skills : Direct Gaze Modulates Newborns ' Recognition of Talking Faces*. *6*(4). https://doi.org/10.1371/journal.pone.0018610

Guidetti, M. (2002). The emergence of pragmatics: Forms and functions of conventional gestures in young French children. *First Language*, *22*(3), 265–285. https://doi.org/10.1177/014272370202206603

Guidetti, M. (2005). Yes or no ? How young French children combine gestures and speech to agree and refuse *. *Journal of Child Language*, *32*, 911–924.

https://doi.org/10.1017/S0305000905007038

Guidetti, M., Fibigerova, K., & Colletta, J.-M. (2014). Gestures and multimodal development. In M. Seyfeddinipur & M. Gullberg (Eds.), *From Gesture in Conversation to Visible Action as Utterance* (pp. 351–370). https://doi.org/10.1075/z.188

Gupta, S., & Sinha, A. (2019). Gestural communication of wild bonnet macaques in the Bandipur National Park, Southern India. *Behavioural Processes*, *168*(May 2018). https://doi.org/10.1016/j.beproc.2019.103956

Hamilton, W. D. (1964). The genetical evolution of social behavior. II. *Journal of Theoretical Biology, 7*(1), 17–52. https://doi.org/10.4324/9780203790427-5

Harold, M. P., & Barlow, S. M. (2013). Effects of environmental stimulation on infant vocalizations and orofacial dynamics at the onset of canonical babbling. *Infant Behavior and Development*, *36*(1), 84–93. https://doi.org/10.1016/j.infbeh.2012.10.001

Harrison, X. A. (2015). A comparison of observation-level randomeffect and Beta-Binomial models for modelling overdispersion in Binomial data in ecology & evolution. *PeerJ*, *2015*(7). https://doi.org/10.7717/peerj.1114

Hattori, Y., & Tomonaga, M. (2020). Rhythmic swaying induced by sound in chimpanzees (Pan troglodytes). *Proceedings of the National Academy of Sciences of the United States of America*, *117*(2), 936–942. https://doi.org/10.1073/pnas.1910318116

Hattori, Y., Tomonaga, M., & Matsuzawa, T. (2013). Spontaneous synchronized tapping to an auditory rhythm in a chimpanzee. *Scientific Reports*, *3*. https://doi.org/10.1038/srep01566

Hawkes, K., Chisholm, J., Fairbanks, L., Johow, J., Kalcher-Sommersguter, E., Liebal, K., & Finlay, B. (2017). *Primate Infancies: Causes and consequences of varying care*. The MIT Press.

Hedges, S. B., Marin, J., Suleski, M., Paymer, M., & Kumar, S. (2015). Tree of life reveals clock-like speciation and diversification. *Molecular Biology and Evolution*, *32*(4), 835–845. https://doi.org/10.1093/molbev/msv037

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Hewes, G. W. (1992). Primate Communication and the Gestural Origin of Language. *Current Anthropology*, *33*(S1), 65–84. https://doi.org/10.1086/204019

Hillbrink, E. (2015). Early developmental changes in the timing of turn-taking: a longitudinal study of mother–infant interaction. *Frontiers in Psychology*, *6*. https://doi.org/doi: 10.3389/fpsyg.2015.01492

Hobaiter, C., Byrne, R. W., & Zuberbühler, K. (2017). Wild chimpanzees' use of single and combined vocal and gestural signals. *Behavioral Ecology and Sociobiology*, *71*(6). https://doi.org/10.1007/s00265-017-2325-1

Hobaiter, Catherine, & Byrne, R. (2011a). The gestural repertoire of the wild chimpanzee. *Animal Cognition*, *14*(5), 745–767. https://doi.org/10.1007/s10071-011-0409-2

Hobaiter, Catherine, & Byrne, R. (2011b). Serial gesturing by wild chimpanzees: Its nature and function for communication. *Animal Cognition*, *14*(6), 827–838. https://doi.org/10.1007/s10071-011-0416-3

Hobaiter, Catherine, & Byrne, R. W. (2014). Report The Meanings of Chimpanzee Gestures. *Current Biology*, *24*(14), 1596–1600. https://doi.org/10.1016/j.cub.2014.05.066

Hopkins, W. D., Taglialatela, J. P., & Leavens, D. A. (2007). Chimpanzees differentially produce novel vocalizations to capture the attention of a human. *Animal Behaviour*, *73*(2), 281–286. https://doi.org/10.1016/j.anbehav.2006.08.004

Horowitz, A. (2009). Attention to attention in domestic dog (Canis familiaris) dyadic play. *Animal Cognition*, *12*(1), 107–118. https://doi.org/10.1007/s10071-008-0175-y

Hostetter, A. B., Cantero, M., & Hopkins, W. D. (2001). Differential use of vocal and gestural communication by chimpanzees (Pan troglodytes) in response to the attentional status of a human (Homo sapiens). *Journal of Comparative Psychology*, *115*(4), 337–343. https://doi.org/10.1037/0735-7036.115.4.337

Hostetter, A. B., Russell, J. L., Freeman, H., & Hopkins, W. D. (2007). Now you see me, now you don't: Evidence that chimpanzees understand the role of the eyes in attention. *Animal Cognition*, *10*(1), 55–62. https://doi.org/10.1007/s10071-006-0031-x

Hsu, H.-C., Fogel, A., & Messinger, D. (2001). Infant non-distress vocalization during mother-infant face-to-face interaction: Factors associated with quantitative and qualitative differences. *Infant Behavior and Development*, *24*, 107–128.

Igualada, A., Bosch, L., & Prieto, P. (2015). Language development at 18 months is related to multimodal communicative strategies at 12 months. *Infant Behavior and Development*, *39*, 42–52. https://doi.org/10.1016/j.infbeh.2015.02.004

Iverson, J. M., & Fagan, M. K. (2004). *Infant Vocal – Motor Coordination : Precursor to the Gesture – Speech System ? 75*(4), 1053–1066.

Iverson, J. M., & Goldin-meadow, S. (2005). Gesture Paves the Way for Language Development. *Psychological Science*, *16*(5), 367–371.

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science*, *16*(5), 367–371. https://doi.org/10.1111/j.0956-7976.2005.01542.x

Jackson-Maldonado, D. (2012). MacArthur-Bates Communicative Development Inventories. *The Encyclopedia of Applied Linguistics*. https://doi.org/10.1002/9781405198431.wbeal0745

Jensen, T. S., Bøggild-Andersen, B., Schmidt, J., Ankerhus, J., & Hansen, E. (1988). Perinatal Risk Factors and First-Year Vocalizations: Influence on Preschool Language and Motor Performance. *Developmental Medicine & Child Neurology*, *30*(2), 153–161. https://doi.org/10.1111/j.1469-8749.1988.tb04746.x

Kagitcibasi, C. (1996). *Family and human development across cultures: A view from the other side*. 10 Industrial Avenue, Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Kaminski, J., Call, J., & Tomasello, M. (2004). Body orientation and face

orientation: Two factors controlling apes' begging behavior from humans. *Animal Cognition*, *7*(4), 216–223. https://doi.org/10.1007/s10071-004-0214-2

Kaplan, G. (2011). Pointing gesture in a bird- merely instrumental or a cognitively complex behavior? *Current Zoology*, *57*(4), 453–467. https://doi.org/10.1093/czoolo/57.4.453

Keller, H., Borke, J., Staufenbiel, T., Yovsi, R. D., Abels, M., Papaligoura, Z., … Su, Y. (2009). Distal and proximal parenting as alternative parenting strategies during infants' early months of life: A cross-cultural study. *International Journal of Behavioral Development*, *33*(5), 412–420. https://doi.org/10.1177/0165025409338441

Keller, H., Voelker, S., & Yovsi, R. D. (2005). Conceptions of parenting in different cultural communities : The case of west african Nso and northern german women. *Social Development*, *14*(1), 158–180.

Kendon, A. (2011). Some modern considerations for thinking about language evolution: A discussion of The Evolution of Language by Tecumseh Fitch. *Public Journal of Semiotics*, *3*(1), 79–108. https://doi.org/10.37693/pjos.2011.3.8832

Kersken, V., Carlos, J., Ulf, G., Adrian, L., & Hobaiter, C. (2018). A gestural repertoire of 1- to 2-year-old human children : in search of the ape gestures. *Animal Cognition*, *0*(0), 0. https://doi.org/10.1007/s10071-018-1213-z

Kersken, V., Gómez, J. C., Liszkowski, U., Soldati, A., & Hobaiter, C. (2019). A gestural repertoire of 1- to 2-year-old human children: in search of the ape gestures. *Animal Cognition*, *22*(4), 577–595. https://doi.org/10.1007/s10071-018-1213-z

Kilbride, P. L. (1980). Sensiromotor behavior of baganda and Samia infants.A controlled comparison. *Journal of Cross-Cultural Psychology*, *11*, 131–152.

