



HAL
open science

Aux prémices de nos conversations : une étude comparative chez les nourrissons humains et les primates non-humains

Bastien Meunier

► **To cite this version:**

Bastien Meunier. Aux prémices de nos conversations : une étude comparative chez les nourrissons humains et les primates non-humains. Neurosciences [q-bio.NC]. Université de Rennes, 2023. Français. NNT : 2023URENB059 . tel-04495987

HAL Id: tel-04495987

<https://theses.hal.science/tel-04495987>

Submitted on 8 Mar 2024

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

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

THESE DE DOCTORAT DE

L'UNIVERSITE DE RENNES

ECOLE DOCTORALE N° 637
Sciences de la Vie et de la Santé
Spécialité : *Neurosciences, Éthologie*

Par

Bastien MEUNIER

Aux prémices de nos conversations : une étude comparative chez les nourrissons humains et les primates non-humains

Thèse présentée et soutenue à Rennes, le 01 décembre 2023
Unité de recherche : UMR 6552

Rapporteurs avant soutenance :

Jean-Pierre Chevrot Professeur émérite ; Université de Grenoble Alpes, EA 609 LIDILEM
Judith M. Burkart Professeure ; University of Zurich, IEA

Composition du Jury :

Présidente : Maya Gratier Professeure ; Université Paris Nanterre, UR 3456 LECD

Examineurs : Jean-Pierre Chevrot Professeur émérite ; Université de Grenoble Alpes, EA 609 LIDILEM
 Judith M. Burkart Professeure ; University of Zurich, IEA
 Maya Gratier Professeure ; Université Paris Nanterre, UR 3456 LECD
 Adrien Meguerditchian Chargé de recherche ; CNRS, UMR 7290 LPC

Directeur de thèse : Alban Lemasson Professeur ; Université de Rennes, UMR 6552 EthoS

Co-encadrante de thèse : Virginie Durier Chargée de recherche ; CNRS, UMR 6552 EthoS

Remerciements

Cette thèse est le résultat de trois années de dur labeur, au cours desquelles j'ai pu rencontrer tant de gens formidables, qu'il m'est difficile de les remercier en quelques lignes. Il est tard, pardonnez-moi si j'oublie quelqu'un.

A Virginie Durier et à Alban Lemasson, pour m'avoir fait confiance dans ce projet, et m'avoir guidé comme vous l'avez fait. Bien sûr, il est impossible de passer en revue tous les moments passés ensemble, et toutes les choses que vous m'avez apprises. Je me rappellerai de ce super duo que vous avez formé pour m'encadrer au mieux. Mille mercis !

A Judith Burkart et Jean-Pierre Chevrot, d'avoir bien voulu être les rapporteurs de cette thèse, et à Maya Gratier et Adrien Meguerditchian pour avoir accepté de siéger dans mon jury.

A Maéva Thomazeau et Justine Thobois, les étudiantes que j'ai eu le plaisir de co-encadrer et qui m'ont apporté une aide précieuse. Rien n'aurait été possible sans ces heures que vous avez passées à suivre le regard de bébés pas toujours réveillés. Chapeau bas pour le travail monstre que vous avez abattu toutes les deux.

A Arnaud Rossard pour m'avoir tant aidé lors de ces repasses effectuées sur les mangabeys !

A Marion Charrier, ce rayon de soleil qui chaque jour illumina le bureau. Tu m'as été d'une telle aide au quotidien qu'il est illusoire d'espérer un jour rembourser les dettes que je te dois. On peut dire que sans ta présence une nuit d'Août à Bielefeld, cette thèse n'aurait pas eu la même allure.

A Antoine l'Azou, cet irremplaçable esthète, avec qui nous avons eu des conversations mémorables, allant de la politique chinoise de l'enfant unique à la poésie iranienne. Merci d'être là pour nous ouvrir à d'autres horizons, même si je ne regarderai probablement jamais de films ouzbeks, il ne faut pas abuser, désolé.

A Francis Mosinski, mon compagnon de route depuis le master, avec qui nous avons rédigé les meilleurs rapports de Travaux Pratiques de tous le far-west. Tout ce qui est technique est toi, et une fois n'est pas coutume, tu m'as sauvé la mise dans l'élaboration de cette thèse. Merci à toi l'ami !

A Lucie Rigaille, ma marraine primatologue, qui est d'un soutien sans faille et d'une gentillesse inégalée. Je suis tellement content que tu sois arrivée là où tu es désormais ! Néanmoins, je ne te remercie pas de confirmer par ton sujet d'étude les blagues de mon père sur la similitude acoustique entre éthologue et proctologue...

A Manon Toutain et Léa Guéguen, je pars devant en éclaireur, ne tardez pas ! Je vous souhaite le meilleur vous qui participez à faire rayonner cette promo CAH si mal comprise !

A ma promo CAH ! Céline, Pierre, Thomas, Salomé, Alexa-Lou, Léo, Julien, Erell, et tant d'autres avec qui j'ai passé de merveilleux instants à Rennes. Je ne vous oublierai jamais !

A toute ma famille pour m'avoir soutenu jusqu'au bout dans ces études interminables dont on ne sort probablement jamais. Merci infiniment Papa et Maman, Julien et Manon pour avoir été un soutien pendant les moments difficiles. Merci Mamie pour toutes ses pensées qui me sont parvenues à un moment où j'en avais tant besoin.

A Caro, qui est l'as de trèfle qui pique mon cœur. Merci pour tout ton soutien. Tu as été d'une si précieuse aide qu'il serait vain de le préciser. Tu le sais !

A toutes les familles qui ont participé à nos études, passé du temps avec des chercheurs dans les nuages sans rien attendre en retour. Cette thèse n'aurait pas été possible sans vous.

A tous les mangabeys qui ont bien voulu se laisser observer et ne pas se faire berner par les repasses que je me suis échiné à leur repasser.

Ah, on m'indique dans l'oreillette qu'il ne me reste plus que cinq minutes pour envoyer cette thèse. Je remercie tous ceux que j'ai oublié, tout le personnel d'EthoS pour les bons moments passés ensemble, tous mes anciens profs qui sont devenus mes collègues. Merci aussi à....*plus de temps*

A mon papy, à qui je dédie cette thèse

« Il faut voir aussi qu'une conversation a sa vie et ses exigences propres. C'est un petit système social qui tend à préserver ses frontières ; c'est un îlot de dépendance et de loyauté, avec ses héros et ses traîtres »

Erving Goffman, *Les rites d'interaction* (1982)

Table des matières

Chapitre I : Introduction	1
1. Les conversations en tant que programme de recherche	3
1.1. La face immergée du langage	3
1.2. Une forme pour une multitude de conversations	4
1.3. Les contraintes du tour de parole sur le langage	7
2. Du tour de rôle au tour de parole : ontogénèse des conversations	10
2.1. Pouvoir converser avant même de savoir parler	10
2.2. Le rôle des proto-conversations dans l'apprentissage du langage	12
2.3. L'impact de l'environnement social	14
3. Des précurseurs chez les autres primates : phylogénèse des conversations	18
3.1. Etudier les origines du langage grâce aux primates non-humains	18
3.2. Des conversations chez les primates non-humains ?	21
3.3. Des proto-conversations au cœur de l'organisation sociale	24
3.4. Des règles apprises socialement ?	25
4. Problématique et objectifs de la recherche	26
4.1. Problématique	26
4.2. Objectifs communs	27
4.3. Objectifs et méthodes des études sur les nourrissons humains	27
4.4. Objectifs et méthodes des études sur les mangabeys à collier (<i>Cercocebus torquatus</i>)	28
Chapitre II : Proto-conversations chez les nourrissons humains	31
Article 1 : Of the importance to reconsider individual variability in infant studies: family traits do impact conversational skills in 6-month-olds	33
Article 2 : Often overlooked, experience at daycare influences the perception of conversational turn-taking in pre-verbal infants.	53
Article 3 : Cross- and same-sex analysis of parent-infant protoconversations: fathers also matters	67
Chapitre III : Proto-conversations chez les mangabeys à collier (<i>Cercocebus torquatus</i>)	83
Article 4 : Social factors drive vocal exchanges in red-capped mangabeys (<i>Cercocebus torquatus</i>)	85
Article 5 : Turn-taking in red-capped mangabey (<i>Cercocebus torquatus</i>): an experimental investigation of a conversation rule in a non-human primate species	99
Chapitre IV : Discussion	115
1. Analyse des proto-conversations	117
1.1. Chez les nourrissons humains	117
1.2. Chez les mangabeys à collier	118
2. Une structure temporelle similaire	119

3.	De la normalité à la normativité : mettre en évidence des normes	121
3.1	Qu'est-ce qu'une norme ?	121
3.2	Des normes conversationnelles chez les nourrissons humains ?.....	121
3.3	Des normes conversationnelles chez les primates non-humains ?	122
4	Homologies et convergences, un puzzle évolutif à reconstruire	124
4.1	La pro-socialité qui motive notre moteur interactionnel	124
4.2	Les relations mère-nourrisson comme « îlots de confiance » temporaires.....	126
4.3	L'hypothèse de l'élevage coopératif.....	128
4.4	Retour aux exigences d'une vie sociale complexe	130
5	Perspectives	130

Chapitre I :

Introduction

1. Les conversations en tant que programme de recherche

1.1. La face immergée du langage

Parler de conversations plutôt que de langage peut interroger, habitués que nous sommes à voir dans ce dernier la compétence cardinale de l'exceptionnalité humaine. Les conversations sont en ce sens vues comme le résultat de la pratique du langage, et notre intérêt pour elles est circonstancié à l'étude de cette capacité. Pourtant, une réflexion amorcée par des philosophes du langage et continuée par tout un pan de la linguistique montre à quel point le langage humain n'a du sens qu'à l'intérieur de conversations.

Deux philosophes du langage de la seconde moitié du XXe siècle ont permis ce développement : John Longshaw Austin (1911-1960) et Paul Grice (1913-1988). Dans *Quand dire c'est faire* (1962, conférences publiées à titre posthume), Austin a mis en évidence que dans notre langage ordinaire, nous cherchions moins à énoncer des vérités en décrivant un état du monde qu'à accomplir une action comme demander, avertir, inviter, promettre, excuser, etc. Austin distingue en fait ce qui est dit littéralement dans l'énoncé (l'acte locutoire), ce qui est sous-entendu par le locuteur et qui est la fonction de l'énoncé (l'acte illocutoire), et ce qui est interprété par les interlocuteurs et est donc le résultat de l'interaction (l'acte perlocutoire). Le sens littéral des énoncés est en effet toujours ambivalent, tant et si bien que nous nous servons constamment d'éléments contextuels pour comprendre ce qui est dit. Par exemple, lorsqu'un locuteur dit « est-ce qu'il y a du sel ? » (exemple repris de Amphaeris et al., 2023), il peut sincèrement s'interroger sur l'existence du sel dans le monde (c'est l'acte locutoire). Pourtant, dit à table, le locuteur veut dire qu'il cherche du sel sur sa table pour en mettre dans son plat (c'est l'acte illocutoire). Les autres personnes ne s'y trompent pas et lui passent généralement directement le sel, chose qui n'est pourtant pas explicitement demandée ici (c'est l'acte perlocutoire). Ainsi, ce qui est accompli ici l'est grâce à un ensemble de pratiques sociales et de processus cognitifs qui dépassent largement ce qui est littéralement dit. L'ensemble des éléments contextuels permettant de préciser le sens d'une locution est l'objet d'étude de la pragmatique, un champ important de la linguistique actuelle.

Grice a, quant à lui, une approche différente bien que complémentaire. Grice (1957) explique que le langage n'est pas simplement l'association entre un signifiant et un signifié, entre un mot et une chose, mais bien l'échange d'intentions de manière ostensible. Il distingue en effet ce qu'il appelle le sens naturel qui est cette simple association entre deux choses différentes et ne nécessite en fait pas d'intentions (la fumée signifie qu'il y a un feu, mais le feu n'a pas eu

l'intention de communiquer, il s'agit juste d'une relation entre deux phénomènes), et le sens non-naturel ou conventionnel qui est à l'œuvre dans le langage humain. Dans ce dernier, non seulement nous communiquons une intention, mais en plus nous manifestons notre intention de vouloir communiquer cette intention, qui s'illustre par la simple utilisation de la parole : en parlant, les autres savent que nous cherchons ouvertement à communiquer et nous savons nous-mêmes que les autres vont essayer de comprendre ce que nous voulons. Le langage n'est donc pas un code, nous ne décryptons pas une suite de symboles car aucune de ces suites ne seraient suffisamment complètes pour enlever toute ambiguïté (cf. l'exemple « est-ce qu'il y a du sel ? » donné plus haut qui peut donner lieu à une multitude d'interprétations). Lorsque nous parlons, nous faisons des inférences pour comprendre les intentions d'autrui, et le sens de chaque phrase dépend de son contexte d'émission, notamment son contexte conversationnel (Grice, 1975). Le sens n'est pas strictement contenu dans les mots qu'on emploie mais émerge de cette dynamique conjointe des interlocuteurs et qui fait que toute conversation est un effort coopératif selon Grice (1975). Il postulera que les conversations humaines impliquent un principe coopératif (les interlocuteurs veulent se comprendre), qui implique lui-même le respect de plusieurs sous-principes : un principe de quantité obligeant les locuteurs à donner ni trop ni pas assez d'informations, un principe de qualité obligeant les locuteurs à dire ce qu'ils croient vrai, un principe de relation obligeant les locuteurs à dire quelque chose dans le thème de la conversation (appelé aussi principe de pertinence, que certains auteurs considèrent comme le principe central regroupant tous les autres ; voir Wilson and Sperber, 1979), et enfin un principe de manière obligeant les locuteurs à éviter les ambiguïtés, à être clairs et ordonnés dans leurs interventions (Grice, 1975).

Ainsi, il apparaît que les conversations ne sont pas le résultat trivial de cette capacité complexe qu'est le langage, mais plutôt que le langage émerge de notre capacité à interagir avec les autres lors de conversations. En ce sens, converser est un des comportements les plus remarquables de notre espèce, qui doit être étudié pour lui-même.

1.2. Une forme pour une multitude de conversations

Si les êtres humains produisent en moyenne 16 000 mots par jours (sans différence significative entre les hommes et les femmes ; Mehl et al., 2007), ces derniers ne sont pas uniformément répartis dans la journée. Ils sont en effet regroupés en petites salves d'une durée en général inférieure à deux secondes (Levinson and Torreira, 2015) en alternance avec celles des autres

pendant un total d'environ deux heures par jour (estimations faites par Levinson, 2019). Durant ces interactions, les êtres humains s'orientent en général face à leur interlocuteur, même lorsqu'ils sont dans un groupe. Bien que le langage puisse être présent en dehors de ces interactions face-à-face (on peut penser aux écrits ou aux émissions radio qui sont des productions langagières), il se niche principalement dans ces dernières qui représentent l'utilisation majoritaire du langage chez tous les êtres humains. La conversation est donc naturellement devenue l'objet d'étude de chercheurs voulant comprendre les sociétés humaines.

Ces conversations ont d'abord été étudiées par un courant de la sociologie nommé interactionnisme symbolique, dont le représentant le plus connu est probablement Erving Goffman (1922-1982). Ce dernier s'est en effet intéressé aux interactions face-à-face entre individus, arguant qu'il se jouait à cette échelle des phénomènes cruciaux non présents chez les individus eux-mêmes ou invisibles à l'échelle macro-sociologique. Il étudiait notamment comment les individus cherchaient à se présenter sous leur meilleur jour dans le champ social en jouant un rôle comme dans les pièces de théâtre (Bonicco, 2007). Un autre courant représenté par Harold Garfinkel est celui de l'ethnométhodologie, ce dernier étudiant les humains au sein de petits groupes à la manière d'un ethnologue pour mettre en évidence comment ils particularisent leur langage, règles et normes à des circonstances locales (Goodwin and Heritage, 1990). Mais c'est surtout un article fondateur de sociologues comme Harvey Sacks, Emanuel Schegloff, et Gail Jefferson (1974) qui a lancé ce champ d'étude nommé Analyse des Conversations (« Conversation Analysis », CA). Prolongeant les travaux de leurs mentors (puisque Sacks et Schegloff ont été les étudiants de Goffman et en contact régulier avec Garfinkel), ces derniers ont observé que les interlocuteurs échangent en effet spontanément des tours de parole dont la longueur et le contenu ne sont pas spécifiés à l'avance, en évitant de parler en même temps (les chevauchements sont rares) ou de laisser de trop longues pauses entre deux locutions. A partir de ces observations, ils conçoivent un modèle pour décrire l'organisation en tour de parole (« turn-taking ») qu'ils observent. Une contribution à une conversation est qualifiée de tour et est composée d'un nombre variable d'unités (« turn construction units », TCUs) à l'issue desquelles se situe un point où le tour peut se terminer et un autre locuteur peut initier un nouveau tour. Ce point est appelé place pertinente de transition (« transition relevance place », TRP). Un tour de parole est aussi composé de techniques d'allocation de tours (« turn allocation techniques ») qui peuvent être de deux types : celles où le locuteur actuel sélectionne le locuteur suivant (en désignant spécifiquement une personne, ou bien en instaurant une règle explicite comme un tour de table lors d'une réunion) et celles

où le locuteur suivant doit s'auto-sélectionner (en évitant de chevaucher ou de laisser une pause s'installer). Lorsqu'un autre locuteur a pris la parole, le locuteur initial est tenu de se comporter comme n'importe quel autre interlocuteur attendant de pouvoir intervenir au moment opportun ou qu'on le désigne. Il est important de noter que dans les cas où le locuteur désigne son successeur, ce dernier a une obligation d'intervenir dans la conversation et doit se justifier s'il ne veut pas le faire, ce qui n'est pas le cas lors d'une conversation où chacun peut intervenir spontanément. Dans ce dernier cas, si personne ne s'auto-sélectionne lors d'une TRP, le locuteur conserve la parole jusqu'à ce que quelqu'un intervienne.