Kita, S. (2003). Pointing: A foundational building block of human communication. In *Pointing* (Psychology, pp. 9–16).

Klein, J. T., Shepherd, S. V., & Platt, M. L. (2009). Social Attention and the Brain. *Current Biology*, *19*(20), R958–R962. https://doi.org/10.1016/j.cub.2009.08.010

Kumar, S., Stecher, G., Suleski, M., & Blair Hedges, S. (2017). TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Molecular Biology and Evolution*, *34*(7), 1812–1819. https://doi.org/10.1093/MOLBEV/MSX116

Kumar, S., Suleski, M., Craig, J. M., Kasprowicz, A. E., Sanderford, M., Li, M., … Hedges, S. B. (2022). TimeTree 5: An Expanded Resource for Species Divergence Times. *Molecular Biology and Evolution*, *39*(8), 1–6. https://doi.org/10.1093/molbev/msac174

Lameira, A. R., & Call, J. (2018). Time-space-displaced responses in the orangutan vocal system. *Science Advances*, *4*(11), 1–8. https://doi.org/10.1126/sciadv.aau3401

Lameira, A. R., Hardus, M. E., Kowalsky, B., de Vries, H., Spruijt, B. M., Sterck, E. H. M., … Wich, S. A. (2013). Orangutan (Pongo spp.) whistling and implications for the emergence of an open-ended call repertoire: A replication

and extension . *The Journal of the Acoustical Society of America*, *134*(3), 2326–2335. https://doi.org/10.1121/1.4817929

Laporte, M. N. C., & Zuberbühler, K. (2011). The development of a greeting signal in wild chimpanzees. *Developmental Science*, *14*(5), 1220–1234. https://doi.org/10.1111/j.1467-7687.2011.01069.x

Lausberg, H., & Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods*, *41*(3), 841–849. https://doi.org/10.3758/BRM.41.3.841

Lavelli, M., & Fogel, A. (2005). Developmental changes in the relationship between the infant's attention and emotion during early face-to-face communication: The 2-month transition. *Developmental Psychology*, *41*(1), 265–280. https://doi.org/10.1037/0012-1649.41.1.265

Leavens, D. A. (2004). Manual deixis in apes and humans. *Interaction Studies*, *5*(3), 387–408. https://doi.org/10.1075/is.5.3.05lea

Leavens, D. A., Hopkins, W. D., & Thomas, R. K. (2004). Referential Communication by Chimpanzees (Pan Troglodytes). *Journal of Comparative Psychology*, *118*(1), 48–57. https://doi.org/10.1037/0735-7036.118.1.48

Leavens, D. A., Hostetter, A. B., Wesley, M. J., & Hopkins, W. D. (2004). Tactical use of unimodal and bimodal communication by chimpanzees, Pan troglodytes. *Animal Behaviour*, *67*(3), 467–476. https://doi.org/10.1016/j.anbehav.2003.04.007

Leavens, D. A., Russell, J. L., & Hopkins, W. D. (2005). Intentionality as measured in the persistence and elaboration of communication by chimpanzees (Pan troglodytes). *Child Development*, *76*(1), 291–306.

Leavens, D. A., Russell, J. L., & Hopkins, W. D. (2010). Multimodal communication by captive chimpanzees (Pan troglodytes). *Animal Cognition*, *13*(1), 33–40. https://doi.org/10.1007/s10071-009-0242-z

Legare, C. H., Clegg, J. M., & Wen, N. J. (2018). A Metatheory for Cognitive Development (or "Piaget is Dead" Revisited). *Child Development*, *89*(6), 2288–2302. https://doi.org/10.1111/cdev.13019

Liebal, K., Call, J., & Tomasello, M. (2004). Use of gesture sequences in chimpanzees. *American Journal of Primatology*, *64*(4), 377–396. https://doi.org/10.1002/ajp.20087

Liebal, K., Call, J., Tomasello, M., & Pika, S. (2004). To move or not to move: How apes adjust to the attentional state of others. *Interaction StudiesInteraction Studies Social Behaviour and Communication in Biological and Artificial Systems*, *5*(2), 199–219. https://doi.org/10.1075/is.5.2.03lie

Liebal, K., Pika, S., & Tomasello, M. (2006). Gestural communication of orangutans (Pongo pygmaeus). *Gesture*, *6*(1), 1–38. https://doi.org/10.1075/gest.6.1.02lie

Lin, H., & Green, J. (2009). Infants' expressive behaviors to mothers and unfamiliar partners during face-to-face interactions from 4 to 10 months. *Infant Behavior and Development*, *32*, 275–285.

Liszkowski, U. (2006). Infant pointing at 12 months: communicative goals, motives, and social-cognitive abilities. In *Roots of human sociality* (Routledge, pp.

153–178).

Liszkowski, U., Albrecht, K., Carpenter, M., & Tomasello, M. (2008). Infants ' visual and auditory communication when a partner is or is not visually attending. *Infant Behavior & Development*, *31*, 157–167. https://doi.org/10.1016/j.infbeh.2007.10.011

Liszkowski, U., Brown, P., Callaghan, T., & Takada, A. (2012). *A Prelinguistic Gestural Universal of Human Communication*. *36*, 698–713. https://doi.org/10.1111/j.1551-6709.2011.01228.x

Liszkowski, U., Brown, P., Callaghan, T., Takada, A., & de Vos, C. (2012). A Prelinguistic Gestural Universal of Human Communication. *Cognitive Science*, *36*(4), 698–713. https://doi.org/10.1111/j.1551-6709.2011.01228.x

Liszkowski, U., Carpenter, M., Henning, A., Striano, T., & Tomasello, M. (2004). Twelve-month-olds point to share attention and interest. *Developmental Science*, *7*(3), 297–307. https://doi.org/10.1111/j.1467-7687.2004.00349.x

Liszkowski, U., Carpenter, M., Striano, T., & Tomasello, M. (2006). 12- and 18-Month-Olds Point To Provide Information for Others. *Journal of Cognition and Development*, *7*(2), 173–187. https://doi.org/10.1207/s15327647jcd0702_2

Liszkowski, U., Carpenter, M., & Tomasello, M. (2007). Pointing out new news, old news, and absent referents at 12 months of age. *Developmental Science*, *10*(2), 1–7. https://doi.org/10.1111/j.1467-7687.2006.00552.x

Liszkowski, U., & Rüther, J. (2021). Ontogenetic origins of infant pointing. *The Oxford Handbook of Human Symbolic Evolution*. https://doi.org/10.1093/oxfordhb/9780198813781.013.31

Locke, J. L. (2006). Parental selection of vocal behavior. *Human Nature*, *17*(2), 155–168. https://doi.org/10.1007/s12110-006-1015-x

Locke, J. L. (2017). Development of the Capacity for Spoken Language. In P. Fletcher & B. MacWhinney (Eds.), *The Handbook of Child Language*. https://doi.org/10.1111/b.9780631203124.1996.x

Locke, J. L., & Bogin, B. (2006). Language and life history: A new perspective on the development and evolution of human language. *Behavioral and Brain Sciences*, *29*(3), 259–280. https://doi.org/10.1017/S0140525X0600906X

Long, H. L., Bowman, D. D., Yoo, H., Burkhardt-Reed, M. M., Bene, E. R., & Kimbrough Oller, D. (2020). Social and endogenous infant vocalizations. *PLoS ONE*, *15*(8 August 2020), 1–18. https://doi.org/10.1371/journal.pone.0224956

Ma, W., Golinkoff, R. M., Houston, D. M., & Hirsh-Pasek, K. (2011). Word learning in infant-and adult-directed speech. *Language Learning and Development*, *7*(3), 185–201. https://doi.org/10.1080/15475441.2011.579839

Maille, A., Engelhart, L., Bourjade, M., & Blois-Heulin, C. (2012). To Beg, or Not to Beg? That Is the Question: Mangabeys Modify Their Production of Requesting Gestures in Response to Human's Attentional States. *PLoS ONE*, (7), e41197. https://doi.org/10.1371/journal.pone.0041197

Martinez, M.-L. (1989). Le socio-constructivisme et l'innovation en français. *Pratiques*, *63*(1), 37–62. https://doi.org/10.3406/prati.1989.1514

Mastin, J. D., & Vogt, P. (2016). Infant engagement and early vocabulary

development: A naturalistic observation study of Mozambican infants from 1;1 to 2;1. *Journal of Child Language*, *43*(2), 235–264. https://doi.org/10.1017/S0305000915000148

McCune, L., Lennon, E. M., & Greenwood, A. (2020). Gestures, grunts, and words: Development in a dynamic system. *First Language*. https://doi.org/10.1177/0142723720966820

Mccune, L., & Zlatev, J. (2015). Dynamic systems in semiotic development: The transition toreference. *Cognitive Development*, *36*, 10. https://doi.org/10.1016/j.cogdev.2015.09.010

Mcgillion, M. L., Herbert, J. S., Pine, J., Vihman, M., DePaolis, R., Keren-Portnoy, T., & Matthews, D. (2017). What Paves the Way to Conventional Language? The Predictive Value of Babble, Pointing, and Socioeconomic Status. *Child Development Perspectives*, 2013–2015.

Meguerditchian, A., Cochet, H., & Vauclair, J. (2011). *From gesture to language: Ontogenetic and phylogenetic perspectives on gestural communication and its cerebral lateralization*.

Meguerditchian, A., Taglialatela, J. P., Leavens, D. A., & Hopkins, W. D. (2014). Why vocal production of atypical sounds in apes and its cerebral correlates have a lot to say about the origin of language. *Behavioral and Brain Sciences*, *72*, 529–546. https://doi.org/doi:10.1017/S0140525X13004135 Adrien

Meguerditchian, A., & Vauclair, J. (2014). Communicative Signaling , Lateralization and Brain Substrate in Nonhuman Primates : Toward a Gestural or a Multimodal Origin of Language ? *Humana. Mente Journal of Philosophical Studies*, *27*, 135–160.

Mehr, S. A., & Krasnow, M. M. (2017). Parent-offspring conflict and the evolution of infant-directed song. *Evolution and Human Behavior*, *38*(5), 674–684. https://doi.org/10.1016/j.evolhumbehav.2016.12.005

Mithen, S. J. (2005). *The singing Neanderthals : the origin of music, language, mind and body* (Weidenfeld and Nicholson, ed.). London: Phoenix.

Molesti, S., Bourjade, M., & Meguerditchian, A. (2017). *Gestural Communication in Olive Baboons (Papio anubis): Repertoire and Properties*. Retrieved from https://hal.archives-ouvertes.fr/hal-01560551

Molesti, S., Meguerditchian, A., & Bourjade, M. (2020). Gestural communication in olive baboons (Papio anubis): repertoire and intentionality. *Animal Cognition*, *23*(1), 19–40. https://doi.org/10.1007/s10071-019-01312-y

Moreno-Núñez, A., Rodríguez, C., & Miranda-Zapata, E. (2020). Getting away from the point: The emergence of ostensive gestures and their functions. *Journal of Child Language*, *47*(3), 556–578. https://doi.org/10.1017/S0305000919000606

Murillo, E., & Casla, M. (2020). Multimodal representational gestures in the transition to multi-word productions. *Infancy*, *26*(1), 104–122. https://doi.org/10.1111/infa.12375

Murillo, E., Montero, I., & Casla, M. (2019). *Rhythmic movements with objects at 9 months are related to proximal deictic gestures at 12 months* .

Murillo, E., Montero, I., & Casla, M. (2021). On the Multimodal Path to Language:

The Relationship Between Rhythmic Movements and Deictic Gestures at the End of the First Year. *Frontiers in Psychology*, *12*(February), 1–8. https://doi.org/10.3389/fpsyg.2021.616812

Murillo, E., Ortega, C., Otones, A., Rujas, I., & Casla, M. (2018). Changes in the synchrony of multimodal communication in early language development. *Journal of Speech, Language, and Hearing Research*, *61*(9), 2235–2245. https://doi.org/10.1044/2018_JSLHR-L-17-0402

Murphy, S., Kochanek, K., Xu, J., & Heron, M. (2010). *National Vital Statistics Reports. Deaths: Final Data for 2007*. *58*(19). Retrieved from http://www.mendeley.com/research/deaths-final-data-for-2002-national-vital-statistics-reports/

Nakata, T., & Trehub, S. E. (2004). Infants' responsiveness to maternal speech and singing. *Infant Behavior and Development*, *27*(4), 455–464. https://doi.org/10.1016/j.infbeh.2004.03.002

Nishida, T. (2012). *Chimpanzees of the lakeshore: Natural history and culture at Mahale* (Cambridge University Press, ed.). https://doi.org/https://doi.org/10.1017/CBO9781139059497

Northrup, J. B., & Iverson, J. M. (2020a). *Multimodal coordination of vocal and gaze behavior in mother-infant dyads across the first year of life*. https://doi.org/10.1111/infa.12369

Northrup, J. B., & Iverson, J. M. (2020b). Multimodal coordination of vocal and gaze behavior in mother–infant dyads across the first year of life. *Infancy*, *25*, 952–972.

Northrup, J. B., & Iverson, J. M. (2020c). The Development of Mother–Infant Coordination Across the First Year of Life. *Developmental Psychology*, *56*(2), 221–236. https://doi.org/10.1037/dev0000867.supp

Oller, D. K. (2000). *The Emergence of the Speech Capacity*.

Oller, D. K., Buder, E. H., Ramsdell, H. L., Warlaumont, A. S., Chorna, L., & Bakeman, R. (2013). Functional flexibility of infant vocalization and the emergence of language. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(16), 6318–6323. https://doi.org/10.1073/pnas.1300337110

Oller, D. K., Eilers, R. E., Neal, A. R., & Cobo-Lewis, A. B. (1998). Late onset canonical babbling: A possible early marker of abnormal development. *American Journal on Mental Retardation*, Vol. 103, pp. 249–263. https://doi.org/10.1352/0895-8017(1998)103<0249:LOCBAP>2.0.CO;2

Oller, D. K., Eilers, R. E., Neal, A. R., & Schwartz, H. K. (1999). Precursors to speech in infancy. *Journal of Communication Disorders*, *32*(4), 223–245. https://doi.org/10.1016/s0021-9924(99)00013-1

Oller, D. K., Griebel, U., & Warlaumont, A. S. (2016). Vocal Development as a Guide to Modeling the Evolution of Language. *Topics in Cognitive Science*, *8*(2), 382–392. https://doi.org/10.1111/tops.12198

Oller, D. K., Ramsay, G., Bene, E., Long, H. L., & Griebel, U. (2021). Protophones, the precursors to speech, dominate the human infant vocal landscape. *Philosophical Transactions of the Royal Society B: Biological Sciences*,

376(1836). https://doi.org/10.1098/rstb.2020.0255

Oller, K., Buder, E., Ramsdell, H., Warlaumont, A., Chorna, L., & Bakeman, R. (2013). Functional flexibility of infant vocalization and the emergence of language. *Proceedings of the National Academy of Sciences*, *110*(16), 6318–6323. https://doi.org/10.1073/pnas.1300337110

Pearce, N. (2016). Analysis of matched case-control studies. *BMJ (Online)*, *352*, 1–4. https://doi.org/10.1136/bmj.i969

Pika, S., & Bugnyar, T. (2011). The use of referential gestures in ravens (Corvus corax) in the wild. *Nature Communications*, *2*(1), 560–565. https://doi.org/10.1038/ncomms1567

Pika, S., & Liebal, K. (2006). *Differences and Similarties Between the Natural Gestural Communication of the Great Apes and Human Children*. 267–274. https://doi.org/10.1142/9789812774262_0034

Pika, S., & Mitani, J. (2006). Referential gestural communication in wild chimpanzees (Pan troglodytes). *Current Biology*, 191–192.

Poss, S. R., Kuhar, C., Stoinski, T. S., & Hopkins, W. D. (2006). Differential Use of Attentional and Visual Communicative Signaling by Orangutans (Pongo pygmaeus) and Gorillas (Gorilla gorilla) in Response to the Attentional Status of a Human. *Expedition Field Techniques BATS*, *7*(September 2006), 1–6. https://doi.org/10.1002/ajp

Povinelli, D. J., Eddy, T. J., Hobson, R. P., & Tomasello, M. (1996). What young chimpanzees know about seeing. *Monographs of the Society for Research in Child Development*, *61*(3). https://doi.org/10.2307/1166159

Povinelli, D. J., Theall, L. A., Reaux, J. E., & Dunphy-Lelii, S. (2003). Chimpanzees spontaneously alter the location of their gestures to match the attentional orientation of others. *Animal Behaviour*, *66*(1), 71–79. https://doi.org/10.1006/anbe.2003.2195

Premack, D., & Woodruff, G. (1978). Premack and Woodruff : Chimpanzee theory of mind. *Behavioral and Brain Sciences*, *4*, 515–526.