Enfin, lorsque le tour de parole dysfonctionne, les interlocuteurs disposent de toute une palette de techniques de réparation, permettant de le réinstaurer. Cela peut être le simple fait de s'arrêter de parler lorsque quelqu'un nous a interrompu pour éviter qu'il y ait plusieurs personnes parlant en même temps, ou bien de commencer des bouts de phrases pendant que quelqu'un est en train de parler pour lui signifier qu'on veut intervenir et qu'il a assez parlé (Sacks et al., 1974). Toutefois, si ces interruptions forcées permettent de rétablir le tour de parole, elles ont un effet sur l'image des individus qui sont impliqués. Les individus qui coupent la parole des autres sont perçus comme ayant une personnalité plus négative et comme étant moins sociables que ceux qui ne coupent pas la parole ou qui se font couper la parole (LaFrance, 1992; Robinson and Reis, 1989a; ter Maat et al., 2010). Les individus qui interrompent régulièrement sont ainsi perçus comme plus dominants et plus influents, mais cette amélioration du statut les rend moins appréciés des autres et a donc un coût (Farley, 2008). Mais la perception de ces coupures peut dépendre d'autres facteurs, tels que le genre. En effet, les hommes interprètent les coupures de parole de manière plus positives que les femmes : ils sont plus enclins à les considérer comme des signes de respect, d'accord et de partage d'opinions, tandis qu'à l'inverse les femmes les considèrent plus comme un manque d'accord et de respect (Chambliss and Feeny, 1992).

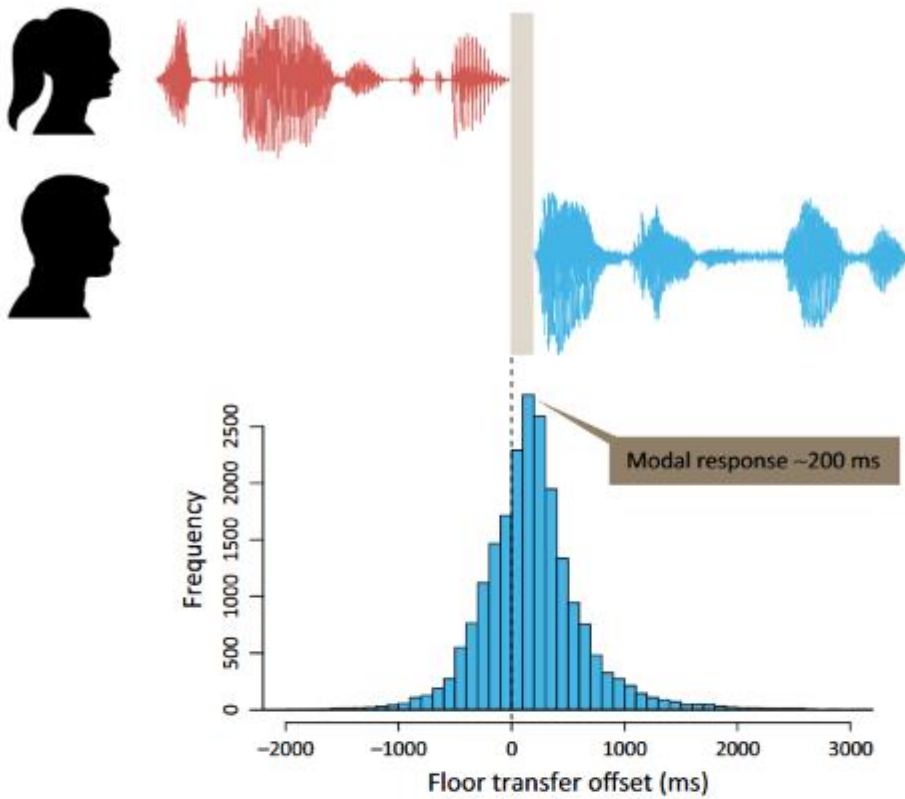
Tous les chevauchements ne conduisent pas à ces mécanismes de réparation. Des études ont montré que les chevauchements sont en fait assez nombreux (Heldner and Edlund, 2010), et la plupart ne sont soit pas perçus du tout car ils sont trop courts (Heldner, 2011), soit ils ne sont pas perçus négativement. Schegloff (2000) a par exemple montré qu'il existait au moins quatre types de chevauchement qui n'étaient pas problématiques. Il y a tout d'abord les chevauchements terminaux (« terminal overlaps ») qui adviennent lorsqu'un locuteur chevauche légèrement la parole d'un locuteur vraisemblablement en train de finir son tour. Il y a ensuite les chevauchements continueurs (« continuers ») qui signifie que le locuteur comprend et/ou approuve ce que le locuteur actuel est en train de dire (les « *hm* » souvent

accompagnés de hochements de têtes). Il y a aussi les accès conditionnels au tour (« conditional access to the turn ») qui vont surtout concerner les contributions complétant un tour actuel, comme lorsqu'un locuteur donne le mot que le locuteur actuel est en train de chercher, ou précise ce qu'il veut dire prédisant un quiproquo. Enfin, il y a les chevauchements choraux (« chordal ou choral ») où les interlocuteurs sont implicitement invités à chevaucher comme c'est le cas lors des rires, des salutations collectives, les départs ou les félicitations. Seuls les chevauchements qui ne rentrent pas dans ces catégories sont considérés comme compétitifs et problématiques pour la conversation (Schegloff, 2000).

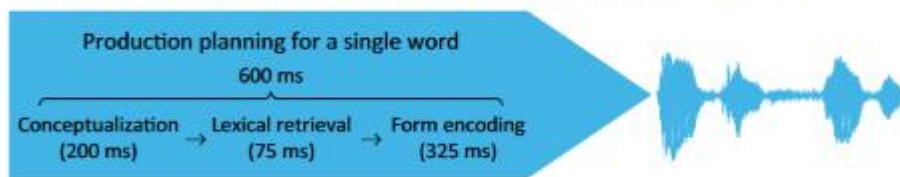
1.3. Les contraintes du tour de parole sur le langage

L'évitement de certains chevauchements et la minimisation des silences qui sont les principales règles caractérisant le tour de parole imposent des contraintes pour le traitement cognitif du langage. En effet, la durée mesurée des intervalles entre les tours est autour de 200ms (Heldner and Edlund, 2010; Levinson and Torreira, 2015), alors que le temps de production ne serait-ce que d'un mot est déjà plus long. Planifier la production d'un mot implique une succession de plusieurs étapes cognitives dont les études de psycholinguistiques ont pu évaluer la durée : cela commence par une phase de conceptualisation d'environ 175ms, puis une phase de sélection lexicale de 75ms, et enfin une phase d'encodage de la forme de 350ms (Indefrey and Levelt, 2004). Ainsi, la planification d'un seul mot prend au moins 600ms, et cette durée de planification s'accroît bien sûr si l'on veut produire une combinaison de mots, ou une phrase. Cette durée excède donc largement la moyenne des intervalles observés entre deux tours. Cela signifie que nous planifions notre tour de parole pendant celui qui précède, ce qui implique que nous saisissons déjà l'intention de notre interlocuteur au milieu de son intervention (qui est en général d'environ deux secondes), et que nous prédisons la fin de cette dernière afin de coordonner notre production verbale (Levinson, 2016; Levinson and Torreira, 2015) (fig.1). Cette prédiction de la fin d'un tour est réalisée grâce à des indices lexicaux, syntaxiques et prosodiques. Par exemple, la présence d'une interjection « et » (indice lexical) et la diminution de l'intensité du ton (indice prosodique) agissent comme des signaux permettant aux locuteurs de commencer leur production plus tôt que s'ils n'étaient pas présents (Barthel et al., 2017). C'est la combinaison d'indices lexico-syntaxiques et prosodiques qui permet une bonne prédiction de la fin du tour, les indices lexico-syntaxiques seuls ne permettant pas cette prédiction, de même que la prosodie, chez les enfants (entre 3 et 5 ans) et chez les adultes (Casillas and Frank, 2017). Nous réalisons donc la prouesse de saisir l'intention de notre

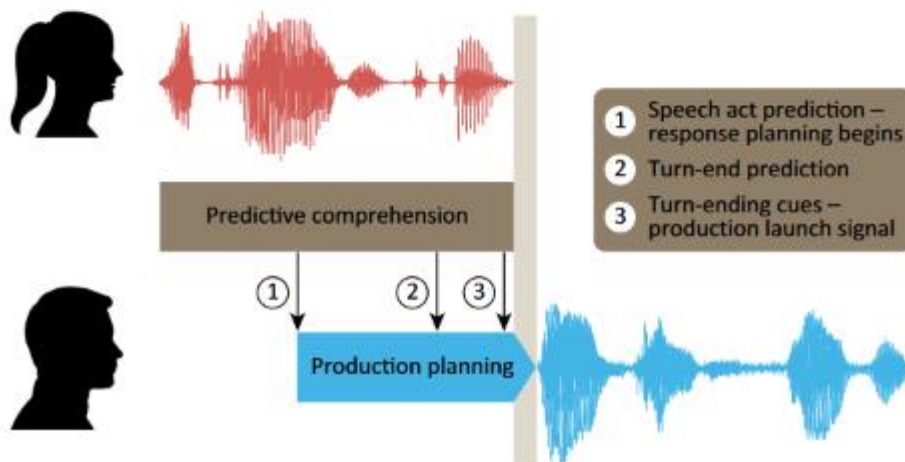
(A) Responses in conversation are fast



(B) Latencies in production are threefold or more longer than the modal gap



(C) Production of response must therefore overlap with comprehension of the incoming turn



Trends in Cognitive Sciences

Figure 1 : schéma de synthèse de la contrainte qu'exerce le tour de parole sur le traitement cognitif du langage, repris de l'article de Levinson (2016).

interlocuteur pendant son tour, de prédire sa fin afin de prendre la parole très rapidement après chaque tour de nombreuses fois, sans effort et sans même nous en rendre compte.

Le tour de parole étant universel, l'évitement des chevauchements et la minimisation des silences se retrouvent dans toutes les sociétés humaines et contraignent de la même manière toutes les langues (Levinson, 2016; Stivers et al., 2009). Ainsi, des langues très différentes dans leurs structures ont des délais de réponse similaires. Dans une étude évaluant le délai de réponse de locuteurs de 10 langues très différentes, le délai moyen était de 208ms, avec une variation allant de 7ms pour le japonais à 469ms pour le danois (Stivers et al., 2009). Malgré cette variation donc, les délais de réponse restent inférieurs au délai nécessaire pour produire une réponse aussi simple soit-elle. En outre, la variation observée ne s'expliquait pas par les structures des langues considérées, puisque les langues où le marqueur de question (ce qui indique que la phrase est une question et appelle une réponse) était en fin de phrase pouvait avoir des délais de réponses bien plus courts que supposé : avoir des marqueurs en fin de phrases oblige le locuteur à attendre la fin de la phrase pour savoir qu'il s'agit d'une question, or c'est le cas du japonais qui a les délais les plus courts (et inversement avec le danois, qui a des marqueurs en début de phrase mais des délais long). Cette variation ne peut donc s'expliquer que culturellement, même si elle est faible puisqu'un délai moyen de 200ms est considéré comme acceptable même par les cultures aux délais de réponses les plus extrêmes (japonais et danois ; Stivers et al., 2009).

L'interrogation qui nous vient est de savoir pourquoi les conversations ont une forme aussi restrictive ? Ne serait-ce pas plus simple si les délais de réponses étaient plus longs, nous évitant de dépenser autant de ressources cognitives à prédire l'acte de langage au beau milieu d'un tour afin de commencer la planification du nôtre ? Notre communication n'en serait-elle plus efficace ? Une hypothèse fonctionnaliste intéressante pourrait être qu'un locuteur produisant des tours courts et exigeant une réponse rapide des autres peut s'assurer que son propos est bien compris et préciser immédiatement dans le cas contraire. Levinson (2006, 2016, 2019) apporte une réponse différente à ces questionnements, via son hypothèse du moteur interactionnel (*Interaction Engine Hypothesis*). Sur la base d'un faisceau d'arguments, cette théorie postule « l'existence d'une capacité d'interaction distincte qui sous-tend notre comportement communicatif et notre langage en particulier ». Cette capacité serait en fait l'assemblage de plusieurs éléments, avec différentes origines évolutives et différents schémas développementaux, et précéderait le langage d'un point de vue phylogénétique et ontogénétique. Ainsi, le langage devrait se compresser à l'intérieur d'une niche qui n'est pas

forcément adaptée pour le recevoir. Etudier le langage ne peut se faire sans l'étude de ce moteur interactionnel, dont on doit reconstituer le puzzle en investiguant sa mise en place chez les nourrissons (ontogénèse) avant de s'interroger sur sa mise en place au cours de l'évolution des primates (phylogénèse).

2. Du tour de rôle au tour de parole : ontogénèse des conversations

2.1. Pouvoir converser avant même de savoir parler

Un des principaux arguments du moteur interactionnel est que les nourrissons interagissent avec leurs parents bien avant de maîtriser le langage. Cela se manifeste dès les premiers jours dans leurs capacités d'imitation : les nouveau-nés sont capables d'imiter les expressions faciales et les gestes des adultes (Meltzoff and Moore, 1977), et même les sons « voyelles » entendus lorsqu'ils ont passé le 1^{er} mois (Kuhl and Meltzoff, 1996). En fait, les nourrissons sont capables de produire une diversité de vocalisations, certaines résultant d'émotions intenses comme les pleurs et les rires, et d'autres émises de manière beaucoup plus flexibles (Oller et al., 2013). Ces dernières sont nommées proto-phones car nécessitant une phonation (avec ou sans articulation) sans toutefois comporter des voyelles ou consonnes clairement identifiables (à la différence du babillage). Les proto-phones regroupent ainsi les cris ou couinements (*squeals*), les sons ressemblant à des voyelles (*vowels-like sounds*), et les grognements (*growls*). Oller et al (2013) ont pu montrer que les affects faciaux des nourrissons étaient majoritairement neutres lorsqu'ils émettaient des proto-phones, mais pouvaient aussi être associés à des affects positifs ou négatifs, contrairement aux pleurs ou aux rires qui étaient uniquement associés à des affects faciaux hautement négatifs ou positifs respectivement. Ce sont ces vocalisations que les nourrissons vont utiliser lors des interactions en face-à-face avec leurs parents, qui combinent des sourires, des contacts et surtout des paroles (Bloom, 1975; Bloom et al., 1987). Les parents vont d'ailleurs bien faire la distinction entre ces proto-phones et les autres vocalisations, en répondant plutôt vocalement aux premières (Gros-Louis et al., 2006) dans des tours de paroles avec un évitement des chevauchements montrant par là qu'ils perçoivent bien ces proto-phones comme des précurseurs du langage (Yoo et al., 2018). Les premiers chercheurs s'intéressant à ces interactions ont d'ailleurs fait explicitement le lien avec les conversations des adultes humains (Bloom, 1975; Ginsburg and Kilbourne, 1988; Snow, 1977; Stern et al., 1975).

Stern et al. (1975) ont ainsi repéré deux modes de communication lors d'interactions entre des mères et leurs nourrissons de 3 à 4 mois : un mode où les deux vocalisaient à l'unisson, et un

autre mode où les deux vocalisaient de manière alternée, le premier étant deux fois plus fréquent que le second. Des études plus récentes ont montré que ce pattern alterné apparaissait beaucoup plus précocement puisqu'on l'observe chez des nouveau-nés à terme et des prématurés bien qu'il n'est là encore pas le mode d'interaction majoritaire (Dominguez et al., 2016; Caskey et al., 2014). Dès le 2^{ème} mois au moins, les nourrissons ont par ailleurs un rôle actif dans ces proto-conversations puisqu'ils sont capables d'initier des tours et d'anticiper la fin de ceux de leurs mères (Gratier et al., 2015). Tout ceci suggère que les nourrissons sont comme préparés à interagir de cette manière, avant qu'interviennent des processus de maturation et d'apprentissage conduisant à des évolutions observables. En effet, au premier mois, ils ne font pas la différence entre des séquences où la mère interagit de manière contingente, des séquences où elle interagit de manière non-contingente, et des séquences où elle imite les expressions faciales de son nourrisson (Striano et al., 2005). Or face aux mêmes interactions à 3 mois, les nourrissons manifestent une sensibilité à la contingence puisqu'ils réagissent plus positivement aux séquences contingentes qu'aux deux autres types. (Striano et al., 2005). Cette sensibilité croissante à la contingence s'illustre aussi par la rapidité avec laquelle le mode d'interaction alterné devient prédominant par rapport au mode d'interaction à l'unisson, entre le 3^{ème} et le 4^{ème} mois (Ginsburg and Kilbourne, 1988). Une étude longitudinale plus récente, ayant observé 12 dyades mères-enfants à 3, 4, 5, 9, 12, et 18 mois, a par ailleurs montré que le nombre de chevauchements diminuait de façon constante, suggérant là encore une maturation et/ou un apprentissage (Hilbrink et al., 2015).

Plus important encore, cette étude a révélé une évolution non-linéaire des délais de réponses des nourrissons : les délais de réponses qui sont aux alentours de 600ms à 5 mois augmentent significativement pour atteindre un pic d'environ 1200 ms à 9 mois. Ce n'est qu'à 18 mois que les enfants retrouvent des délais de réponses équivalents à ceux qu'ils avaient à 5 mois (Hilbrink et al., 2015). Les auteurs s'attendaient bien à un allongement du délai de réponse causé par la difficulté pour un enfant d'utiliser le langage au sein des tours de paroles (il s'agissait d'une des prédictions de l'hypothèse du moteur interactionnel, Levinson figurant parmi les auteurs de cet article), mais dans ce cas-là cet allongement serait arrivé plus tard, à partir du 12^{ème} mois, lorsque l'enfant commence à combiner plusieurs mots ensemble. L'hypothèse la plus simple pouvant expliquer la précocité de ce phénomène est que les nourrissons de cet âge auraient acquis une compréhension du tour de parole et attendraient la fin du tour de leur mère avant de vocaliser pour éviter les chevauchements. Une autre hypothèse moins parcimonieuse serait que les nourrissons de cet âge comprendraient « la fonction communicative et de partage d'une

relation sociale », causant cet allongement de délai. Quoiqu'il en soit, si cet allongement des délais n'est pas causé par le développement du langage en tant que tel, ce dernier peut devenir ultérieurement un facteur aggravant le phénomène. Les enfants semblent en effet lutter pour répondre de manière verbale aussi vite que les adultes dans les conversations jusqu'à 4 ans, âge en dessous duquel on observe des délais plus long (Casillas and Frank, 2017). La question de l'intégration du langage dans les tours de rôle déjà en place au cours de l'ontogénèse reste donc en suspens.

2.2. Le rôle des proto-conversations dans l'apprentissage du langage

Les nourrissons ont toujours fasciné les chercheurs par la facilité avec laquelle ils parviennent à apprendre une langue à partir de rien. Ce processus implique une succession d'étapes importantes qui se déroulent dès la première année, où le nourrisson doit apprendre à utiliser simultanément des compétences perceptuelles, computationnelles et sociales (Kuhl, 2004). Surtout, ce processus semble très robuste, puisque la quasi-totalité des êtres humains franchissent ces différentes étapes et parviennent à maîtriser le langage, malgré la grande diversité des modes de vies.

Pour ce faire, les nourrissons disposent durant la première moitié de leur première année d'une capacité qui peut paraître comme un super-pouvoir aux yeux des adultes : ils peuvent distinguer tous les phonèmes de toutes les langues, même celles auxquelles ils n'ont pas été exposés (Werker and Tees, 1984). Cette perception universelle décline à mesure de l'exposition à une langue particulière, phénomène nommé restriction perceptuelle, qui est en fait une spécialisation de l'enfant pour sa langue maternelle. La restriction de la perception aux sons spécifiques à la langue maternelle commencent à partir du 6^{ème} mois tant et si bien que les nourrissons de 10-12 mois performant significativement moins bien que ceux de 6-8 mois lors d'expériences de discrimination de sons non-spécifiques à la langue maternelle (Werker and Tees, 1984). Un double apprentissage statistique se met en place (Kuhl, 2004). Les nourrissons catégorisent d'abord les fréquences des sons maternels sur la base de la régularité à laquelle ils les entendent. Ensuite, parmi cette gamme de sons maternels qu'ils perçoivent, les nourrissons apprennent à reconnaître ceux qui sont le plus souvent associés pour segmenter le flux linguistique entendu en plusieurs mots à apprendre (Goodsitt et al., 1993). Du point de vue de la production spécifique, le décalage temporel vis-à-vis de la perception est logique puisqu'il faut percevoir avant de produire. Ainsi, les nourrissons ne produisent les sons spécifiques à leur

langue maternelle qu'à partir de 10 mois. Avant, la production vocale des nourrissons est universelle : à 3 mois ils commencent à produire des sons ressemblant aux voyelles et à 7 mois commence le babillage canonique, qui va permettre aux nourrissons d'exercer leur appareil phonatoire nécessaire pour produire des sons spécifiques à une langue (Kuhl, 2004). L'apprentissage du langage est un processus cumulatif où chaque acquisition va déterminer la perception et l'apprentissage lors de l'étape suivante (fig. 2).

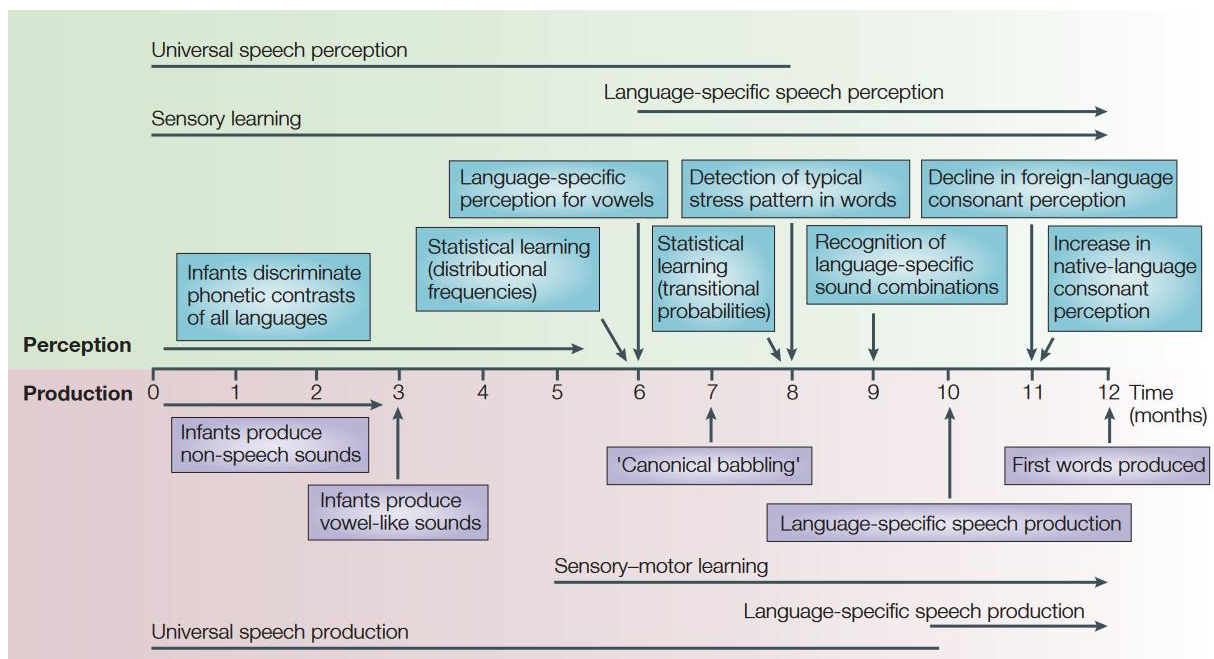


Figure 2 : schéma de synthèse présentant l'agenda développemental de la production et de la perception du langage au cours de la première année, repris de l'article de Kuhl (2004).

Cet apprentissage ne peut se faire que dans le cadre d'interactions avec les parents, dans ces proto-conversations précédemment décrites. En effet, Kuhl et al. (2003) ont montré que chez des nourrissons américains de 9 mois exposés à du Mandarin pendant douze sessions de 25min, soit avec un intervenant humain soit via des vidéos, seuls les nourrissons exposés à l'intervenant humain montraient des signes d'apprentissage. Ceci met en évidence que les nourrissons sont surtout sensibles à l'apport langagier provenant de personnes avec lesquelles ils interagissent en direct, et que le langage apporté par un intermédiaire est moins déterminant. Dans une autre expérience, des nourrissons de 8 mois étaient soit exposés à des interactions contingentes avec leurs mères soit à des interactions non-contingentes : certaines mères ne devaient pas réagir immédiatement aux vocalisations spontanées de leurs nourrissons mais à des vocalisations pré-

enregistrées qu'elles entendaient dans un casque audio (et qui n'étaient donc pas synchronisées avec les vocalisations de leur enfant) (Goldstein et al., 2003). Les nourrissons exposés à des interactions contingentes ont produit un plus grand nombre de vocalisations et de meilleure qualité, c'est-à-dire plus de vocalisations ressemblant à du langage (babillage canonique). Ceci met en évidence qu'au-delà d'interagir en direct avec un humain, c'est bien la contingence des réponses de ce dernier qui est déterminant pour inciter un nourrisson à vocaliser et ainsi favoriser son apprentissage du langage. Enfin, au cours des proto-conversations, les parents parlent d'une manière différente à leur nourrisson qu'ils ne le font aux cours de conversations avec des adultes : ils adoptent une voix plus aigüe en exagérant les intonations et en faisant davantage de répétitions. Ce « parler-bébé » (*Infant-Directed Speech* en anglais, IDS) est préféré par tous les nourrissons par rapport au ton employé dans des conversations entre adultes (The ManyBabies Consortium, 2020) et facilite l'apprentissage du langage. La quantité de ce type de discours entendu par les nourrissons prédit ainsi mieux leur acquisition postérieure du vocabulaire que les autres types de discours auxquels ils sont aussi exposés, comme les discours dirigés vers les adultes (Shneidman et al., 2013). Au final, tous les types de discours ne sont pas déterminants pour l'apprentissage du langage du nourrisson, ce dernier sélectionne le discours qui lui est adressé, au cours d'interactions en direct avec des interlocuteurs qui répondent à ces propres productions. De tous les facteurs influençant son environnement, seuls ceux qui influencent ces interactions vont avoir un impact significatif sur le développement du langage.

2.3. L'impact de l'environnement social

La quantité et la qualité des proto-conversations varient d'un foyer à l'autre, ce qui peut avoir un impact précoce et durable sur le développement des compétences langagières de l'enfant. Bigelow (1998) a par exemple montré que le style d'interaction des parents sur leur enfant a une influence très précoce, puisque les nourrissons de 5 mois manifestent déjà une préférence pour le niveau de contingence temporelle de la mère, même lorsque celui-ci est plutôt bas par rapport à celui des autres mères. Si cela peut paraître anodin, le style d'interaction des parents étant pendant longtemps la seule référence des enfants, il suffit que cette dernière soit en décalage par rapport à la norme pour que les enfants soient en retard et aient des difficultés à s'intégrer socialement lorsqu'ils iront à la crèche ou à l'école. La popularité des enfants à la crèche est en effet lié à leur style d'interaction : les enfants appréciés de leurs pairs sont ceux qui sont plus à même d'initier et de maintenir un discours cohérent, de répondre de manière

contingente et de s'adresser à un individu en particulier, en comparaison des enfants rejetés (Hazen and Black, 1989). Une autre étude a d'ailleurs montré que le style d'interaction de ces enfants rejetés, qui effectuent des tours de paroles trop longs, interrompent fréquemment et échouent à répondre aux demandes, pouvait être rapporté au grand nombre d'interactions avec leurs parents où ces derniers ne leur laissent pas le temps de répondre (Black and Logan, 1995a). Étudier les facteurs de variation des interactions parents-enfants au sein des foyers est donc d'une importance capitale pour comprendre les variations de production et donc probablement de perception des autres interactions.

L'expérience de l'enfant dépend d'un environnement qui comporte de nombreux éléments l'influençant à différents niveaux et à différents moments de son existence. Selon la théorie bioécologique (Bornstein and Arterberry, 2022), le nourrisson est intégré dans un système d'environnements et d'expériences, certains étant plus proches et ayant des effets plus directs sur les nourrissons (comme les parents, les membres de la fratrie, les pairs), d'autres étant plus distants et ayant des effets plus indirectes (la culture, la classe sociale, la religion) (fig. 3). La complexité de ce système rend difficile l'évaluation de tout lien causal entre la performance de l'enfant et les différents facteurs existants.

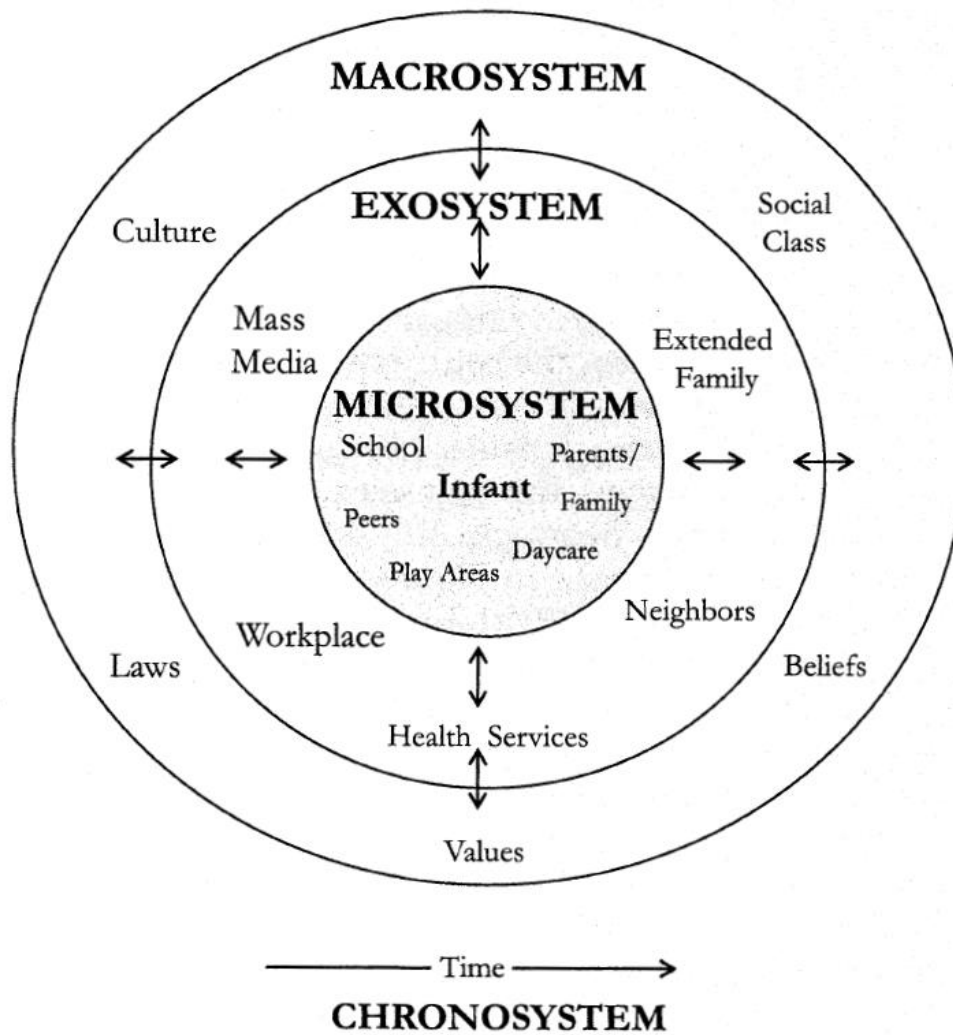


Figure 3 : schéma de synthèse des différents niveaux d'influences de l'environnement d'un nourrisson selon la théorie bioécologique, repris du livre de Bornstein et Arterberry (2022).

Néanmoins, des facteurs sont bien connus dans la littérature pour avoir une influence sur le développement du langage. C'est le cas du statut socio-économique du foyer, défini par le niveau académique et le métier exercé par les parents. Hart et Risley (2003) ont été parmi les premiers à quantifier le retard des enfant de 3 ans issus de famille au statut socio-économique bas par rapport à leurs homologues issus de foyers au statut socio-économique élevé : les premiers avaient entendu 30 millions de mots de moins à 3 ans que les seconds selon les estimations. Plusieurs études ont confirmé ce résultat en montrant que les enfants de familles au statut socio-économique bas avaient moins de vocabulaire et de compétences syntaxiques que ceux de familles au statut socio-économique élevé au même âge (Fernald et al., 2013; Hart and Risley, 2003; Hoff, 2003; Huttenlocher et al., 2010; Rowe, 2008). Comme nous l'avons vu

plus haut, les enfants sont surtout sensibles aux discours qui leur ait directement adressé, et c'est donc la motivation des parents à parler à leur enfant qui peut expliquer ces différences dans le développement. Par exemple, les parents d'un statut socio-économique haut vont plus parler à leurs enfant dans le but de susciter chez eux des réponses vocales (motivent l'interaction) que les parents d'un statut socio-économique bas, qui vont quant à eux plutôt parler à leur enfant dans le but d'orienter leur comportement (motivent l'action) (Hoff, 2006). Surtout, il a été montré que les premiers pratiquent plus le « parler-bébé » (Schwab and Lew-Williams, 2016), qui est crucial pour le développement du langage et des aptitudes conversationnelles.