Proops, L., & McComb, K. (2010). Attributing attention: The use of human-given cues by domestic horses (Equus caballus). *Animal Cognition*, *13*(2), 197–205. https://doi.org/10.1007/s10071-009-0257-5

Pusey, A. E. (1990). Behavioural Changes at Adolescence in Chimpanzees. *Behaviour*, *115*(3–4), 203–246. https://doi.org/10.1163/156853990X00581

Reynolds Losin, E. A., Russell, J. L., Freeman, H., Meguerditchian, A., & Hopkins, W. D. (2008). Left hemisphere specialization for oro-facial movements of learned vocal signals by captive chimpanzees. *PLoS ONE*, *3*(6), 1–7. https://doi.org/10.1371/journal.pone.0002529

Reynolds, V. (2006). The Chimpanzees of the Budongo Forest. *International Journal of Primatology*, *27*(5). https://doi.org/10.1007/s10764-006-9075-7

Rodrigues, E. D. (2021). *Operationalizing Intentionality in Primate Communication : Social and Ecological Considerations Content courtesy of Springer Nature , terms of use apply . Rights reserved . Content courtesy of Springer Nature , terms of use apply . Rights reserved* .

Rodrigues, E. D., Marôco, J., & Frota, S. (2021). Communicative gestures in 7–12-

month infants: A phylogenetic comparative approach. *Infant and Child Development*, *30*(5), 1–11. https://doi.org/10.1002/icd.2262

Rodríguez, C., Moreno-Núñez, A., Basilio, M., & Sosa, N. (2015). Ostensive gestures come first: Their role in the beginning of shared reference. *Cognitive Development*, *36*, 142–149. https://doi.org/10.1016/j.cogdev.2015.09.005

Rossano, F. (2013). Gaze in Conversation. In *The handbook of conversation analysis* (p. 308).

Rüther, J., & Liszkowski, U. (2020). Ontogenetic Emergence of Cognitive Reference Comprehension. *Cognitive Science*, *44*(7). https://doi.org/10.1111/cogs.12869

Salomo, D., & Liszkowski, U. (2013). *Sociocultural Settings In fl uence the Emergence of Prelinguistic Deictic Gestures*. *84*(4), 1296–1307. https://doi.org/10.1111/cdev.12026

Scaife, M., & Bruner, J. (1975). *The capacity for joint attention in the infant*. *253*, 265–266.

Schneider, C., Call, J., & Liebal, K. (2012). Onset and Early Use of Gestural Communication in Nonhuman Great Apes. *American Journal of Primatology*, *74*(1), 102–113. https://doi.org/10.1002/ajp.21011

Scola, C., Bourjade, M., & Jover, M. (2015). Social interaction is associated with changes in infants' motor activity. *Socioaffective Neuroscience & Psychology*, *5*(1), 28256. https://doi.org/10.3402/snp.v5.28256

Scott-Phillips, T. C. (2015). Nonhuman primate communication, pragmatics, and the origins of language. *Current Anthropology*, *56*(1), 56–80. https://doi.org/10.1086/679674

Searle, J. . (1965). What is a speech act. *Perspectives in the Philosophy of Language: A Concise Anthology*, *2000*, 253–268.

Shapiro-Mendoza, C., Tomashek, K., Kotelchuck, M., Barfield, W., Weiss, J., & Evans, S. (2006). Risk Factors for Neonatal Morbidity and Mortality Among "Healthy," Late Preterm Newborns. *Seminars in Perinatology*, *30*(2), 54–60. https://doi.org/10.1053/j.semperi.2006.02.002

Siegal, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences (2nd ed)* (McGraw Hil). Boston.

Sperber, D., & Wilson, D. (1995). *Relevance Communication and Cognition Second Edition* (Second Edi). Oxford/Cambridge: Blackwell Publishers.

Spinelli, M., Fasolo, M., & Mesman, J. (2017). Does prosody make the difference? A meta-analysis on relations between prosodic aspects of infant-directed speech and infant outcomes. *Developmental Review*, *44*, 1–18. https://doi.org/10.1016/j.dr.2016.12.001

Stern, D. (1999). Vitality contours: The temporal contour of feelings as a basic unit for constructing the infant's social experience. *Early Social Cognition: Understanding Others in the First Months of Life*, 67–80.

Stern, D., Beebe, B., Jaffe, J., & Benett, S. (1977). The infant's stimulus world during social interaction: A study of caregiver behaviors with particular reference to repetition and timing. *Studies in Mother-Infant Interaction*, 177–202.

Stoel-Gammon, C. (2011). Relationships between lexical and phonological

development in young children. *Journal of Child Language*, *38*(1), 1–34. https://doi.org/10.1017/S0305000910000425

Stolt, S., Lehtonen, L., Haataja, L., & Lapinleimu, H. (2012). Development and predictive value of early vocalizations in very-low-birth-weight children: A longitudinal study. *Clinical Linguistics and Phonetics*, *26*(5), 414–427. https://doi.org/10.3109/02699206.2011.648365

Stumpf, R. M., Emery Thompson, M., Muller, M. N., & Wrangham, R. W. (2009). The context of female dispersal in Kanyawara chimpanzees. *Behaviour*, *146*(4–5), 629–656. https://doi.org/10.1163/156853909X413853

Super, C., & Harkness, S. (2015). Charting infant development: Milestones along the way. In *The Oxford handbook of human development and culture: An interdisciplinary perspective* (L.A Jensen, pp. 79–93). https://doi.org/https://doi.org/10.1093/oxfordhb/9780199948550.013.6

Symonds, M. R. E., & Moussalli, A. (2011). A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behavioral Ecology and Sociobiology*, *65*(1), 13–21. https://doi.org/10.1007/s00265-010-1037-6

Taglialatela, J. P., Russell, J. L., Schaeffer, J. A., & Hopkins, W. D. (2008). Communicative Signaling Activates "Broca's" Homolog in Chimpanzees. *Current Biology*, *18*(5), 343–348. https://doi.org/10.1016/j.cub.2008.01.049

Taglialatela, J. P., Russell, J. L., Schaeffer, J. A., & Hopkins, W. D. (2011). *Chimpanzee Vocal Signaling Points to a Multimodal Origin of Human Language*. *6*(4), 1–7. https://doi.org/10.1371/journal.pone.0018852

Tempelmann, S., Kaminski, J., & Liebal, K. (2011). Focus on the essential: All great apes know when others are being attentive. *Animal Cognition*, *14*(3), 433–439. https://doi.org/10.1007/s10071-011-0378-5

Theobald, E. (2018). Students are rarely independent: When, why, and how to use random effects in discipline-based education research. *CBE Life Sciences Education*, *17*(3), 1–12. https://doi.org/10.1187/cbe.17-12-0280

Tomasello, M. (1995). Understanding the self as social agent. In *Advances in Psychology* (Vol. 112). https://doi.org/10.1016/S0166-4115(05)80024-X

Tomasello, M. (2008). *Origins of Human Communication* (The MIT Press, Ed.). https://doi.org/https://doi.org/10.7551/mitpress/7551.001.0001

Tomasello, M., & Call, J. (2018). Thirty years of great ape gestures. *Animal Cognition*, (0123456789). https://doi.org/10.1007/s10071-018-1167-1

Tomasello, M., Call, J., Nagell, K., Olguin, R., & Carpenter, M. (1994). The learning and use of gestural signals by young chimpanzees: A trans-generational study. *Primates*, *35*(2), 137–154. https://doi.org/10.1007/BF02382050

Tomasello, M., Call, J., Warren, J., Frost, G. T., Carpenter, M., & Nagell, K. (1997). The ontogeny ofchimpanzee gestural signals: A comparison across groups and generations. *Evolution of Communicati*, *1*(2), 223–259. https://doi.org/10.1075/eoc.l.2.04tom

Townsend, S. W., Koski, S. E., Byrne, R. W., Slocombe, K. E., Bickel, B., Boeckle, M., … Manser, M. B. (2017). Exorcising Grice's ghost: an empirical approach to studying intentional communication in animals. *Biological Reviews*, *92*(3),

1427–1433. https://doi.org/10.1111/brv.12289

Trevarthen, C., & Aitken, K. J. (2003). Intersubjectivité chez le nourrisson: Recherche, théorie et application clinique. *Devenir*, *15*(4), 309–428. https://doi.org/10.3917/dev.034.0309

Trivers, R. L. (1974). *Parent-Offspring Conflict. 264*, 249–264.