Un autre facteur très étudié est celui du sexe de l'enfant. Plusieurs études montrent un léger avantage, à un âge donné, pour les filles en termes de production et de compréhension du langage (Berglund et al., 2005; Eriksson et al., 2012; Galsworthy et al., 2000; Zambrana et al., 2012). Si des facteurs endogènes peuvent être invoqués pour expliquer ces différences observées, l'influence du sexe dépend aussi de la manière dont les parents vont interagir avec leurs enfants sur la base de leur sexe, et donc comment les stéréotypes de genre peuvent impacter le développement du langage. En effet, plusieurs études relatent que les parents parlent plus aux petites filles qu'aux petits garçons (Clearfield and Nelson, 2006; Leaper et al., 1998; Lovas, 2011; Sung et al., 2013). La qualité des interactions varie aussi en fonction du sexe de l'enfant : par exemple, les mères de petits garçons font plus de commentaires tandis que les mères de petites filles font plus d'interprétations (Clearfield and Nelson, 2006), ou encore les mères s'engagent plus dans des jeux d'exploration avec les petits garçons alors qu'elles font plus de jeux symboliques avec les petites filles (Bornstein et al., 1999). Comme la perception du sexe de l'enfant par les parents compte pour beaucoup dans le développement du langage et des aptitudes conversationnelles, le sexe est susceptible d'interagir avec d'autres facteurs tels que le statut socio-économique des parents : on observe des différences de maîtrise de la langue en faveur des petites filles dans les classes socio-économiques basses et pas dans les classes socio-économiques hautes (Barbu et al., 2015).

Aussi, plusieurs études relatent une influence de l'ordre de naissance sur le développement du langage. En effet, au sein de fratrie, les enfants entrent « en compétition » pour avoir l'attention de leurs parents qui ne passent pas autant de temps à interagir avec chacun d'eux qu'ils ne passeraient de temps à interagir avec un enfant unique (Jones and Adamson, 1987). Outre le nombre, la qualité des interactions avec les parents qu'ont les enfants vivant dans une fratrie diffère par rapport aux interactions qu'a un enfant unique : il y a notamment plus d'interactions

triadiques, centrées plutôt sur une activité que sur le langage lui-même comme dans les interactions dyadiques (Oshima-Takane and Robbins, 2003). Ainsi, l'environnement d'un cadet diffère de celui d'un aîné puisque ce dernier aura eu des interactions plus nombreuses et de qualité différente au début de son existence. C'est ce qui est confirmé par la littérature, où les aînés ont un avantage en termes de vocabulaire et de compréhension du langage sur les cadets au même âge (Berglund et al., 2005; Dunn and Shatz, 1989; Hoff-Ginsberg, 1998; Pine, 1995; Zambrana et al., 2012). Les cadets, en revanche, développent plus de compétences interactives, avec plus d'habilité à produire des réponses contingentes (Hoff-Ginsberg, 1998), probablement car ils doivent plus lutter pour avoir de l'attention que les aînés.

Comme stipulé par Bornstein et Arterberry (2022), les facteurs d'influence sont très nombreux, si bien qu'un répertoire exhaustif serait impossible ici (pour une revue plus complète voir Hoff, 2006). Si les études sur le développement du langage se sont plutôt focalisées sur les grandes étapes communes dans le but de décrire un schéma général de son apprentissage, étudier les facteurs de variations permet de mieux comprendre les mécanismes sous-jacents en plus de dresser un schéma plus pertinent (Fernald et al., 2013).

3. Des précurseurs chez les autres primates : phylogénèse des conversations

3.1. Etudier les origines du langage grâce aux primates non-humains

Les linguistes ont longtemps vu le langage comme le propre de l'humain, en discontinuité avec les systèmes de communication des animaux non-humains. La société de linguistique de Paris avait notamment ajouté en 1865 dans ses règlements l'interdiction de soumettre la moindre communication concernant l'origine du langage pour éviter les « querelles stériles » et « thèses farfelues ». Aujourd'hui, cette question est sérieusement étudiée scientifiquement par plusieurs disciplines dont certains linguistes (voir par exemple le concept de « Formal monkey linguistic » ; Schlenker et al., 2016). Certains biologistes, adeptes de l'approche comparative, voient dans l'étude de la communication animale l'opportunité de découvrir les précurseurs de certaines propriétés de notre langage, et donc de mettre en évidence des continuités. Les primates non-humains suscitent en ce sens un intérêt particulier du fait de leur proximité phylogénétique avec les humains.

La communication gestuelle des primates non-humains a d'abord été pointé comme une origine probable du langage humain pour sa flexibilité d'utilisation et d'apprentissage (Hewes, 1973).

Plusieurs études ont ainsi montré qu'il était possible d'apprendre de nombreux signes gestuels à des grands singes élevés par des humains afin de pouvoir communiquer avec eux (Gardner and Gardner, 1969; Patterson, 1978; Terrace et al., 1979). En outre, les primates non-humains utilisent de nombreux gestes de manière intentionnelle (Byrne et al., 2017; Tomasello and Call, 2019). Surtout, dans le cerveau humain, les aires de Broca et Wernicke essentielles pour le langage se situent dans l'hémisphère gauche ; or certaines espèces de primates non-humains utilisent préférentiellement leurs bras droits pour effectuer des gestes communicatifs (Meguerditchian et al., 2013, 2011), ce qui implique l'utilisation prononcée de leur hémisphère gauche. Il y aurait donc un lien entre l'utilisation de gestes communicatifs latéralisés à droite chez des espèces proches de l'humain, le développement de l'hémisphère gauche et l'apparition du langage articulé.

La communication vocale aussi suscite l'intérêt des scientifiques dans la recherche des origines du langage. Des études pionnières chez les singes vervets ont en effet mis en évidence la production par les individus de différents cris d'alarmes en fonction du type de prédateur (aigle, léopard, serpent), provoquant des stratégies de défense différentes (Seyfarth et al., 1980a, 1980b). Cela les a conduits à parler de « communication sémantique », bien que le terme plus parcimonieux de « vocalisation fonctionnant par référence » ait été choisi par la suite (Macedonia and Evans, 1993), ouvrant la voie à de nombreuses études relatant l'existence de tels systèmes de communication qui seraient des précurseurs de la sémantique humaine (Cäsar et al., 2012; Kirchof and Hammerschmidt, 2006; Pereira and Macedonia, 1991; Schel et al., 2013; Wheeler, 2010; Zuberbühler, 2000). Des critiques ont néanmoins nuancé ce parallèle, arguant que ces vocalisations pourraient simplement être liées à des états internes que les autres individus associeraient à certaines menaces par conditionnement (Price et al., 2015; Wheeler and Fischer, 2012). Mais le débat n'est pas totalement tranché. En effet, certains primates non-humains comme les mones de Campbell ou les cercopithèques hocheurs combinent plusieurs vocalisations pour modifier le sens d'une manière qui rappelle la syntaxe humaine (Ouattara et al., 2009; Zuberbühler, 2002). Bien que ces propriétés proto-syntaxiques demeurent éloignées de la récursivité de la syntaxe humaine (Hauser et al., 2002), elles suggèrent que les primates non-humains sont capables de communiquer vocalement un peu plus que des états internes. Des preuves d'intentionnalité dans l'émission de ces cris ont même été mises en évidence, chez les chimpanzés par exemple, où les individus exposés à un serpent vocalisaient plus en présence de congénères qui n'avaient pas vu le serpent ou qui n'avaient pas entendu les premiers cris (Crockford et al., 2012). Ces évidences de proto-syntaxe et d'intentionnalité révèlent une

flexibilité dans l'usage qui contraste avec le fort déterminisme génétique du répertoire des primates non-humains (Hammerschmidt and Fischer, 2013). Cette flexibilité pourrait être similaire avec celle de la communication gestuelle, notamment car certains auteurs considèrent aussi que les gestes communicatifs sont innés bien qu'utilisés de manières flexibles (Byrne et al., 2017). L'hypothèse de l'origine vocale du langage humain et l'hypothèse de l'origine gestuelle du langage humain ne sont en tout cas pas exclusives, certains auteurs proposant une origine multimodale (Levinson and Holler, 2014; Prieur et al., 2020).

Au-delà de ces questions portant sur les modalités ou le caractère inné des signaux communicatifs, un autre débat oppose au moins deux conceptions de la communication vocale : l'une se basant sur la théorie de l'information de Shannon-Weaver où l'émetteur coderait une information dans une vocalisation à destination d'un receveur qui n'aurait plus qu'à décoder l'information (Shannon, 1948), et l'autre se basant sur la théorie de la manipulation où l'émetteur cherche à modifier le comportement des receveurs par les effets directs des vocalisations (ex : vocalisations stridentes pour gêner et interrompre un agresseur) ou indirectement par conditionnement (ex : un dominant qui produit un cri qui fait s'écarter les subordonnés car ils ont associé cette vocalisation aux agressions passées) (Owren and Rendall, 2001). De toute évidence, aucune des vocalisations des primates non-humains connues n'encodent un sens précis comme peut le faire le langage humain. Mais des études confirment que les primates non-humains utilisent des éléments contextuels pour inférer le sens des vocalisations perçues (Arnold and Zuberbühler, 2013; Wheeler and Hammerschmidt, 2013). Ces faits sont déterminants car prendre en compte des éléments contextuels pour déterminer le sens d'une vocalisation (qui peut être utilisée très variablement) implique des capacités cognitives autrement plus complexes, plus proches de celles impliquées dans le langage humain (Fischer and Price, 2016). Selon Scott-Phillips (2015), étant donné que le langage humain repose essentiellement sur des aspects pragmatiques d'inférences d'intentions ostensibles (cf. la théorie du sens de Grice évoquée dans la première partie), nous faisons fausse route lorsque nous étudions chez d'autres espèces animales les aspects de surface du langage tels que la sémantique ou la syntaxe. Ce sont les processus cognitifs sous-jacents tels que l'intentionnalité et l'ostension qu'il convient d'étudier pour découvrir, à ce niveau, la différence entre les êtres humains et les autres espèces animales qui sous-tend le reste.

Ici aussi, nous nous plaçons à un niveau en deçà des questions de sémantique et de syntaxe puisque nous étudions comment les individus interagissent entre eux vocalement au sein de conversations. Étant donné l'importance qu'ont les conversations pour notre espèce, et étant

donné que les nourrissons semblent préparés à interagir de manière conversationnelle, ce comportement a des origines évolutives que l'on peut retracer. De même que l'on peut étudier les proto-conversations de nourrissons ne disposant pas encore du langage, on peut étudier les proto-conversations de primates non-humains ne disposant pas eux aussi du langage.

3.2. Des conversations chez les primates non-humains ?

Chez les primates non-humains, des vocalisations peuvent être émises essentiellement selon trois patterns : de manière isolée (un individu produit un simple cri), en série de cris consécutifs (un individu répète plusieurs fois une même vocalisation) ou en interaction (au moins deux individus vocalisent dans une même fenêtre temporelle) (Pougnault et al., 2020a). Parmi ces vocalisations produites en interactions, on peut en distinguer plusieurs types. Il y a les réponses phoniques désorganisées (i), qu'on observe par exemple lorsqu'un cri d'alarme est émis par un individu et suscite des cris en raison de la menace de la part des autres membres du groupe d'une manière chaotique. Il y a les chœurs (ii), qui concernent les cas où plusieurs individus produisent simultanément la même vocalisation, en synchronisant leur production les uns avec les autres (Yoshida and Okanoya, 2005). Il y a aussi les duos (iii), qui impliquent seulement deux individus de sexes opposés. Les duos (parfois appelés chants) sont souvent de longues séquences de différents cris synchronisés (en chevauchement ou alternance) entre les deux partenaires avec des patterns stéréotypés. On retrouve surtout ce pattern chez des espèces monogames, qui émettent des vocalisations à longue distance dans un contexte de reproduction ou de défense de territoire (Geissmann and Orgeldinger, 2000). Enfin, il y a les échanges vocaux (parfois appelés antiphonies) (iv) qui peut impliquer seulement deux individus ou plus. Ce type de pattern se distingue surtout des autres en ce qu'il implique une diversité de partenaires récurrents : des individus de tout âge et de tout sexe peuvent y contribuer (Pika et al., 2018), même si au sein d'un groupe certains partenaires sont plus probables. Les échanges vocaux sont surtout présents dans les taxons où on observe des associations de long terme entre les congénères, et donc ils sont particulièrement présents chez les primates (Yoshida and Okanoya, 2005).

Chaque espèce de primate dispose d'un répertoire de plusieurs vocalisations déterminées génétiquement (Hammerschmidt and Fischer, 2013). Ces vocalisations sont émises dans des contextes différents et selon des patterns différents : il y a des cris d'alarme dont nous avons déjà abondamment parlé, des cris de nourrissage émis en présence d'une source alimentaire

(Slocombe and Zuberbühler, 2006), des cris de menace émis de manière isolée ou en série de cris consécutifs par un individu, des cris de détresse qui sont des réponses phoniques désorganisées émis dans le cas d'une agression d'un membre du groupe (Gouzoules et al., 1984), ou encore des cris de copulation (Zanoli et al., 2021). Le type de vocalisation qui est impliqué dans les échanges vocaux sont appelés cris de contact (contact calls) ou sociaux/affiliatifs (social calls, affiliative calls). Selon plusieurs auteurs, les cris de contacts auraient une fonction dans la coordination spatiale des membres d'un groupe : ils permettraient d'initier et/ou de réguler les mouvements de groupe (Neumann and Zuberbühler, 2016; Sperber et al., 2017), de coordonner la fusion de sous-groupes (Briseño-Jaramillo et al., 2022), et/ou de maintenir un contact entre individus isolés et non-visibles les uns des autres à cause du couvert forestier (Digweed et al., 2007). D'autres auteurs suggèrent une fonction de création et de maintien de lien social, sans critère d'éloignement ou de visibilité. Ces vocalisations de faible intensité sont en effet susceptibles d'être produites à n'importe quel moment de la journée dans des contextes plutôt paisibles. Ils sont même fréquents en captivité où il n'y a pas nécessité pour les individus de se coordonner spatialement. Ils sont en outre la plupart du temps échangés (en antiphonie), c'est-à-dire qu'après une vocalisation d'un individu, une vocalisation d'un autre individu est émise en un temps assez court pour qu'on puisse considérer qu'il lui répond. Snowdon et Cleveland (1984) ont été les premiers à rapprocher ces échanges vocaux aux conversations humaines, en observant trois individus produire successivement ces vocalisations de contact en alternance les uns des autres, et toujours dans le même ordre. D'autres études pionnières menées sur les singes-écureuils ont été dans le même sens en révélant l'existence de patterns temporels dans les échanges vocaux entre individus affiliés : les individus répondaient aux vocalisations des autres dans un délai court de 0.5 sec, et les vocalisations émises en réponse à une première étaient acoustiquement différentes (pic de fréquence plus important) suggérant une forme de questions-réponses (Biben et al., 1986; Masataka and Biben, 1987).

Pour que la comparaison avec les conversations humaines soit pertinente, il faut que ces échanges vocaux suivent un pattern de tour de parole. Les biologistes voulant étudier les interactions vocales des animaux non-humains doivent ainsi transposer la formalisation du tour de parole opéré par les sociologues et leur Analyse des Conversations (Logue and Stivers, 2012). Trois critères peuvent être retenus (Henry et al., 2015) : une alternance de tours entre plusieurs individus (i), des délais qui sont minimisés et donc courts (ii), et un évitement des chevauchements qui sont donc peu nombreux (iii). L'alternance des tours (i) est présente bien que limitée par rapport à l'humain, puisque les échanges vocaux sont souvent présents sous la

forme la plus simple : un individu vocalise une fois en réponse à une vocalisation d'un autre (Lemasson et al., 2011). Bien que minoritaire, il n'est pas rare d'observer le premier individu à avoir vocalisé répondre à la réponse de l'autre individu, ou bien qu'un troisième individu réponde à son tour, mais cela va rarement plus loin que deux alternances entre plus de trois individus. Les marmousets se singularisent sur cet aspect puisqu'on observe de nombreuses alternances chez eux (Takahashi et al., 2013), mais ces alternances en continue concernent des cris de contact à plus longue distance (lorsque les individus sont isolés visuellement) ; pour les cris de contact à courte-distance, l'alternance de tours est comme pour les autres espèces plus limitée (Landman et al., 2020; Yamaguchi et al., 2009). Une expérience de repasse par haut-parleur a montré la pertinence de cette règle d'alternance chez les mones de Campbell, puisque les échanges vocaux repassés qui rompaient cette règle d'alternance ont suscité moins d'attention visuelle que les échanges qui la satisfaisaient (Lemasson et al., 2011). La minimisation des délais de réponses (ii) est aussi un critère satisfait car chez les espèces étudiées, les individus respectent un délai de réponse moyen de l'ordre d'une durée d'un cri, et le délai maximal pour caractériser une vocalisation comme étant une réponse se situe entre une et trois secondes. Ainsi, le délai maximal de réponse varie en fonction des espèces, avec des délais ne dépassant pas une seconde chez les macaques japonais, les mones de Campbell et les singes-écureuils (Katsu et al., 2018; Lemasson et al., 2010; Masataka and Biben, 1987), tandis que chez les singes araignées, les gorilles et les bonobos ils peuvent aller jusqu'à trois secondes (Briseno-Jaramillo et al., 2018; Lemasson et al., 2018; Levréro et al., 2019). Les délais de réponses moyens peuvent varier à la marge au sein d'une même espèce : selon les populations chez les macaques japonais (Sugiura, 1993), ou au sein d'un même groupe selon l'affinité entre les individus chez les singes-écureuils (Biben et al., 1986). En ce qui concerne l'évitement des chevauchements (iii), ce critère est aussi satisfait si on en croit la fréquence de distribution des délais de réponse observés chez les espèces étudiées montrant que les chevauchements sont rares. Toutefois, contrairement aux productions vocales humaines qui sont assez longues pour pouvoir être interrompues volontairement, les cris de contact sont assez courts (quelques centaines de millisecondes) et donc la probabilité que des chevauchements soit présents peut interroger. En revanche, une expérience de repasse menée sur des gorilles captifs a montré que ces derniers manifestaient moins d'intérêt pour un échange vocal où deux cris de deux individus se chevauchaient par rapport à un échange vocal avec un intervalle inter-cris habituel de 500ms (Pougnault et al., 2020b).

3.3. Des proto-conversations au cœur de l'organisation sociale

La quasi-totalité des primates non-humains vivent en groupe social, c'est-à-dire que les individus vont nouer des relations sociales de long terme sur la base d'interaction répétées (Hinde, 1976). Plusieurs dimensions peuvent caractériser ces relations sociales, telles que la parenté, l'affinité, la dominance, le statut reproducteur. La gestion des relations sociales qu'implique la vie en groupe nécessite certaines capacités cognitives, des ressources et du temps. Selon Dunbar (1991, 1995, 2003), chez les primates non-humains, la cohésion d'un groupe (et sa taille) dépend de la force des liens affiliatifs entre ses individus, qui résulte du temps qu'ils passent à se toiletter. Ce temps passé à se toiletter entre en compétition avec d'autres fonctions vitales, comme se nourrir et se reproduire, et est donc limité par ces nécessités. Dunbar observe que le temps alloué au toilettage social est variable en fonction des espèces de primates non-humains, et est corrélé avec la taille du groupe mais ne dépasse jamais 20% de leur temps d'éveil. Mais surtout, il suggère que la fonction du toilettage social aurait pu être partiellement transférée aux interactions vocales, qui permettraient de créer du lien à distance avec plusieurs individus en même temps et durant une autre activité, ce qui la rendrait beaucoup plus efficiente comparée au toilettage. Le langage humain servirait aussi à remplir cette fonction, de manière encore plus efficiente car il permet aux interlocuteurs d'échanger des informations personnelles plus précises : il a d'ailleurs été confirmé que nos relations sociales et nos expériences passées était le sujet principal de nos conversations (plus de la moitié de notre temps à converser est consacré à relater ce genre d'informations éminemment sociales) (Dunbar et al., 1997).

Cette idée que les échanges vocaux ont la même fonction de création de lien que le toilettage social a été confirmée par plusieurs études menées sur différentes espèces, où les partenaires des échanges vocaux étaient les mêmes que les partenaires de toilettage social. (Arlet et al., 2015; Briseno-Jaramillo et al., 2018; Kulahci et al., 2015; Lemasson et al., 2018, 2006; Levréro et al., 2019). Ces échanges vocaux peuvent d'ailleurs servir à initier une interaction affiliative, puisqu'ils sont plus souvent suivis de toilettages ou de contacts que des cris isolés (Katsu et al., 2016). L'affinité entre les individus peut influencer d'autres paramètres de ces échanges vocaux. En effet, chez les singes-écureuils, les délais de réponses de ces échanges sont plus courts entre deux individus affiliés qu'entre deux individus qui ne le sont pas, probablement en raison de la motivation des individus à répondre à leurs partenaires préférés (Biben et al., 1986). Chez les mones de Campbell, les individus produisent généralement plusieurs variants

acoustiques du type de cri consacré aux échanges vocaux, et les individus liés d'affinité qui ont donc des interactions vocales régulières sont les plus susceptibles de partager ces variants entre eux (Lemasson et al., 2003). La présence de ces variants dans le groupe évoluent au cours du temps, et des variants peuvent changer, apparaître ou disparaître, en fonction de la dynamique sociale du groupe (Lemasson and Hausberger, 2004).

D'autres facteurs que l'affinité peuvent jouer un rôle dans la régulation de ces échanges vocaux. L'âge est un de ces facteurs. Les individus les plus âgées sont ainsi les partenaires d'échanges préférés chez les mones de Campbell, puisque leurs vocalisations suscitent plus de réponses de la part de leurs congénères malgré le fait qu'ils vocalisent moins (Lemasson et al., 2010). Chez les gorilles, les individus préfèrent échanger vocalement avec des congénères qui ont un âge proche du leur (Lemasson et al., 2018). Le statut de dominance peut aussi être un facteur déterminant chez certaines espèces. Les individus dominants peuvent être des partenaires préférés et recevoir plus de réponses (Digweed et al., 2007), ou bien répondre eux-mêmes plus aux vocalisations des autres (Lemasson et al., 2018).

3.4. Des règles apprises socialement ?

Plus que des patterns, certains paramètres de ces échanges vocaux agissent comme de véritables règles dont le non-respect est perçu par les individus : c'est le cas de l'alternance des interlocuteurs, le non-chevauchement des vocalisations, et aussi de la convergence acoustique des vocalisations de l'échange (Lemasson et al., 2011; Pognault et al., 2020b; Bouchet et al., 2017). Plusieurs études ont mis en évidence un rôle de l'apprentissage dans la perception et dans la maîtrise de ces échanges vocaux. Lors d'expériences de repasse effectuées chez les mones de Campbell par exemple, seuls les juvéniles ne percevaient pas la différence entre un échange vocal qui respectait la règle d'alternance et un échange vocal qui ne la respectait pas (Lemasson et al., 2011). Ceci confirme des observations montrant que les juvéniles ne respectaient pas le tour de rôle puisqu'ils chevauchaient 12 fois plus les vocalisations des autres que les adultes, et étaient plus prompts à répéter la même vocalisation sans attendre une réponse. Chez les macaques japonais aussi, une expérience de repasse a révélé que les juvéniles ne reconnaissaient pas la règle de la convergence acoustique : ces derniers ne faisaient pas la différence entre, d'une part, des échanges vocaux où les vocalisations avaient des modulations de fréquences similaires et, d'autre part, des échanges vocaux qui n'étaient pas ajustés acoustiquement (Bouchet et al., 2017). Des observations ont en outre montré que les juvéniles

de macaques japonais ne produisent pas de vocalisations selon un tour de rôle contrairement aux femelles adultes, ces dernières produisant plus de vocalisations en réponse à une autre malgré qu'elles produisent au total moins de vocalisations que les juvéniles (Lemasson et al., 2013). On retrouve le même phénomène chez les singes araignées avec des juvéniles qui échangent vocalement moins que les adultes et qui respectent moins la règle de la convergence acoustique (Briseno-Jaramillo et al., 2018). Chez les ouistitis communs, non seulement l'échange vocal est un comportement appris, mais en plus, l'apprentissage est guidé par les parents (Chow et al., 2015a). En effet, quand les petits produisent une vocalisation qui chevauche celles de leurs parents, ces derniers sont moins susceptibles de répondre. De plus, quand les petits répondent aux vocalisations des parents par une vocalisation différente non-appropriée pour un échange vocal, ces derniers les interrompent en vocalisant par-dessus (Chow et al., 2015a).

4. Problématique et objectifs de la recherche

4.1. Problématique

Comprendre un comportement implique notamment de connaître le schéma développemental de ce dernier au cours de la vie des individus (ontogénèse), ainsi que sa mise en place au cours l'évolution (phylogénèse) (Tinbergen, 1963). Les deux sont souvent intimement liés puisque les comportements se développant le plus précocement chez tous les individus d'une même espèce sont susceptibles d'avoir une histoire évolutive profonde, justifiant une approche comparative avec des espèces proches. Le fait que les conversations sont largement répandues dans les cultures humaines en plus de se développer très précocement suggère en effet que des espèces de primates non-humains disposent de précurseurs de ce type de comportement. Comme nous l'avons, les échanges vocaux de cris de contact à courte distance sont de bons candidats, car ils ont une structure temporelle et une fonction de création de lien similaire aux conversations humaines. Pour examiner cette hypothèse, il est difficile de comparer ces échanges vocaux impliquant des vocalisations simples aux conversations humaines où des interlocuteurs sont susceptibles de transmettre des informations précises grâce au langage. Il est plus pertinent de les comparer avec les proto-conversations pratiquées par les nourrissons préverbaux, sachant que ces dernières sont les précurseurs du langage d'un point de vue de l'ontogénèse. Nous pouvons donc nous demander ce qui, dans ces deux formes de proto-

conversations, est similaire (et ce qui diffère) d'un point de vue de la structure temporelle et du lien avec des facteurs sociaux.

4.2. Objectifs communs

Nous avons donc décidé d'étudier deux modèles biologiques : des nourrissons préverbaux humains, et des primates non-humains adultes de l'espèce des mangabeys à collier (*Cercocebus torquatus*). Pour ces deux modèles, nous avons suivi la même logique en étudiant d'une part la production des proto-conversations grâce à des observations d'interactions spontanées, et d'autre part la perception des proto-conversations grâce à des expérimentations. L'observation d'interactions spontanées nous permettent d'identifier des patterns temporels en lien avec des facteurs sociaux, afin d'explorer la flexibilité sociale des individus étudiés. Les expérimentations nous permettent de tester la pertinence de certains patterns (observés par nous ou par d'autres) en mesurant pour les deux modèles les temps de regard, qui sont des indicateurs comportementaux très utilisés dans les expériences étudiant la perception chez les humains et les primates non-humains (pour les humains voir Aslin, 2007; pour les primates non-humains voir Zuberbühler, 2014). Cela nous permet en outre de voir si les patterns testés agissent comme des règles connues par les individus qui les pratiquent.

4.3. Objectifs et méthodes des études sur les nourrissons humains

Si de nombreuses études se sont intéressées à la façon dont les nourrissons interagissaient avec leurs parents, relativement peu d'études ont été menées sur la perception des interactions entre humains par les nourrissons. On peut citer les études menées sur la prédiction de la fin des tours de paroles mais elles concernent des enfants qui commençaient déjà à parler (par exemple : Casillas and Frank, 2017). Chez des nourrissons préverbaux, l'étude de Augusti et al. (2010) a montré une préférence des nourrissons de 6 mois pour des interactions où les interlocuteurs étaient face-à-face par rapport aux interactions dos-à-dos. S'intéresser à la perception des nourrissons préverbaux des conversations a donc un intérêt à la fois pour notre approche comparative mais aussi pour les disciplines étudiant le développement du langage et la cognition sociale chez les nourrissons. Nous nous sommes donc demandé dans quelle mesure des nourrissons préverbaux comprenaient la règle du tour de parole. Ainsi, nous avons mis en place deux expériences où nous avons présenté à chaque sujet trois saynètes vidéo de

conversation entre deux interlocutrices : : une saynète où B parle juste après que A ait fini son tour (tour de parole), une saynète où A donne explicitement la parole à B en s'adressant à elle (passage de parole), et une saynète où B coupe la parole de A (coupure de parole). Notre prédiction était que les nourrissons allaient percevoir différemment la situation de coupure par rapport aux deux autres car seule celle-ci présente une rupture de la règle du tour de parole. Lors de ces deux expérimentations, nous avons étudié le lien de la perception des nourrissons avec des facteurs socio-démographiques reconnus pour influencer le développement du langage. La première expérimentation concerne des nourrissons de 6 mois seulement, et étudie la perception de ces derniers en lien avec leur sexe, le statut socio-économique des parents, et la présence de fratrie. La seconde expérimentation étend le dispositif à une palette d'âge plus large puisqu'elle concerne des nourrissons de 3 à 11 mois afin d'étudier l'influence de l'âge sur la perception, et rajoute un facteur socio-démographique en plus des autres : le mode de garde du nourrisson. Enfin, nous avons aussi observé quelques-uns de ces nourrissons en interaction avec leurs parents pendant 10min (soit la mère, soit le père, soit les deux mais l'un après l'autre). Bien qu'il existe de nombreuses études portant sur l'observation d'interactions spontanées, la plupart portent sur des interactions mère-enfant et beaucoup plus rarement sur des interactions père-enfant. Nous avons donc voulu voir si le pattern de tour de rôle différait en fonction du sexe du parent, en relation avec le sexe du nourrisson.

4.4. Objectifs et méthodes des études sur les mangabeys à collier (*Cercocebus torquatus*)

Chez les primates non-humains, la littérature scientifique sur les échanges vocaux en tant que proto-conversation est moins vaste que celle sur les proto-conversations parents-nourrissons. De plus en plus d'études parlent explicitement de tour de rôle (*turn-taking*) pour parler d'alternance de vocalisations entre individus différents, mais ces alternances peuvent concerner des chœurs ou des duos, qui ont des fonctions non-analogues aux conversations humaines. Si nous nous focalisons sur les échanges vocaux de cris de contact à courte-distance, les études sont moins nombreuses et concernent un nombre encore trop restreint d'espèces. Il est important d'étudier ce type très particulier d'échanges chez un grand nombre d'espèces de primates non-humains pour mettre en évidence les points communs et les différences et les mettre en relation avec les spécificités de leurs systèmes sociaux. Dans cette optique, nous avons étudié les échanges vocaux chez les mangabeys à collier (*Cercocebus torquatus*).

Les mangabeys à collier font partie des Papioninés avec les géladas, les babouins, les macaques, les drills et les mandrills. Ils sont reconnaissables pour le pelage blanc entourant le cou et la tête avec des poils roux sur le dessus de la tête. Il y a un dimorphisme sexuel, avec des mâles adultes pesant 10 kg et mesurant 60 cm, et des femelles adultes plus petites pesant 6 kg et mesurant 50 cm (Gautier-Hion et al., 1999; Swedell, 2010). Ils vivent dans les forêts et les mangroves d'Afrique de l'Ouest bordant le Golfe de Guinée, et se déplacent de manière semi-terrestre (Cooke, 2012). Ils se nourrissent principalement de graines et de fruits, et parfois d'invertébrés. Ils vivent dans de larges groupes multimâle/multifemelle (jusqu'à environ 80 individus) avec des dynamiques de fission-fusion saisonnière en fonction de la disponibilité en fruit (les sous-groupes étant plutôt composé d'une trentaine d'individus) (Dolado et al., 2016). Leur organisation sociale est qualifiée de semi-despotique/semi-tolérante car elle a à la fois des caractéristiques des sociétés dites despotiques et des sociétés dites tolérantes. En effet, les mangabeys à collier ont une hiérarchie forte et linéaire avec des agressions fréquentes unidirectionnelles (comme dans les sociétés despotiques) ; mais les comportements agressifs sont de faible intensité avec les individus dominants qui se positionnent en périphérie du groupe, et il y a des liens affiliatifs forts entre les individus avec des toilettages réciproque (comme dans les sociétés tolérantes) (Dolado and Beltran, 2012, 2011). Enfin, leur répertoire vocal est connu et composé de 8 types de cris, dont un cri de contact à courte distance déjà repéré comme étant émis en contexte affiliatif (Bouchet et al., 2012, 2010). Pour son organisation sociale complexe et son répertoire vocal, le mangabey à collier est une espèce parfaite pour le genre d'études que nous avons mené.