Tronick, E. (2003). Things still to be done on the still-face effect. *Infancy*, *4*(4), 475–482. https://doi.org/10.1207/S15327078IN0404_02

Tronick, E., Als, H., Adamson, L., Wise, S., & Brazelton, T. B. (1978). The Infant's Response to Entrapment between Contradictory Messages in Face-to-Face Interaction. *Journal of the American Academy of Child Psychiatry*, *17*(1), 1–13. https://doi.org/10.1016/S0002-7138(09)62273-1

Van de Pol, M., & Wright, J. (2009). A simple method for distinguishing within- versus between-subject effects using mixed models. *Animal Behaviour*, *77*(3), 753–758. https://doi.org/10.1016/j.anbehav.2008.11.006

Volterra, V., & Antinucci, F. (1979). Negation in child language: A pragmatic study. *Developmental Pragmatics*, 281–303.

Vygotsky, L. S. (1978). *Mind in society: Development of higher psychological processes*. Harvard University Press.

Wacewicz, S., & Zywiczynski, P. (2016). *Language & Communication The multimodal origins of linguistic communication*. *54*, 1–8. https://doi.org/10.1016/j.langcom.2016.10.001

Waddington, C. (1942). Epigenotype. *Endeavour*.

Waddington, C. (2012). The epigenotype. 1942. *International Journal of Epidemiology*, *41*, 10–13. https://doi.org/10.1093/ije/dyr184

Waddington, C. (2014). *The epigenetics of birds*. Cambridge University Press.

Wallez, C., Schaeffer, J., Meguerditchian, A., Vauclair, J., Schapiro, S. J., & Hopkins, W. D. (2012). Contrast of hemispheric lateralization for oro-facial movements between learned attention-getting sounds and species-typical vocalizations in chimpanzees: Extension in a second colony. *Brain and Language*, *123*(1), 75–79. https://doi.org/10.1016/j.bandl.2012.07.002

Warlaumont, A. S., Richards, J. A., Gilkerson, J., & Oller, D. K. (2015). A social feedback loop for speech development and its reduction in autism. *Psychol Sci.*, *25*(7), 1314–1324. https://doi.org/10.1177/0956797614531023.A

Watts, D., & Pussey, A. (1993). *Behavior of juvenile and adolescent great apes*. New York: Pereira,M. E. Fair- banks, L. A.

Wittig, R. M., & Boesch, C. (2003). Food competition and linear dominance hierarchy among female chimpanzees of the Taï National Park. *International Journal of Primatology*, *24*(4), 847–867. https://doi.org/10.1023/A:1024632923180

Wörmann, V., Holodynski, M., Kärtner, J., & Keller, H. (2014). The Emergence of Social Smiling: The Interplay of Maternal and Infant Imitation During the First Three Months in Cross-Cultural Comparison. *Journal of Cross-Cultural Psychology*, *45*(3), 339–361. https://doi.org/10.1177/0022022113509134

Wu, Z., & Gros-Louis, J. (2014). Infants' prelinguistic communicative acts and maternal responses: Relations to linguistic development. *First Language*, *34*(1),

72–90. https://doi.org/10.1177/0142723714521925

Wu, Z., & Gros-Louis, J. (2015). Caregivers provide more labeling responses to infants' pointing than to infants' object-directed vocalizations. *Journal of Child Language*, *42*(3), 538–561. https://doi.org/10.1017/S0305000914000221

Wu, Z., & Gros-Louis, J. (2017). The Value of Vocalizing: 10-Month-Olds' Vocal Usage Relates to Language Outcomes at 15 Months. *Infancy*, *22*(1), 23–41. https://doi.org/10.1111/infa.12150

Yale, M. E., Messinger, D. S., Cobo-Lewis, A. B., & Delgado, C. F. (2003). The Temporal Coordination of Early Infant Communication. *Developmental Psychology*, *39*(5), 815–824. https://doi.org/10.1037/0012-1649.39.5.815

Zywiczynski, P., Gontier, N., & Wacewicz, S. (2017). The evolution of (proto-)language: Focus on mechanisms. *Language Sciences*, *63*, 1–11. https://doi.org/10.1016/j.langsci.2017.06.004

8.
ANNEXES

Annexe 1 : Formulaire de consentement

UNIVERSITÉ TOULOUSE Jean Jaurès

La communication du bébé avant l'apparition du langage

FORMULAIRE DE CONSENTEMENT

Titre du projet : La communication du bébé avant l'apparition du langage

Chercheur responsable scientifique du projet

Dr. Marie Bourjade, Maitre de conférences en psychologie, Laboratoire Cognition, Langues, Langage et Ergonomie (CLLE), Université Toulouse Jean Jaurès, 5 Allée Antonio Machado, 31058 Toulouse cedex 9. Téléphone : 05.61.50.35.41 ; @ : marie.bourjade@univ-tlse2.fr

Autre chercheur impliqué :
Mawa Dafreville, Doctorante en psychologie du développement, Laboratoire Cognition, Langues, Langage et Ergonomie (CLLE), Université Toulouse Jean Jaurès, 5 Allée Antonio Machado, 31058 Toulouse cedex 9. Téléphone : 06.31.29.75.57 ; @ : mawa.dafreville@univ-tlse2.fr

Lieu de recherche : Au domicile de l'enfant et du parent

But du projet de recherche

Vous êtes invités à participer à une étude ayant pour but de comprendre la communication du bébé lors d'une activité essentielle de sa vie quotidienne : le repas (à savoir le déjeuner, le goûter ou le dîner). Au fur-et-à-mesure que l'enfant grandit, il développe des moyens de communication de plus en plus complexes jusqu'à l'acquisition des premiers mots. Cette étude vise, en somme, à retracer le développement communicatif du bébé de 7 à 18 mois.

Comment se déroule l'étude (méthodologie)

Si vous acceptez de participer à cette étude, votre enfant sera observé à votre domicile, en interaction avec vous au cours du repas. Cette activité sera filmée, de même que vous. Chaque observation durera environ 60 minutes, pendant lesquelles le chercheur aura essentiellement un rôle d'observateur et ne prendra pas part à l'activité vous impliquant, votre enfant et vous. Cette étude impliquera deux observations d'environ une heure repartie sur une période de deux mois. Ces visites se feront au moment le plus approprié pour votre enfant et vous. Chaque observation fera l'objet d'une prise de rendez-vous préalable par téléphone et sera donc fixée à votre convenance.

Vos droits de vous retirer de la recherche en tout temps

Votre participation à cette recherche est basée sur le volontariat. Vous pouvez décider de stopper votre participation et celle de votre enfant à tout moment au cours de l'étude, sans avoir besoin de donner de raison particulière.

Vos droits à la confidentialité et au respect de la vie privée

Afin d'analyser les comportements de votre enfant, un enregistrement vidéo de l'activité observée sera effectué et conservé. Les nom(s) et prénom(s) des participants, ou toute autre information permettant de les identifier, ne seront pas associés à l'enregistrement vidéo, ni à toute autre transcription créée à partir de la vidéo. Les données obtenues seront traitées avec la plus entière confidentialité, et seront conservées dans un endroit sécurisé auquel seuls les chercheurs associés à cette recherche auront accès. Après l'enregistrement, conformément aux dispositions de la loi Informatique et Libertés, vous pourrez exercer vos droits d'accès, et de rectification ou suppression de cet enregistrement auprès du responsable scientifique du projet.

UNIVERSITÉ TOULOUSE Jean Jaurès

La communication du bébé avant l'apparition du langage

Bénéfices

En termes d'avancées scientifiques, les résultats de cette étude permettront de mieux comprendre comment l'enfant réussit son entrée dans le langage, en combinant gestes, vocalisations et mots de manière intentionnelle.

Risques possibles

Afin de limiter le risque que violation de la confidentialité, toute information permettant l'identification de votre enfant est conservée dans un endroit fermé à clé, et les données sont conservées sur un ordinateur protégé par mot de passe.

Diffusion

Les résultats de cette étude seront diffusés de façon strictement anonyme dans des conférences et publiées dans des articles de revues académiques. Toute donnée permettant une identification des participants sera supprimée à la fin du projet. Les enregistrements vidéo ne pourront être reproduits entièrement ou en partie pour des présentations orales ou écrites résultant de cette étude sans votre accord préalable (voir document d'autorisation facultative d'utilisation des droits d'image ci-dessous). Les nom(s) et prénom(s) des participants ne seront pas utilisés dans des présentations orales ou écrites résultant de cette étude.