Ainsi, nous avons voulu confirmer l'existence d'échanges vocaux selon un pattern de tour de rôle chez les mangabeys à collier avant d'étudier l'influence de facteurs sociaux sur ces derniers. Nous avons donc observé des groupes d'individus captifs à la Station biologique de Paimpont en notant les comportements sociaux et les vocalisations émis par les individus. Nous avons mené des analyses au niveau individuel pour voir s'il y avait un lien entre le statut social des individus et leur participation aux échanges vocaux, et des analyses au niveau dyadique pour voir si les dyades d'échanges dépendaient de facteurs sociaux. Ensuite, nous avons mené une expérience de repasse pour voir si le pattern temporel de tour de rôle observé avait du sens pour les mangabeys. Nous avons pour cela présenté à chaque individu trois types d'échanges vocaux construits artificiellement : un échange vocal où un individu B répond rapidement à la vocalisation d'un individu A (500ms), un échange vocal où un individu B met plus de temps à répondre à la vocalisation d'un individu A (1500ms), et un échange vocal raté où la vocalisation

d'un individu B chevauche de 50% la vocalisation d'un individu A. Nous avons mesuré les temps de regards de chacun des individus après ces repasses pour évaluer les différences en fonction des trois conditions. Notre prédiction était que les individus testés allaient réagir différemment à la repasse présentant un chevauchement par rapport aux deux autres car elle est la seule à être une rupture de la règle du tour de rôle. Mais une différence de réaction entre la repasse où il y a un délai de réponse court avec celle où il y a un délai de réponse long nous intéressait aussi, car cela signifierait que les individus ne perçoivent pas tous les échanges vocaux de la même manière.

Chapitre II :

Proto-conversations chez les nourrissons humains

Article 1 :

Of the importance to reconsider individual variability in infant studies: family traits do impact conversational skills in 6-month-olds.

Bastien Meunier¹, Stéphanie Barbu¹, Alice Rabiller¹, Nicolas Dollion², Alban Lemasson^{1,3}, Virginie Durier¹

¹ Univ Rennes, CNRS, Normandie Univ, EthoS (Éthologie animale et humaine) - UMR 6552, Rennes, France

² EA 6291 Laboratoire C2S (Cognition Santé Société), Université de Reims Champagne-Ardenne, 51100, Reims, France

³ Institut Universitaire de France, 1 rue Descartes, 75231 Paris, Cedex 05, France

Synthèse de l'article 1

Contexte

Le tour de parole est un pattern universel des conversations humaines qui se caractérise par un échange rapide de tours de parole entre les interlocuteurs et par l'absence de chevauchement. Le langage est intégré dans cette compétence conversationnelle acquise bien avant, pendant la petite enfance, à travers les interactions quotidiennes qu'a le nourrisson avec ses parents. La précocité de cette compétence et son lien avec le langage nous permettent de tester l'influence de l'environnement social sur le développement du langage avant son apparition, chez des nourrissons préverbaux.

Méthode

Nous avons donc étudié la perception du tour de parole chez des nourrissons de 6 mois, en mesurant leurs regards pendant la présentation vidéo de trois situations conversationnelles différentes : une situation où la locutrice A donne explicitement la parole à la locutrice B (passage de parole), une situation où la locutrice B prend elle-même la parole juste après la locutrice A (tour de parole), et une situation où la locutrice B coupe la parole de la locutrice A (coupure de parole). Nous avons analysé le lien des temps de regards de 51 nourrissons avec des facteurs socio-démographiques comme le statut socio-économique, la présence d'une fratrie, et le sexe du nourrisson testé.

Résultats

Les nourrissons ont globalement plus regardé la deuxième locutrice après qu'elle ait coupé la parole à la première que dans les autres situations conversationnelles. Cette sensibilité dépendait du statut socio-économique de la famille ainsi que de la présence d'une fratrie au sein du foyer. En effet, seuls les nourrissons issus des familles au statut socio-économique haut et les nourrissons qui n'avaient pas de fratrie ont réagi différemment à la situation de coupure de parole par rapport aux autres situations. Ainsi, nos résultats mettent en évidence l'influence de l'expérience sociale précoce sur une compétence cruciale pour le développement du langage.

Of the importance to reconsider individual variability in infant studies: family traits do impact conversational skills in 6-month-olds.

Bastien Meunier¹, Stéphanie Barbu¹, Alice Rabiller¹, Nicolas Dollion², Alban

Lemasson^{1,3}, Virginie Durier¹

¹ Univ Rennes, CNRS, Normandie Univ, EthoS (Éthologie animale et humaine) - UMR 6552, Rennes, France

² EA 6291 Laboratoire C2S (Cognition Santé Société), Université de Reims Champagne-Ardenne, 51100, Reims, France

³ Institut Universitaire de France, 1 rue Descartes, 75231 Paris, Cedex 05, France

Abstract

This study investigated, in preverbal infants, the influence of family characteristics on the turn-taking rule perception, a conversational skill crucial for later language development. Looking times of 6-month-old infants (N=51, 27 girls) following video presentations of 3 different conversational situations were assessed. Infants varied in their sensitivity to these situations, depending on their family socioeconomic status (SES) and the presence of siblings. Only infants from high SES families or singletons reacted differently to the violation of a turn-taking compared to other conversational patterns (i.e. without overlap between locutors). These results suggest that differences related to social experience can be observed as early as 6 months, that is well before infants master language, even for a pragmatic language skill.

Keywords: preverbal infants, pragmatics acquisition, social context, turn-taking

Introduction

If human language fascinates by its way of expressing the most sophisticated thoughts through a system of abstract signs, we cannot understand its main issues without studying its most daily and ordinary use. Language is indeed mostly performed in conversations between several actors following a turn-taking pattern; which has been first described by Sacks et al. (1974) as: speakers rapidly exchange turns whose length and content are not specified in advance, and do so while avoiding overlaps and long pause. Turn-taking has been found in every human cultures, despite some marginal variations (in the average time between two turns for example) (Stivers et al., 2009), and to different degrees in other animal species (Pika et al., 2018), suggesting a cognitive basis as deep as language, if not deeper.

This depth is confirmed by the ontogeny as the first turn-taking interactions occur in infancy well before language production (Murray, 1985). Indeed, infants start to respond vocally in alternation with their parents from their 3rd month (Bloom, 1975; Stern et al., 1975), a pattern that becomes predominant in the following weeks compared to concomitant vocalizations (Ginsburg and Kilbourne, 1988). During these early interactions, infants are already active participants as they themselves anticipate and initiate turns (Gratier et al., 2015). Turn-taking skills continue to increase after 5 months, with infants overlapping less and less (Hilbrink et al., 2015). According to Levinson (2016), the precedence of these pragmatic skills over linguistics skills confirms that language is only a part of a larger whole named “interaction engine”, a universal features from which language develops.

Learning to interact according to a turn-taking pattern is of paramount importance for infants. On the one hand, it helps to develop language because infants need the socio-interactional matrix surrounding language to motivate them to engage with it. Kuhl et al. (2003) showed evidence of learning in 9-month-old infants exposed to live sessions with an adult speaker, evidence that was not found when infants were exposed to recorded sessions through a screen or loudspeakers. Moreover, in 8-month-old infants, contingent social interactions (verbal and non-verbal) with mothers promote more complex vocal behavior than non-contingent interactions (Goldstein et al., 2003). This shows that language input alone is not enough to learn to speak: live human interactions are needed, where the timing of the responses and the link the response have with what precedes it is what matters. On the other hand, turn-taking acts as a rule in our society, and its violation leads to being perceived as “less sociable” (Robinson and Reis, 1989) and as having a “negative and strong personality” (ter Maat et al., 2010). Acquiring these conversation skills early is therefore crucial: young children who are the most accepted

by their peers are those able to initiate and maintain coherent speech across interaction contexts (Hazen and Black, 1989). Rejected children perform lengthy turns, interrupt peers and respond noncontingently to requests. This interaction style could be related to a high proportion of interactions involving failed requests observed in their homes (parents do not leave time to their children to respond) (Black and Logan, 1995). The shaping of interaction style by parents is attested as early as 5 months of age, where infants already prefer their mother's level of temporal contingency to other levels, even when the former is rather low (Bigelow, 1998).

Several family and individual characteristics have been identified as significant predictors of variations in language development. Family socioeconomic status (SES), defined by a combination of parents' education and profession, is a major example. Indeed, many studies have reported differences in children's language production and perception development from the second year, always in a sense that children from low SES had less vocabulary and syntactic skills than those from high SES (Fernald et al., 2013; Hart and Risley, 2003; Hoff, 2003; Huttenlocher et al., 2010; Rowe, 2008). The birth order is also known as a factor that influence language development, as later-born children have to compete with their older siblings to get parent's attention. Thus, within families, first-born children develop vocabulary and language comprehension faster than later-borns (Berglund et al., 2005; Dunn and Shatz, 1989; Hoff-Ginsberg, 1998; Pine, 1995; Zambrana et al., 2012). More broadly, being singletons or having siblings can be a source of variations across homes in terms of quantity and quality of interactions. Sex is also a well-studied factor in language development literature. Studies show a slight advantage for girls in terms of both production and comprehension of language skills (Berglund et al., 2005; Eriksson et al., 2012; Galsworthy et al., 2000; Zambrana et al., 2012). In addition, sex is known to be a factor that can interact with others, such as SES (Barbu et al., 2015) or presence of siblings (Havron et al., 2019).

Despite such great variability among young children in language outcome and language experiences, studies concerning infants remain quite silent on this variability. The researchers' desire to identify and highlight a general pattern of language development contributed to overlook potential sources of variability (Fernald et al., 2013). Recruiting children is also difficult, expensive, and time-consuming, leading most studies to deal with small sample sizes. In addition to causing false conclusions and replication problems in infant studies (Oakes, 2017; The ManyBabies Consortium et al., 2020), these small sample sizes do not allow to simultaneously explore the contribution of several factors of variability at the same time. Finally, the recruitment bias towards convenience samples means that most studies on infants are conducted on a very specific sample of the population (children from higher-SES families)

(Fernald et al., 2013), casting doubt on the relevance of the common language development pathway. In contrast, investigating factors of variability allow us to understand underlying mechanisms and thus to draw a more precise developmental pathway.

Thus, we conducted our study on a rather large sample of 51 infants (after exclusion) to cover a wide economic strata with 4 SES-classes (from low to high), to include a relevant number of singletons as well as infants with older sibling(s), as well as a balanced ratio of girls and boys. Six months is the onset of a crucial pragmatic skill for human beings: the ability to manage turns within conversational prototypes. Therefore, we chose this age to investigate the impact of social characteristics on this onset. We used audiovisual sketches of conversations because they combine lexical, syntactic, prosodic, and non-verbal cues together, which corresponds to interactions in naturalistic contexts and which is relevant for an infant who can't yet understand meaning based on lexico-syntactic cues only. Among studies on conversational skills in the first year, studies on perception are rare compare to those on production (Augusti et al., 2010; Bakker et al., 2011) and they did not consider factors of variability as in studies of language development from the second year. Our aim was to overcome this by studying the extent to which social experience of 6-month-old infants already shape an important language feature.

Methods

Participants

Sixty-five 6-month-old infants participated in our study. Fourteen were excluded for fussing or crying during the experiment (6), due to technical problem (1) or insufficient looking time, i.e. looking time on each video <50% or not looking at the pictures (7). Therefore, statistical analyses are based on data from 51 infants (age: mean = 6;9, range = 6;1 – 6;26). All participants were French monolingual, born fullterm and had not been diagnosed with a developmental disorder at the time of the experiment.

Sociodemographic characteristics considered for the analysis were: the infant's sex (27 girls – 24 boys), the presence of siblings in the family (37 with sibling(s) – 14 singletons) and the socio-economic status of the family. The latter was estimated using the educational level and the profession of both parents, and was divided into four categories: both parents with high-school level or less (low SES, 12); at least one parent with bachelor degree (undergraduate) (middle SES, 6); one parent with a postgraduate degree (middle-high SES, 12); both parents with a postgraduate degree (high SES, 21).

Families were contacted by mailing. Parents willing to participate came at the lab with their infant. Prior to the infant's participation in the study, parents were informed of its aim and procedure and were asked to sign a consent form.

Procedure

Set-up: Infants were seated in a car seat into a white cubicle with a screen (24") placed 60 cm in front of them (DELL U2410). A computer (DELL Precision M4600) was connected to the screen to broadcast the test video sequence. A camera (Sony HDR-SR12) placed above the screen was connected to another screen outside the cubicle to allow parents and experimenters to monitor the infant all along the experiment and to end it in case of fussing. The camera also recorded the infant behavior during the video sequence and images were analyzed off-line.

Test (Fig. 1): Infants watched a video sequence consisting of three sketches, each preceded and followed by pictures for 10 sec. These pictures represented the two women playing in the sketch. Each one being displayed on the side of the screen she appeared during the sketch.

In each sketch, two women told different short non-familiar stories invented for the purpose of the study while watching the camera (i.e. as if they spoke to the infant). They smiled and spoke in infant-directed speech. Each sketch staged one of three different conversational situations. In the situation 1 (S1), the first speaker told her story, stopped speaking at the end, turned toward the second woman and said "*Et toi ?*" (*And you?*). The second woman then turned back toward the camera and told her story. This situation was called "Turn-giving". It was the only situation where the women looked at each other. In the other situations, they only looked at the camera and therefore at the infant. In the situation 2 (S2), the first speaker told her story and stopped speaking at the end. Then, the second woman started telling her story. This situation was called "Turn-taking". In the situation 3 (S3), the first speaker told her story. Without waiting for the end, the second woman started telling her own story, therefore overlapping the first woman speech for 1.4 seconds (± 0.2) on average, i.e. during 3 to 5 words. This situation was called "Overlapping".

Within a video sequence, each sketch staged a different pair of women telling a different pair of stories, so each infant saw 6 women relating 6 different stories. Thirty-six video sequences were built with each situation appearing in various places within the three sketches sequence (i.e. S1 could be seen first, second or third = sketch order), with each pair of women staging one of the three situations (i.e. Pair 1 could play S1, S2 or S3) and with each woman playing first or second speaker role (so with the second speaker on the left or the right side of the screen).

Sketches lasted less than 30 sec (mean duration in seconds \pm s.d., S1: 28.2 ± 1.2 ; S2: 26.7 ± 2.1 ; S3: 23.3 ± 1.4).

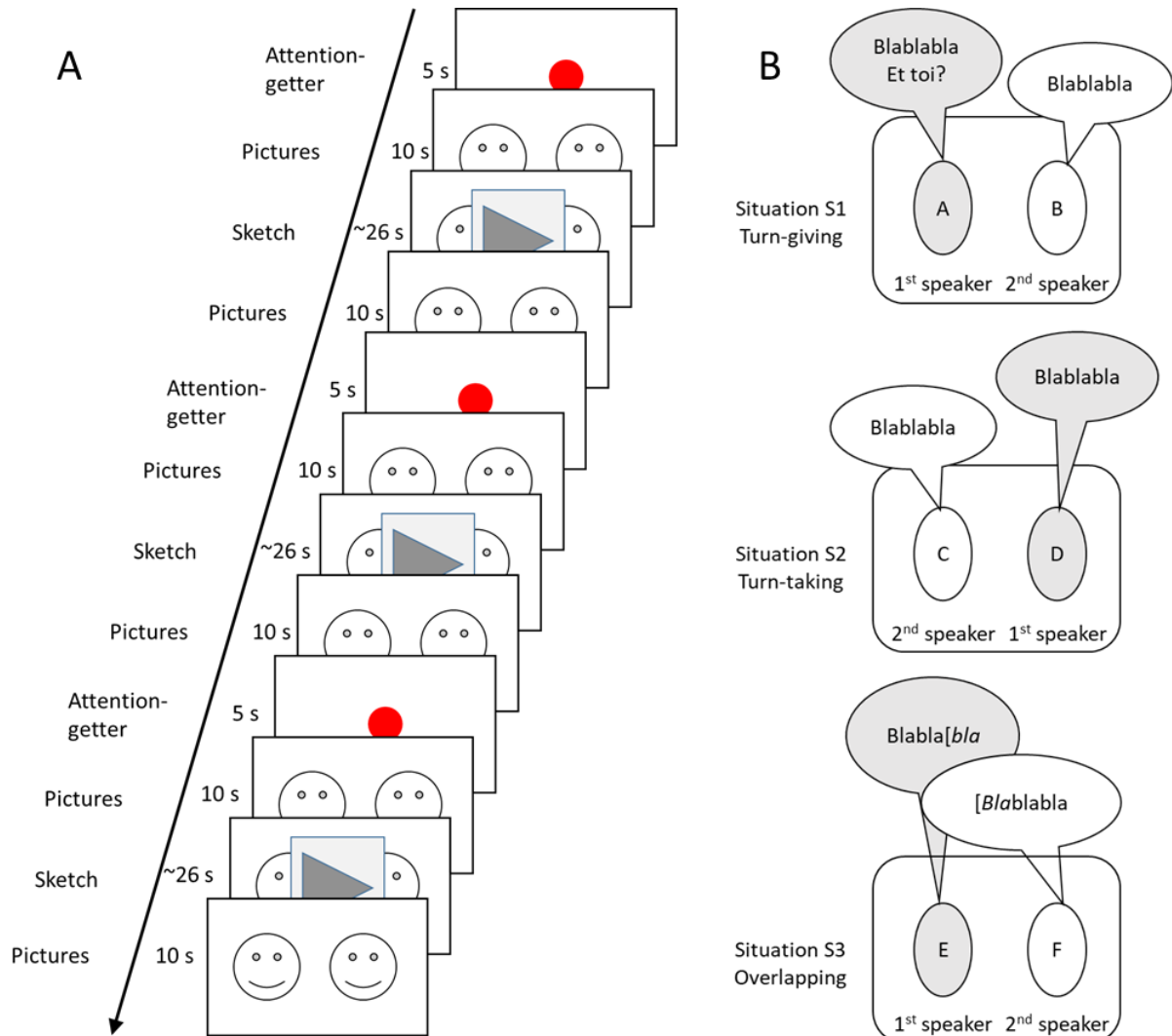


Fig. 1: Experimental procedure to assess the perception of turn taking violation in 6-month-old infants. A) Time scale of a video sequence with three sketches (one for each situation), each preceded and followed by the pictures of the women playing in the sketch. Each batch “pictures-sketch-pictures” was preceded by an attention-getter, i.e. a flashing red circle associated with a rebounding sound. B) Schemes of the three situations. Letters represents different women. In gray the woman speaking first during the sketch, in white the woman speaking second, in italics the overlapping part in situation S3. The first speaker was randomly on the right or on the left side of the screen.