Vos droits de poser des questions en tout temps

Vous pouvez poser des questions au sujet de la recherche en tout temps en communiquant avec le responsable du projet par courrier électronique à mawa.dafreville@univ-tlse2.fr ou par téléphone au 06 31 29 75 57.

Consentement à la participation

En signant le formulaire de consentement, vous certifiez que vous avez lu et compris les renseignements ci-dessus, qu'on a répondu à vos questions de façon satisfaisante et qu'on vous a avisé que vous étiez libre d'annuler votre consentement ou de vous retirer de cette recherche en tout temps, sans préjudice.

Pôle Toulouse Midi - Pyrénées – Comité d'Ethique pour les Recherches Non Interventionnelles (CERNI)

A remplir par le participant

J'ai lu et compris les renseignements ci-dessus et j'accepte de plein gré de participer à cette recherche.

Nom, Prénom – Date – Signature

Annexe 2 **: Autorisation de droit à l'image**

UNIVERSITÉ TOULOUSE Jean Jaurès

AUTORISATION D'UTILISATION DES DROITS D'IMAGE

AUTORISATION D'UTILISATION DES DROITS D'IMAGE

Centre d'étude toulousain du jeune enfant (*Baby Lab Toulouse*)
Maison des Sciences de l'Homme et de la Société (USR 3414 – CNRS), Université Toulouse Jean Jaurès
5 allées A. Machado, 31058 Toulouse Cedex 9

Au cours de nos études, nous procédons à l'enregistrement vidéo de votre enfant et vous. La procédure standard du *Baby Lab* est de conserver ces vidéos de façon confidentielle et séparée des informations permettant toute identification exceptée l'image, avec un accès limité aux chercheurs ayant signé la charte éthique du *Baby Lab*. Cependant, il peut être utile pour notre travail de diffuser de façon sélective tout ou une partie d'un enregistrement vidéo vers une audience scientifique et/ou publique. Par exemple, il peut être utile de montrer des images d'enfants détendus, non déstabilisés voire amusés par la caméra lors de nos observations à des parents qui se posent la question d'y participer également. Il peut également être utile de montrer à d'autres scientifiques à quoi ressemble notre *Baby Lab*, et illustrer les résultats que nous leur avançons. Pour ces raisons, nous vous demandons la permission, facultative, d'utiliser l'enregistrement de vous et votre enfant. Nous n'utiliserons jamais ces droits à des fins commerciales, et nous n'indiquerons jamais le nom réel de votre enfant avec ces images.

Nom et prénom du parent : _____
Prénom de l'enfant : _____

Autorise le Centre d'études toulousain du jeune enfant (*Baby Lab* Toulouse) à utiliser les enregistrements de mon enfant, effectués lors de notre visite, pour le ou les objectifs suivants (voir informations détaillées en bas de page) :

a) Formation scientifique	Oui / Non
b) Illustration/Démonstration scientifique	Oui / Non
c) Promotion du *Baby Lab* (site internet, lettre d'information etc.)	Oui / Non
Par ailleurs, j'accepte que mon visage apparaisse sur la vidéo	Oui / Non

Parent
Signature: _____ Date: _____

Informations détaillées:

a) **Formation scientifique** : lorsque des membres du personnel, des étudiants et des collègues d'autres universités en visite à la faculté de psychologie de Toulouse Jean Jaurès regardent une vidéo de votre enfant, afin de se former à nos méthodes scientifiques. La formation scientifique est encadrée par un chercheur du *Baby Lab* qui a signé la charte éthique du *Baby Lab*. La ou les personnes formée(s) ne détien(nen)t pas le droit d'utiliser l'enregistrement en dehors de la formation.

b) **Illustration/Démonstration scientifique** : lorsque les enregistrements sont utilisés pour illustrer nos méthodes dans une audience scientifique. Cela inclut la publication d'articles académiques, l'impression d'images sur des posters scientifiques ou l'usage d'enregistrements vidéo lors de présentations orales en conférences scientifiques nationales et internationales, ainsi que lors de visites dans d'autres institutions universitaires. Les droits d'accès aux données brutes seront uniquement transférés à des journaux scientifiquement réputés, pour une utilisation unique, dans un article relatif à l'étude pour laquelle ces enregistrements ont été effectués.

c) **Promotion du *Baby Lab* (site internet, lettre d'information etc.)** : lorsque des enregistrements sont utilisés pour la promotion du *Baby Lab* au grand public. Cela comprend (de façon non exhaustive) l'utilisation pour le site internet du *Baby Lab*, dans les lettres d'informations distribuées aux parents ayant participé à nos études, et pour tout matériel servant au recrutement de nouveaux participants (flyers, posters...) qui seront distribués dans des maternités, centres pédiatriques, crèches etc.

Si vous souhaitez faire remonter une inquiétude ou une réclamation, vous pouvez contacter directement la responsable scientifique du *Baby Lab* Toulouse, Dr Hélène Cochet (mail: helene.cochet@univ-tlse2.fr, tél: 05-61-50-39-70, ou par courrier à l'adresse ci-dessus).

Annexe 3 : Amendement au consentement

UNIVERSITÉ TOULOUSE Jean Jaurès

La communication du bébé avant l'apparition du langage

AMENDEMENT AU CONSENTEMENT
DE L'ETUDE SUR LA COMMUNICATION DU BEBE AVANT L'APPARITION DU LANGAGE

Nous tenons à vous remercier pour votre participation à cette étude sur la communication du bébé. Cette étude a pour but d'étudier la communication de votre enfant avant l'apparition de ses premiers mots. Le bébé, avant apparition concrète du langage, communique avec plusieurs modalités sensorielles (telles que le toucher, la vision, l'ouïe...) c'est ce qu'on appelle la communication multimodale du bébé. Notre hypothèse est que le choix des modalités sensorielles du bébé se fait de façon de plus en plus tactique, au cours de son développement, vis-à-vis du contexte d'interaction et de l'état de son partenaire de communication.

Vous êtes un des principaux partenaires de communication de votre enfant. C'est pourquoi nous vous demandons l'autorisation d'analyser l'interaction entre vous et votre enfant, la façon dont chacun d'entre vous réagit aux signaux de communication de l'autre.

En signant le formulaire de consentement, vous certifiez que vous avez lu et compris les renseignements ci-dessus, qu'on a répondu à vos questions de façon satisfaisante et qu'on vous a avisé que vous étiez libre d'annuler votre consentement.

Pôle Toulouse Midi - Pyrénées – Comité d'Ethique pour les Recherches Non Interventionnelles (CERNI)

J'ai lu et compris les renseignements ci-dessus et je confirme de plein gré notre participation, mon enfant et moi, à cette recherche.

Nom, Prénom – Date – Signature

Annexe 4 : Questionnaire de pré-visite

Questionnaire de pré-visite

Date:	Profession des parents
Nom et prénom de l'enfant :	Mère .....
.....	Père .....
ID :	
Date de naissance :	
Sexe :.....	

Allaitement

Avez-vous allaité votre enfant ?	OUI - NON
Si oui pendant combien de temps ?	.....

Etat de santé de votre enfant

Votre enfant a-t-il eu des antécédents médicaux particuliers ?	OUI - NON
Si oui, lesquels ?	....
Suit-il un traitement médical actuellement ?	....

Configuration familiale

Votre enfant vit-il avec ses deux parents ?	OUI - NON
Outre son père, vit-il avec d'autres adultes que vous ? Qui ? *(ex : grands-parents, oncle/tante, beau-père/mère)*	OUI – NON .....
A-t-il des frères et sœurs ?	OUI - NON
Si oui, de quel(s) âge(s) ?	....

	.....
Outre des frères et sœurs, est-il entouré par d'autres enfants à son domicile? Qui et de quel(s) âge(s) ?	OUI – NON ....

Mode de garde

Quel est le mode ? de garde ?

Centre maternel	Crèche familiale	Halte-garderie itinérante	Micro crèche
Centre multi-accueil	Crèche parentale	Jardin d'enfants	Au domicile parental
Crèche collective	Halte-garderie	Au domicile d' une assistante maternelle	

Autres remarques

.....

.....

.....

.....

.....

.....

Annexe 5 : Questionnaire de visite 1

Questionnaire de Visite 1

Date: .....

Nom et prénom de l'enfant :

.....

ID : .....

Date de naissance : .....

Sexe : .....

Ce questionnaire a pour but de nous renseigner sur les habitudes de votre enfant.

Sommeil

Combien d'heures de sommeil en moyenne dort votre enfant par nuit?