Measure and statistical analysis

We measured the time each infant spent looking at each woman during the pictures’ display. We then calculated the ratio of time spent looking at the 2nd speaker over the time spent looking

at any of the speakers ($\text{ratio_LT} = \text{LT speaker 2} / (\text{LT speaker 1} + \text{LT speaker 2})$) for each display. Finally, we calculated the difference of these ratios between the picture displayed after the sketch and the picture displayed before ($\delta_LT = \text{ratio_LT AFTER} - \text{ratio_LT BEFORE}$), in order to assess if the behavior of the 2nd speaker during the sketch had any impact on the infant's looking time toward this person. We chose to use ratios rather than raw data to lower the potential impact of short overall looking time toward both women.

The impact of several experimental and sociodemographic factors on δ_LT was assessed using a Linear Mixed Model (LMM). The infants' identity was assigned as a random factor. Fixed effects were the situation displayed in the sketch (S1, S2 or S3), the sketch order (seen 1st, 2nd or 3rd), the pair of women (3 pairs), the position of the 2nd speaker (on the left or the right side of the screen), the infant's sex (F or M), the family SES (4 modalities) and the presence of sibling(s) (Yes or No). Interactions between Situation and the other factors were also set as fixed effects into the model. The model was built using lmerTest package (Kuznetsova et al, 2017) in R 3.5.0 (2018) and it was written as follows:

$$\delta_LT \sim \text{Situation} + \text{Sketch order} + \text{Pair of women} + \text{Position 2}^{\text{nd}} \text{ speaker} + \text{Sex} + \text{SES} + \text{Sibling} + \text{Situation} : (\text{Sketch order} + \text{Pair of women} + \text{Position 2}^{\text{nd}} \text{ speaker} + \text{Sex} + \text{SES} + \text{Sibling}) + (1|\text{Infant_id})$$

Post-hoc tests were performed using the emmeans package (Lenth, 2019) in R 3.5.0 (2018).

Results

Our analyses revealed that 6-month-olds were sensitive to turn-taking violation. An influence of the family environment on infants' conversational skills the preverbal period was present: only infants from high-SES family and singletons reacted to overlapping during sketches, with an increased LT toward the second speaker.

The situation factor had a significant effect on the δ_LT when considered alone (LMM, Situation: $\chi^2 = 12.88$, $df = 2$, $p = 0.002 < 0.001$, Table S1). Furthermore, two sociodemographic factors interacted with the situation factor.

First, family SES had a major impact on infants' reactions to the violation of the turn-taking rule. Indeed, we found a significant interaction between Situation and family SES, assessed through both parent's educational level and profession (LMM, Situation x SES: $\chi^2 = 14.01$, $df = 6$, $p = 0.03$; Table S1). Only infants from high-SES families reacted differently to the three situations (Fig.2, see Table S2 for post-hoc comparisons). They looked longer at the second speaker after her speech overlapped with the first speaker than in the two other situations

respecting the overlap avoidance (S3 vs. S1: t .ratio = -3.34, df = 80.4, p = 0.008; S3 vs. S2: t .ratio = -4.11, df = 80.6, p = 0.001). A similar pattern was observed for middle-high-SES infants when comparing S3 (overlapping) to S2 (turn-taking) (t .ratio = -2.79, df = 81, p = 0.026). Infants from the other two lower-SES categories did not show different looking reactions in relation to the situations (all p > 0.23).

Table S1: Effect of experimental and sociodemographic factors on the difference of looking time on the 2nd speaker after and before a sketch. Results of the LMM with all fixed effects and interactions (package lmerTest (Kuznetsova et al, 2017) in R 3.5.0 (2018)). Significant differences are indicated by * for p <0.05 and ** for p <0.01.

Fixed factors and interactions	Chi ²	df	P value
Situation	12.9	2	0.002 **
Sketch order	1.3	2	0.531
Pair of women	3.2	2	0.204
Position of the 2 nd speaker	1.9	1	0.164
Infant's sex	0.4	1	0.541
Family SES	5.6	3	0.133
Presence of sibling	1.4	1	0.229
Situation:Sketch order	6.9	4	0.138
Situation:Pair of women	2.1	4	0.711
Situation:Position of 2 nd speaker	4.2	2	0.125
Situation:Infant's sex	0.6	2	0.741
Situation:Family SES	14	6	0.03 *
Situation:Presence of sibling	9.2	2	0.01 *

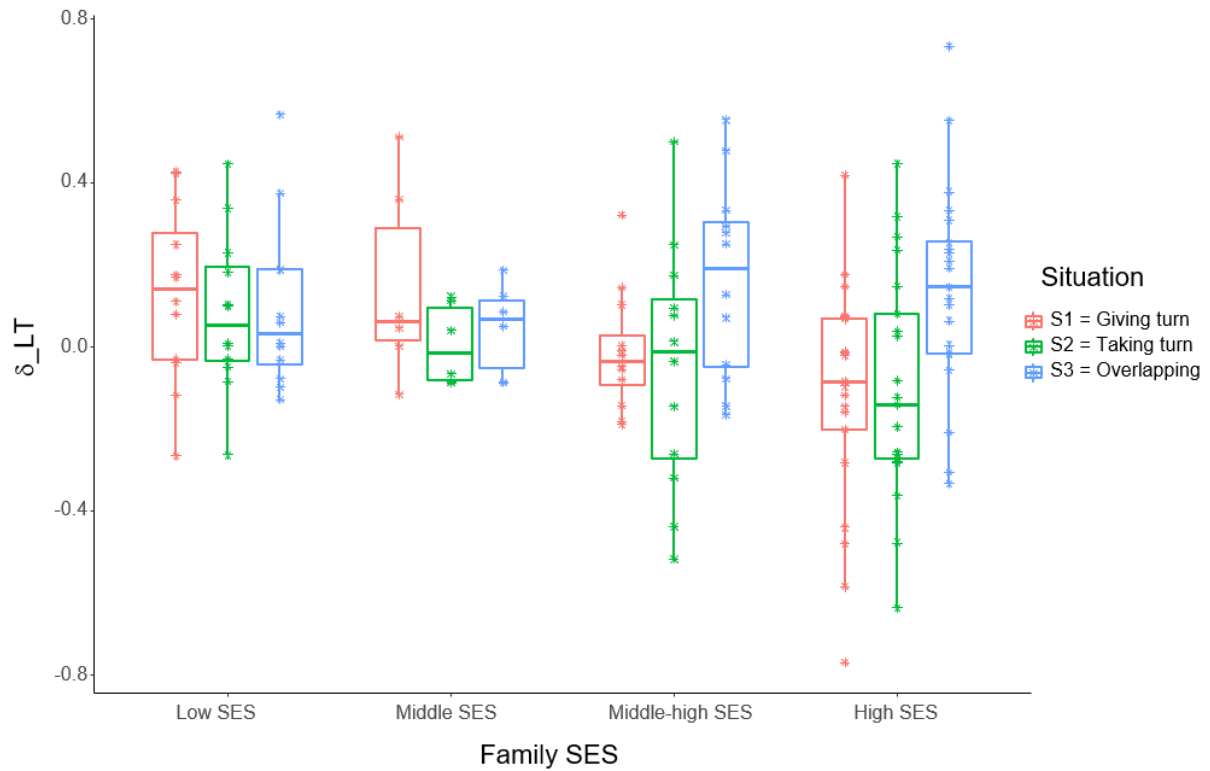


Fig. 2: Difference of looking time on the 2nd speaker between before and after the sketch (δ_{LT}) in relation to the situation and the family SES. Crosses = single data points; dots = estimated marginal means for each category; bars = 95% confidence intervals (package emmeans (Lenth, 2019) in R 3.5.0 (2018); S1 (red) = Turn-giving; S2 (green) = Turn-taking; S3 (blue) = Overlapping.

Table S2: Post-hoc multiple comparisons for the interaction effect between the situation seen in a sketch and the family SES on the difference of looking time on the 2nd speaker after and before the sketch. S1 = Turn-giving; S2 = Turn-taking; S3 = Overlapping; Low SES = both parents with high-school level or less; Middle SES = at least one parent with bachelor degree; Middle-high SES = one parent with a postgraduate degree; High SES = both parents with a postgraduate degree. Significant differences are indicated by * for $p < 0.05$ and by ** for $p < 0.01$ (package emmeans (Lenth, 2019) in R 3.5.0 (2018)).

Comparisons	t.ratio	df	p.value
<i>High SES</i>			
S1D vs S2D	0.765	80.1	0.670
S1D vs S3D	-3.342	80.4	0.008*
S2D vs S3D	-4.114	80.6	0.001*
<i>Middle-high SES</i>			
S1C vs S2C	1.801	80.6	0.227
S1C vs S3C	-1.045	81.1	0.598
S2C vs S3C	-2.788	81	0.026*
<i>Middle SES</i>			
S1B vs S2B	0.557	79.8	0.772
S1B vs S3B	0.435	79.4	0.797
S2B vs S3B	-0.120	79.9	0.905
<i>Low SES</i>			
S1A vs S2A	0.870	79.9	0.663
S1A vs S3A	1.160	81.8	0.598
S2A vs S3A	0.281	81.4	0.851

Second, whether infants had siblings or not influenced their responses to the different situations. The δ_LT varied significantly between the three situations depending on the presence or absence of sibling(s) (LMM, Situation x Sibling: $\chi^2 = 9.21$, $df = 2$, $p = 0.01$; Table S1). Post-hoc comparisons evidenced that only singletons had different δ_LT according to the situation (Fig.3, Table S3). Indeed, for these infants, δ_LT differed significantly between S2 (turn-taking) on one side and S1 (turn-giving; $t.ratio = 2.88$, $df = 80.5$, $p = 0.015$) and S3 (overlapping; $t.ratio = -3.13$, $df = 81.2$, $p = 0.014$) on the other side. Singletons looked longer at the second speaker after her speech overlapped with the first speaker. They also looked longer at her after she was given the turn by the first speaker. When speakers spoke one after the other without

any interaction, singletons did not change their looking time on the second speaker. Infants with sibling(s) did not show different looking reactions to the three situations.

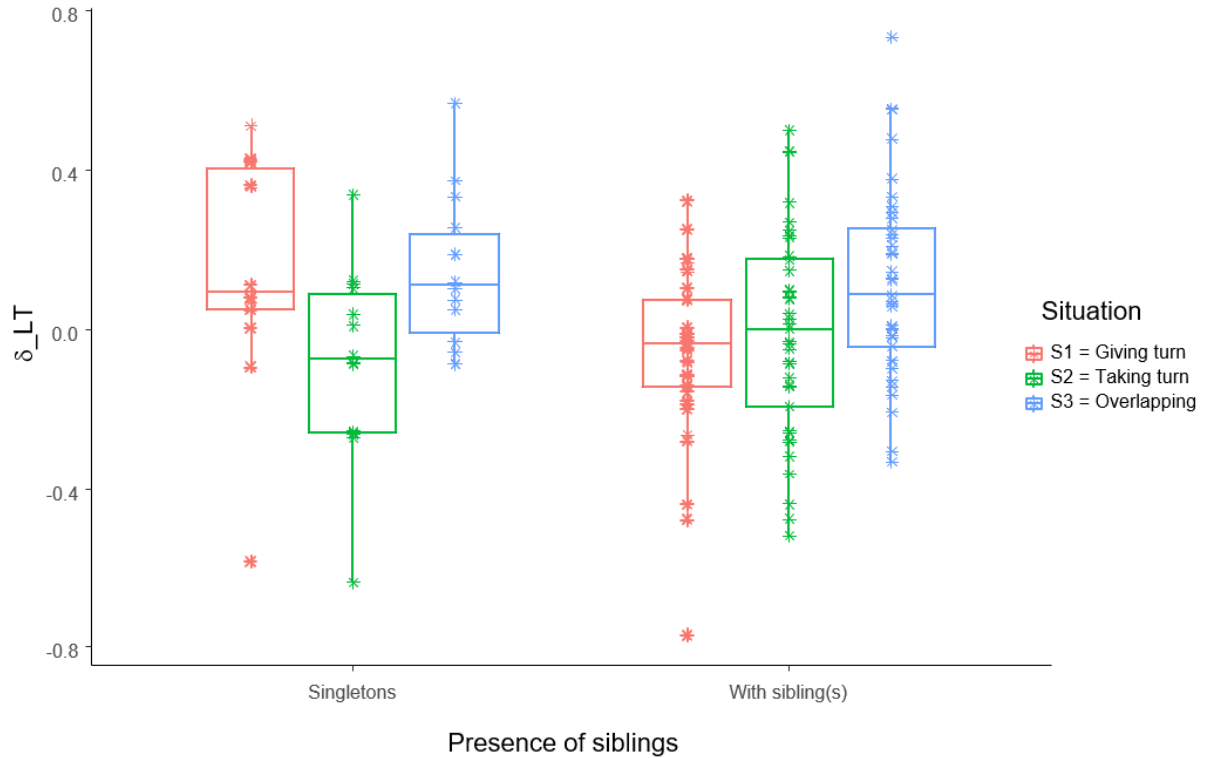


Fig. 3: Difference of looking time on the 2nd speaker between before and after the sketch in relation to the situation and the presence of sibling(s) in the infant’s family. Crosses = single data points; dots = estimated marginal means for each category; bars = 95% confidence intervals (package emmeans (Lenth, 2019) in R 3.5.0 (2018); S1 (red) = Turn-giving; S2 (green) = Turn-taking; S3 (blue) = Overlapping

Table S3: Post-hoc multiple comparisons for the interaction effect between the situation seen in a sketch and the presence of siblings in the family on the difference of looking time on the 2nd speaker after and before the sketch. S1 = Turn-giving turn; S2 = Turn-taking turn; S3 = Overlapping; Singletons = absence of sibling in the family at the time of experiment; With sibling(s) = presence of sibling(s) in the family. Significant differences are indicated by * for $p < 0.05$ (package emmeans (Lenth, 2019) in R 3.5.0 (2018)).

Comparisons	t.ratio	df	p.value
<i>Singletons</i>			
S1 vs S2	2.881	80.5	0.015 *
S1 vs S3	-0.331	81.9	0.770
S2 vs S3	-3.133	81.2	0.014 *
<i>With sibling(s)</i>			
S1 vs S2	-0.905	79.3	0.553
S1 vs S3	-1.197	79.2	0.470
S2 vs S3	-0.293	79.7	0.770

None of the other experimental or socio-demographic factors tested had a significant effect when considered alone (LMM, Sex: $\chi^2 = 0.37$, $df = 1$, $p = 0.54$; Family SES: $\chi^2 = 5.6$, $df = 3$, $p = 0.13$; Siblings: $\chi^2 = 1.44$, $df = 1$, $p = 0.23$; Situations' Order: $\chi^2 = 1.26$, $df = 2$, $p = 0.53$, Pair of women: $\chi^2 = 3.18$, $df = 2$, $p = 0.2$; Position of the 2nd speaker: $\chi^2 = 1.93$, $df = 1$, $p = 0.16$). It should be noted that the infant's sex did not influence the looking time toward the 2nd speaker on the pictures whether or not we took the Situation into account. Infants' sex did not influence their reaction to turn-taking violation in our study.

Thus, some infants reacted differently in relation to the compliancy or not of the 2nd speaker with a major conversation rule, i.e. turn-taking, during the sketch. The attention (looking time) toward a speaker depended on her behavior during the sketch but also on some sociodemographic characteristics of the infants' family.

Discussion

Our study demonstrated that 6-month-old infants are sensitive to turn-taking violation, and that this sensitivity depends on family SES and presence of sibling, but not on sex. On one hand, high and middle-high SES infants looked more at the 2nd speaker after the overlapping situation than after the two other situations (for high SES infants) or just after the turn-taking situation

(for middle-high SES). Low and middle-low SES did not show any differences between situations. On the other hand, singletons looked more at the second speaker after overlapping and turn-giving situations compared to turn-taking situation, whereas infants with sibling(s) did not.