.....

Quelles sont approximativement ses heures de coucher et de lever ?

Lever	Coucher

Combien de siestes fait-il par jour ? Pourriez-vous en préciser l'heure et la durée ?

.....

.....

Combien de temps votre enfant prend-il pour s'endormir ?

.....

Les jours où il prend un peu plus de temps que d'habitude pour s'endormir, que faites-vous ?

.....

.....

.....

Ces derniers temps (cette dernière semaine) avez-vous constaté un changement au vu de vos dernières réponses concernant le sommeil de votre enfant ? Si oui, s'explique-t-il ? (Maladie, présence d'invités...)

.....

.....

Diversification alimentaire

Avez-vous déjà diversifié l'alimentation de votre enfant ? **OUI – NON**

A partir de quand avez-vous commencé la diversification alimentaire ?

.....

Aviez-vous établi un ordre d'introduction des aliments particulier ? **OUI – NON**

Si oui lequel ?

.....

Par quel repas avez-vous commencé ? (le midi, le soir, le goûter)

.....

Rythme alimentaire

Quelles sont approximativement les heures auxquelles votre enfant prend ses repas?

	A domicile	En garde (crèche, assistante maternelle, ...)
Petit-déjeuner		
Déjeuner		
Dîner		

Annexe 6 : Questionnaire de visite 2

Questionnaire de Visite 2

Date: .....

Nom et prénom de l'enfant :

.....

ID : .....

Date de naissance : .....

Sexe : .....

Ce questionnaire a pour but de nous renseigner sur votre ressenti concernant votre participation à cette étude. Vos réponses pourraient nous être d'une aide précieuse puisqu'elles nous permettraient d'appréhender au mieux nos prochaines observations.

Comportement de votre enfant

Par rapport à d'habitude, trouvez-vous que votre enfant interagissait avec vous de la même façon ?

OUI – NON

Si non, qu'avez-vous constaté de différent dans son comportement ?

.....

.....

.....

.....

.....

Votre ressenti

Les observations et/ou les questions posées par l'observateur vous ont-elles gênée ? OUI – NON

Si oui, pourriez-vous l'expliquer ?

.....

.....

.....

.....

.....

Vos impressions concernant cette étude

A quels repas, trouvez-vous que votre enfant communique le plus avec vous ? (Petit-déjeuner, collations, déjeuner, dîner) Sauriez-vous l'expliquer ?

.....

.....

.....

.....

.....

Selon vous, à partir de quel âge votre enfant a-t-il véritablement commencé à communiquer avec vous ? Sauriez-vous l'expliquer?

.....

.....

.....

.....

.....

.....

.....

.....

Quand vous êtes au téléphone, comment votre enfant réagit-il ? Auriez-vous constaté une différence de comportement au fur-et-à-mesure que votre enfant grandissait ?

.....

.....

.....

.....

.....

.....

.....

.....

.....

Annexe 7 : Matériel additionnel de la discussion générale

Chez le jeune chimpanzé

Table S1 : Nombre de gestes produits par les jeunes chimpanzés en fonction de la condition d'attention visuelle maternelle et la tranche d'âge

Modalité perçue	Signaux	Objet	ENFANT			JUVENILE			ADOLESCENT			Total général
			Attentive	Inattentive	Total	Attentive	Inattentive	Total	Attentive	Inattentive	Total	
Audibles			**17**	**13**	**30**	**13**	**17**	**30**	**17**	**17**	**34**	**94**
	Big loud scratch	x	1	1	2	9	12	21	16	6	22	45
	Dangle		10	6	16	1		1				17
	Dangle: with shake			3	3		1	1				4
	Drum	x								1	1	1
	Hit object/ground	x								2	2	2
	Leaf clipping						1	1		1	1	2
	Locomote: gallop						1	1				1
	Object move	x	1		1	3	1	4				5
	Object shake	x		1	1							1
	Stomp 2-feet object/ground	x	4		4							4
	Stomp object/ground	x	1	1	2							2
	Stomping 2-feet object/ground	x		1	1		1	1	1	7	8	10
Tactiles			**8**	**29**	**37**	**11**	**31**	**42**	**4**	**2**	**6**	**85**
	Bite					1	1	2	1		1	3
	Grab		1	5	6							6
	Grab-pull			2	2							2
	Hit other		2	5	7							7
	Hit with object	x		1	1							1
	Hitting other			2	2		1	1				3
	Push		1	3	4		1	1				5
	Push (directed)		1	3	4	1	15	16				20
	Stomp 2-feet other			3	3							3
	Stomp other			2	2							2
	Stomping 2-feet other			1	1							1
	Stroke: mouth					2	5	7	2		2	9

	Tap other					1	1				1
	Touch		1	1	2	6	7	13	1		1	16
	Touch: long other		2	1	3	1		1		2	2	6
Silencieux-visuels			**3**	**7**	**10**	**31**	**23**	**54**	**21**		**21**	**85**
	Beckon					1		1				1
	Head stand			1	1		1	1				2
	Look					1		1				1
	Object in mouth			1	1							1
	Present (directed)		1		1	7	4	11	14		14	26
	Present: genitals					1		1				1
	Raise arm			1	1	6	4	10				11
	Reach		1	1	2	12	3	15	7		7	24
	Reach - palm					2	4	6				6
	Roll over						2	2				2
	Shake arm						1	1				1
	Spin: side roulade						1	1				1
	Spin: somersault			2	2	1		1				3
	Swing		1	1	2		3	3				5
Total général			**28**	**49**	**77**	**55**	**71**	**126**	**42**	**19**	**61**	**264**

Note : ***Pour reconnaître un proto-ajustement chez le chimpanzé****, la méthode était descriptive à partir du tableau de signaux ci-dessus. Avant de s'atteler à identifier les signaux proto-ajustés, il s'agissait au préalable de ne sélectionner que les signaux produits plus de fois que la moyenne pour chaque modalité de signal. Pour l'unimodal, les signaux silencieux-visuels proto-ajustés étaient moins produits en condition d'inattention qu'en condition d'attention visuelle. Pour l'intermodal, les signaux proto-ajustés étaient les signaux audible-ou-tactiles qui n'étaient pas plus produits en condition d'inattention qu'en condition d'attention visuelle.*

Chez l'enfant préverbal

Note : ***Pour reconnaître un proto-ajustement chez l'enfant préverbal****, la méthode était descriptive à partir des graphiques et des valeurs de probabilité (partie 4.2). Pour l'unimodal, les signaux silencieux-visuels proto-ajustés étaient moins produits en condition d'inattention qu'en condition d'attention visuelle. Pour l'intermodal, les signaux proto-ajustés étaient les signaux audibles dont la proportion (par rapport aux signaux silencieux-visuels) étaient grande en condition d'inattention qu'en condition d'attention visuelle ou alors que la répartition des signaux était significativement différente d'une condition visuelle à l'autre (e.g. discrimination du babillage dans la partie 4.2) mais avec un pattern inversé (i.e. différence de proportions entre signaux audible et silencieux-visuels plus grande en condition d'attention qu'en condition d'inattention visuelle).*

Table S2 **Nombre de signaux oraux produits par l'enfant par condition d'attention visuelle maternelle quand la mère est silencieuse**

		Attentive	Inattentive	Total	Part%
[7-10]	signaux oraux	610	316	926	
	bruits de bouche	99	59	158	17%
	babillage	65	14	79	9%
	non-cv	446	243	689	74%
	verbal	0	0	0	0%
[11-14]	signaux oraux	636	374	1010	
	bruits de bouche	138	73	211	21%
	babillage	87	91	178	18%
	non-cv	410	210	620	61%
	verbal	1	0	1	0%
[15-20]	signaux oraux	508	228	736	
	bruits de bouche	67	53	120	16%
	babillage	81	37	118	16%
	non-cv	316	124	440	60%
	verbal	44	14	58	8%
toutes	signaux oraux	1754	918	2672	
	bruits de bouche	304	185	489	18%
	babillage	233	142	375	14%
	non-cv	1172	577	1749	65%
	verbal	45	14	59	2%

Table S3. Valeurs de probabilité des résultats aux tests d'ajustement par signal quand la mère s'adresse à l'enfant (condition avec LAE). Les valeurs grisées sont les résultats non-significatifs et les valeurs soulignées indiquent quand la significativité des signaux *avec* coordination du regard est différente de celle considérant les signaux *avec* et *sans* coordination du regard.

	Avec et *sans* coordination du regard			*Avec* coordination du regard		
	[7-10]	**[11-14]**	**[15-20]**	**[7-10]**	**[11-14]**	**[15-20]**
Ajustement unimodal						
Mouvements corporels	0.006	0.007	0.047	0.012	0.026	<u>0.188</u>
Gestes déictiques	0.023	0.001	0.047	0.047	0.001	0.047
Proximaux	0.188	*0.001*	*0.047*	0.750	*0.001*	*0.094*
Distaux	*0.047*	*0.012*	*0.047*	<u>0.094</u>	*0.003*	*0.047*
Ajustement intermodal						
GA avec objet	0.111	1	1	0.25	1	1
GA sans objet	1	1	1	1	1	1
Bruits de bouche	0.503	1	1	1	0.402	1
Non-CV	1	0.751	1	1	*0.056*	0.812
Babillage	1	0.019	0.573	1	0.026	0.462
Verbal	1	1	1	1	1	0.712

Figure S1. Carte développementale de l'ajustement *unimodal* et *intermodal* des signaux de l'enfant préverbal âgé de 7 à 20 mois quand la mère s'adresse à son enfant. Les gestes *proto-ajustés* sont des signaux qui avant émergence montrent une forme partielle ou inachevée d'ajustement. *(1) le Langage Adressé à l'Enfant (LAE) ; (2) le babillage n'est pas ajusté par la majorité des enfants mais les enfants qui l'ajustent montrent un pattern très marqué (voir figure S3 ci-dessous).*

Figure S3**. Proportion de signaux oraux (par rapport aux signaux silencieux-visuels) par individu, tranche d'âge et condition d'attention visuelle maternelle quand la mère s'adresse à son enfant (p-valeurs : *<0,05, **<0,01, ***<0,001).** Les grands points noirs représentent la proportion

moyenne par individu. La médiane (lignes horizontales), les quartiles (boîtes), les centiles (2,5 % et 97,5 %) et les valeurs aberrantes (points noirs) sont indiqués. Les profils d'utilisation sont comparés d'une condition visuelle maternelle à une autre.

Annexe 8 : Liste des publications et des communications réalisées pendant la thèse

Articles publiés ou en révision dans des revues internationales

Dafreville, M., Hobaiter, C., Guidetti, M., Sillam-Dussès, D., Bourjade, M. Sensitivity to the communicative partner's attentional state: a developmental study in the wild chimpanzee (*Pan troglodytes schweinfurthii*) (2021*). American Journal of Primatology*

Bourjade, M., **Dafreville, M.**, Scola C., Jover M., Six-month-old infant's communication in a comparative perspective: how do maternal attention and interaction matter? (*submitted to Journal of Experimental Child Psychology*)

Dafreville, M., Guidetti, M., Bourjade, M. Attention-sensitive signalling by 7-to-20-month-old infants in a comparative perspective (*submitted to Child Devvelopment*)

Dafreville, M., Guidetti, M., Bourjade, M. Beyond pointing gestures: Attention-sensitive communication in 7-to-20-month-old infants develops pragmatic skills (*submitted to Developmental Psychology*)

Sélection de communications dans des colloques internationaux et nationaux

Dafreville, M., Scola, C., Jover, M., Bourjade, M. (2019, September). *Tactical Use of Multimodal Communication by Six-month-old Human Infants*. Oral presentation at the 6th Protolang Conference, Lisbon, Portugal.

Dafreville, M., Guidetti, M., Bourjade, M. (2022, July). *Sensitivity to the mother's multimodal attention state: A developmental study on mother-infant dyads in naturalistic setting*. Graphic presentation at the International Congress of Infant Studies (ICIS), Ottawa, Canada.

Dafreville, M., Bourjade, M., Guidetti, M. (2020, Novembre). *Communication Multimodale et Attention Sociale : vers une Hypothèse Multimodale des Origines Du Langage Humain*. Communication orale au Laboratoire d'Ethologie Expérimentale et Comparée (LEEC), Paris, France. (invitée)

Dafreville, M., Bourjade, M., Guidetti, M. (2020, Janvier). *Communication Multimodale et Attention Sociale : vers une Hypothèse Multimodale des Origines Du Langage Humain.* Communication affichée à International Society for Gesture Studies (ISGS) – France*,* Paris, France.

Annexe 9 : Enseignements réalisés pendant la thèse

Année	Intitulé du cours	Grade	Lieu	Heures ETD
2021-2022	*Une* introduction à la psychologie du développement (nouvelle PY00201V)	L1	UT2J	48h
2020-2021	Approfondissement et illustrations des conceptions psychologiques de la pensée (PY00203V)	L1	UT2J	12h
2019-2021	Communication & Expression	L2	INSA	36h
2018-2022	Techniques d'observation en SHS (PY00405V)	L2	UT2J	96h
2018-2021	Conceptions psychologiques actuelles de la pensée (PY00201V)	L1	UT2J	42h

Résumé

Communication sensible à l'attention maternelle dans une perspective comparative : approche développementale et multimodale

S'inscrivant dans l'approche *evo-devo*, certaines hypothèses évolutives sur les origines du langage suggèrent que le contrôle volontaire de la voix pourrait être apparu dans la lignée humaine comme un moyen pour le nourrisson de maintenir l'attention et l'investissement du parent. Cette thèse avait pour but d'étudier le développement de la communication du jeune primate à l'aune de ces hypothèses. Pour cela, nous avons quantifié les signaux produits par les « jeunes » humains (N= 30) et chimpanzés (N= 11) envers leur mère selon son état d'attention visuelle (attentive/inattentive). L'observation en condition écologique de la *communication sensible à l'attention* a ensuite été testée à l'aide de descripteurs d'ajustement *unimodal* et *intermodal*, respectivement définis par la capacité du jeune primate à traiter les mères visuellement inattentives (i) en évitant de produire des signaux silencieux-visuels et/ou (ii) en favorisant la production de signaux perceptibles par l'utilisation de signaux audibles et tactiles. Chez le jeune chimpanzé l'ajustement *intermodal* émergeait avant l'*unimodal*, tandis que chez l'enfant préverbal l'ajustement *unimodal* émergeait avant l'*intermodal*. De plus, chez les dyades humaines, le Langage Adressé à l'Enfant (LAE) influençait de manière concurrente ces deux ajustements ; l'ajustement *unimodal* était plus marqué lorsque la mère s'adressait à l'enfant, tandis que l'ajustement *intermodal* était plus marqué lorsque la mère était silencieuse. En testant ces ajustements sur des signaux spécifiques, nos résultats montraient que les signaux *canoniques* du développement du langage (e.g babillage, geste de pointage) n'étaient pas les seuls à s'ajuster à l'attention visuelle maternelle. Enfin, l'ajustement *unimodal* serait prédicteur de compétences lexicales. Tous ces résultats sont discutés au regard des hypothèses *evo-devo* et dans une perspective comparative sur l'évolution de la communication au sein de la lignée primate.

Mots clefs : modalité sensorielle, attention visuelle, dyades mère-enfant, développement, évolution, primates

Abstract

Attention-sensitive signalling in a comparative perspective: a developmental and multimodal approach in mother-infant dyads

From the *evo-devo* perspective, some hypotheses about the origins of language suggest that the volitional control of the voice may have appeared in the human lineage as a means to maintain parental attention and investment. This thesis aimed at expanding our knowledge of the primates' communication development within this theoretical perspective. We quantified the infants' attention-sensitive signaling indexed to maternal visual attention and inattention in mother-infant dyads of humans (N=30) and wild chimpanzees (N=11). The naturalistic observation of attention-sensitive signalling was tested through *unimodal* and *cross-modal* adjustments, respectively defined by the capacity of young primates to address visually inattentive mothers (i) by avoiding silent-visual signals (i.e. communication mismatches) and/or (ii) favouring communication matches through audible-or-contact signals. The young chimpanzees displayed *cross-modal* adjustment before *unimodal* adjustment while the human infants displayed *unimodal* adjustment before *cross-modal* adjustment. Further analysis in human dyads showed that maternal *Infant-directed speech* (IDS) affected both descriptors in concurrent directions; *unimodal* adjustment was steeper and appeared earlier when the mother was talking to the infant while *cross-modal* adjustment was steeper when the mothers were silent. In addition, by testing these adjustments on specific signals, we found that *canonical markers* of language development (e.g. babbling, pointing gestures), but also signals from an extended repertoire were used by infants to adjust to their mother's visual attention. Finally, *unimodal* adjustment may predict subsequent lexical skills. All these results are discussed in line with evo-devo hypotheses and comparative studies on the evolution of communication within the primate lineage.

Keywords: sensory modality, visual attention, mother–infant dyads, development, evolution, primates