We found no difference between boys and girls in their looking time across situations, which was somewhat surprising given the girls' advantage in language production and perception observed later in development. This is probably due to the fact that sex generally has a moderate impact, and often interacts with other factors. Indeed, Barbu et al. (2015) showed that, in children aged 2 to 6 years old, low SES boys' performances were lower than low SES girls' performances, but that there was no differences between boys and girls in high-SES. Sex can also interact with birth order, as only older brothers seem to influence negatively language skills of later-born children (Havron et al., 2019). Unfortunately, such interactions could not be tested with our sample size, emphasizing the crucial need of very large samples to investigate the many layers of variability factors and their relationships.

However, sibling(s) and family SES influenced the measured responses in the same direction than in the language development literature. Indeed, it is well known that language production and perception is delayed in children with older sibling(s) or from the lowest family SES (Fernald et al., 2013; Hart and Risley, 2003; Hoff, 2003; Huttenlocher et al., 2010; Rowe, 2008). Here, infants with sibling(s) or from lowest family SES were not sensitive to turn-taking violation. They may not have developed this conversational skill yet. Furthermore, our results highlight that variability is present much earlier than the first language productions, confirming the developmental precedence of perception over production.

The influence of sibling(s) on language development is probably due to the fact that siblings have to compete to get parents' attention. Indeed, mothers of siblings produce the same amount of utterances during dyadic interactions with one of their children than during triadic interactions with their two children (Jones and Adamson, 1987). The direct consequence is that children with sibling(s) have less focused interactions with their parents than singletons. In addition to a decrease in the quantity of input, the family composition also impact its quality, as in triadic interactions with two siblings the mothers' speech is more focused around the children's activities, whereas it is more about language itself in dyadic interactions (Oshima-Takane and Robbins, 2003). Within a sibling group, first-born children accumulate more focused interactions with their parents compared to later-born children, because they were singletons during a certain amount of time. This can explain why first-born children are ahead in vocabulary and grammar compared to later-born children at the same age (Hoff-Ginsberg,

1998). In contrast, later-born children develop more interactive skills: they are more likely to produce contingent conversational responses (Hoff-Ginsberg, 1998). Interestingly, we showed that, at 6 months, singletons higher sensitivity to conversational rules than same-aged infants with sibling(s). This can be explained by differences in quality and quantity of social interactions they have with their parents, and also by the sources of conversational input for the later-born infants: the sibling as well as the parents. Despite being older than the later-born infant, siblings are likely to be young children who have not yet fully mastered turn-taking as the adults. There could be more interruptions in households with siblings than in households with singletons, leading later-born to consider the overlapping situation as any other type of interactions at 6 months.

According to Hoff (2003), the influence of SES on early language development could be primarily due to the motivation of high SES parents to talk with their children, rather than the richness of the vocabulary they use in these interactions. Indeed, high SES mothers talk more often to their infants and more with the purpose of eliciting and maintaining conversations through topic-continuing replies than low SES mothers, who talk less often and more in order to orient their infants' behavior (Hoff, 2006). The parents' SES is therefore also a good predictor of purely conversational skills related to turn-taking, which could explain the influence found in our results well before language mastering. Beyond these conversational skills, SES also influences parents' ability to get their infants' attention through what is usually called motherese, parentese, or Infant/Child Directed Speech (IDS or CDS). IDS/CDS is indeed a high-pitched speech employed by adult towards infants/children, with short utterances, exaggerated intonations and more repetitions. This speech style is preferred by children (The ManyBabies Consortium, 2020) and has a key role in language development since it predicts vocabulary acquisition while overheard speech (Adult Directed Speech listen by children) do not (Shneidman et al., 2013). Above all, IDS/CDS vary according to the parents' SES, with higher-SES using it more than the lower-SES (Schwab and Lew-Williams, 2016). Ramírez et al. (2020) even demonstrated that increasing parent's IDS through coaching interventions enhanced the amount of turn-taking between parents and infants as early as 6 months of age. As IDS can induce more turn-taking production, it could then explain the difference observed here at the same age in turn-taking perception.

The disparity in interactions directed toward infants between families is arguably the primary source of inequality, and undoubtedly one of the most important if cumulative effects are considered. Of course, the developmental pathway of language is quite robust, and ethnographic data reporting language development in communities where parents talk relatively little to their

children might lead to a more nuanced picture. Nevertheless, these studies also indicate that even when this kind of interactions are rare, parents remain the primary sources of input compared to other adults or children (Cristia et al., 2019), and the types of words used during these child directed speech predict later vocabulary better than overheard interactions (Shneidman and Goldin-Meadow, 2012). In WEIRD (Western, Educated, Industrialized, Rich, Democratic) societies, a delay or lack of conversational skills compared to other children is problematic for social integration with peers (making friends), for later academic success, and thus for the general well-being of children and their parents. Therefore, considering a greater diversity in large samples is crucial both from a theoretical and an applied point of view. Only large and diverse samples can allow us to analyze various factors of variability and to understand the mechanisms underlying language acquisition. By revealing differences notably due to family characteristics as early as 6 months in a language development milestone, our study, like many others, stresses the need to consider sociodemographic factors in infant studies (Golinkoff et al., 2018).

Acknowledgements

We would like to thank the 6 young women who agreed to perform in the conversation sketches. We would also like to extend our warmest thanks to the many babies and their parents who came to the lab to take part in our experiment.

References

- Augusti, E., Melinder, A., Gredebäck, G., 2010. Look Who's Talking: Pre-Verbal Infants' Perception of Face-to-Face and Back-to-Back Social Interactions. *Front Psychol.* 1, 161. <https://doi.org/10.3389/fpsyg.2010.00161>
- Bakker, M., Kochukhova, O., von Hofsten, C., 2011. Development of social perception: A conversation study of 6-, 12- and 36-month-old children. *Infant Behav. Dev.* 34, 363–370. <https://doi.org/10.1016/j.infbeh.2011.03.001>
- Barbu, S., Nardy, A., Chevrot, J.-P., Guellai, B., Glas, L., Juhel, J., Lemasson, A., 2015. Sex Differences in Language Across Early Childhood: Family Socioeconomic Status does not Impact Boys and Girls Equally. *Front. Psychol.* 6. <https://doi.org/10.3389/fpsyg.2015.01874>
- Berglund, E., Eriksson, M., Westerlund, M., 2005. Communicative skills in relation to gender, birth order, childcare and socioeconomic status in 18-month-old children. *Scand. J. Psychol.* 46, 485–491. <https://doi.org/10.1111/j.1467-9450.2005.00480.x>
- Bigelow, A.E., 1998. Infants' sensitivity to familiar imperfect contingencies in social interaction. *Infant Behav. Dev.* 21, 149–162. [https://doi.org/10.1016/S0163-6383\(98\)90060-1](https://doi.org/10.1016/S0163-6383(98)90060-1)
- Black, B., Logan, A., 1995. Links between Communication Patterns in Mother-Child, Father-Child, and Child-Peer Interactions and Children's Social Status. *Child Dev.* 66, 255. <https://doi.org/10.2307/1131204>

- Bloom, K., 1975. Social elicitation of infant vocal behavior. *J. Exp. Child Psychol.* 20, 51–58. [https://doi.org/10.1016/0022-0965\(75\)90025-9](https://doi.org/10.1016/0022-0965(75)90025-9)
- Cristia, A., Dupoux, E., Gurven, M., Stieglitz, J., 2019. Child-Directed Speech Is Infrequent in a Forager-Farmer Population: A Time Allocation Study. *Child Dev.* 90, 759–773. <https://doi.org/10.1111/cdev.12974>
- Dunn, J., Shatz, M., 1989. Becoming a Conversationalist despite (Or Because of) Having an Older Sibling. *Child Dev.* 60, 399–410. <https://doi.org/10.2307/1130985>
- Eriksson, M., Marschik, P., Tulviste, T., Almgren, M., Pérez-Pereira, M., Wehberg, S., Umek, L., Gayraud, F., Kovačević, M., Gallego, C., 2012. Differences between girls and boys in emerging language skills: Evidence from 10 language communities. *Br. J. Dev. Psychol.* 30, 326–43. <https://doi.org/10.1111/j.2044-835X.2011.02042.x>
- Fernald, A., Marchman, V.A., Weisleder, A., 2013. SES differences in language processing skill and vocabulary are evident at 18 months. *Dev. Sci.* 16, 234–248. <https://doi.org/10.1111/desc.12019>
- Galsworthy, M.J., Dionne, G., Dale, P.S., Plomin, R., 2000. Sex differences in early verbal and non-verbal cognitive development. *Dev. Sci.* 3, 206–215. <https://doi.org/10.1111/1467-7687.00114>
- Ginsburg, G.P., Kilbourne, B.K., 1988. Emergence of vocal alternation in mother-infant interchanges. *J. Child Lang.* 15, 221–235. <https://doi.org/10.1017/S0305000900012344>
- Goldstein, M.H., King, A.P., West, M.J., 2003. Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proc. Natl. Acad. Sci.* 100, 8030–8035. <https://doi.org/10.1073/pnas.1332441100>
- Golinkoff, R.M., Hoff, E., Rowe, M.L., Tamis-LeMonda, C., Hirsh-Pasek, K., 2018. Talking with children matters: Defending the 30 million word gap. Brookings. URL <https://www.brookings.edu/blog/education-plus-development/2018/05/21/defending-the-30-million-word-gap-disadvantaged-children-dont-hear-enough-child-directed-words/> (accessed 3.15.23).
- Gratier, M., Devouche, E., Guellai, B., Infanti, R., Yilmaz, E., Parlato, E., 2015. Early development of turn-taking in vocal interaction between mothers and infants. *Front. Psychol.* 6. <https://doi.org/10.3389/fpsyg.2015.01167>
- Hart, B., Risley, T.R., 2003. The Early Catastrophe. The 30 Million Word Gap by Age 3. *Am. Educ.* 27, 4–9.
- Havron, N., Ramus, F., Heude, Barbara, Forhan, Anne, Cristia, A., Peyre, H., the EDEN Mother-Child Cohort Study Group, Annesi-Maesano, I., Bernard, J.Y., Botton, J., Charles, M.A., Dargent-Molina, P., de Lauzon-Guillain, B., Ducimetière, P., De Agostini, M., Foliguet, B., Forhan, A., Fritel, X., Germa, A., Goua, V., Hankard, R., Heude, B., Kaminski, M., Larroque, B., Lelong, N., Lepeule, J., Magnin, G., Marchand, L., Nabet, C., Pierre, F., Slama, R., Saurel-Cubizolles, M.J., Schweitzer, M., Thiebaugeorges, O., 2019. The Effect of Older Siblings on Language Development as a Function of Age Difference and Sex. *Psychol. Sci.* 30, 1333–1343. <https://doi.org/10.1177/0956797619861436>
- Hazen, N.L., Black, B., 1989. Preschool Peer Communication Skills: The Role of Social Status and Interaction Context. *Child Dev.* 60, 867–876. <https://doi.org/10.2307/1131028>
- Hilbrink, E.E., Gattis, M., Levinson, S.C., 2015. Early developmental changes in the timing of turn-taking: a longitudinal study of mother–infant interaction. *Front. Psychol.* 6. <https://doi.org/10.3389/fpsyg.2015.01492>
- Hoff, E., 2006. How social contexts support and shape language development☆. *Dev. Rev.* 26, 55–88. <https://doi.org/10.1016/j.dr.2005.11.002>

- Hoff, E., 2003. The Specificity of Environmental Influence: Socioeconomic Status Affects Early Vocabulary Development Via Maternal Speech. *Child Dev.* 74, 1368–1378. <https://doi.org/10.1111/1467-8624.00612>
- Hoff-Ginsberg, E., 1998. The relation of birth order and socioeconomic status to children's language experience and language development. *Appl. Psycholinguist.* 19, 603–629. <https://doi.org/10.1017/S0142716400010389>
- Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., Hedges, L.V., 2010. Sources of Variability in Children's Language Growth. *Cognit. Psychol.* 61, 343–365. <https://doi.org/10.1016/j.cogpsych.2010.08.002>
- Jones, C.P., Adamson, L.B., 1987. Language Use in Mother-Child and Mother-Child-Sibling Interactions 12.
- Kuhl, P.K., Tsao, F.-M., Liu, H.-M., 2003. Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proc. Natl. Acad. Sci.* 100, 9096–9101. <https://doi.org/10.1073/pnas.1532872100>
- Levinson, S.C., 2016. Turn-taking in Human Communication – Origins and Implications for Language Processing. *Trends Cogn. Sci.* 20, 6–14. <https://doi.org/10.1016/j.tics.2015.10.010>
- Murray, L., 1985. Emotional regulations of interactions between two-month-olds and their mothers. *Soc. Percept. Infants* 177–197.
- Oakes, L.M., 2017. Sample Size, Statistical Power, and False Conclusions in Infant Looking-Time Research. *Infancy* 22, 436–469. <https://doi.org/10.1111/infa.12186>
- Oshima-Takane, Y., Robbins, M., 2003. Linguistic Environment of Secondborn Children. *First Lang.* 23, 21–40. <https://doi.org/10.1177/0142723703023001002>
- Pika, S., Wilkinson, R., Kendrick, K.H., Vernes, S.C., 2018. Taking turns: bridging the gap between human and animal communication. *Proc. R. Soc. B Biol. Sci.* 285, 20180598. <https://doi.org/10.1098/rspb.2018.0598>
- Pine, J.M., 1995. Variation in Vocabulary Development as a Function of Birth Order. *Child Dev.* 66, 272. <https://doi.org/10.2307/1131205>
- Ramírez, N.F., Lytle, S.R., Kuhl, P.K., 2020. Parent coaching increases conversational turns and advances infant language development. *Proc. Natl. Acad. Sci.* 117, 3484–3491. <https://doi.org/10.1073/pnas.1921653117>
- Robinson, L.F., Reis, H.T., 1989. The effects of interruption, gender, and status on interpersonal perceptions. *J. Nonverbal Behav.* 13, 141–153. <https://doi.org/10.1007/BF00987046>
- Rowe, M.L., 2008. Child-directed speech: relation to socioeconomic status, knowledge of child development and child vocabulary skill*. *J. Child Lang.* 35, 185–205. <https://doi.org/10.1017/S0305000907008343>
- Sacks, H., Schegloff, E.A., Jefferson, G., 1974. A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language* 50, 696–735. <https://doi.org/10.2307/412243>
- Schwab, J.F., Lew-Williams, C., 2016. Language learning, socioeconomic status, and child-directed speech. *WIREs Cogn. Sci.* 7, 264–275. <https://doi.org/10.1002/wcs.1393>
- Shneidman, L.A., Arroyo, M.E., Levine, S.C., Goldin-Meadow, S., 2013. What counts as effective input for word learning? *J. Child Lang.* 40, 672–686. <https://doi.org/10.1017/S0305000912000141>
- Shneidman, L.A., Goldin-Meadow, S., 2012. Language input and acquisition in a Mayan village: how important is directed speech? *Dev. Sci.* 15, 659–673. <https://doi.org/10.1111/j.1467-7687.2012.01168.x>
- Stern, D.N., Jaffe, J., Beebe, B., Bennett, S.L., 1975. Vocalizing in unison and in alternation: Two modes of communication within the mother- infant dyad. *Ann. N. Y. Acad. Sci.* 263, 89–100. <https://doi.org/10.1111/j.1749-6632.1975.tb41574.x>

- Stivers, T., Enfield, N.J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., Ruiter, J.P. de, Yoon, K.-E., Levinson, S.C., 2009. Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci.* 106, 10587–10592. <https://doi.org/10.1073/pnas.0903616106>
- ter Maat, M., Truong, K.P., Heylen, D., 2010. How Turn-Taking Strategies Influence Users' Impressions of an Agent, in: Allbeck, J., Badler, N., Bickmore, T., Pelachaud, C., Safonova, A. (Eds.), *Intelligent Virtual Agents, Lecture Notes in Computer Science*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 441–453. https://doi.org/10.1007/978-3-642-15892-6_48
- The ManyBabies Consortium, 2020. Quantifying Sources of Variability in Infancy Research Using the Infant-Directed-Speech Preference: *Adv. Methods Pract. Psychol. Sci.* 3, 24–52. <https://doi.org/10.1177/2515245919900809>
- The ManyBabies Consortium, Frank, M.C., Alcock, K.J., Arias-Trejo, N., Aschersleben, G., Baldwin, D., Barbu, S., Bergelson, E., Bergmann, C., Black, A.K., Blything, R., Böhlend, M.P., Bolitho, P., Borovsky, A., Brady, S.M., Braun, B., Brown, A., Byers-Heinlein, K., Campbell, L.E., Cashon, C., Choi, M., Christodoulou, J., Cirelli, L.K., Conte, S., Cordes, S., Cox, C., Cristia, A., Cusack, R., Davies, C., de Klerk, M., Delle Luche, C., Ruiter, L. de, Dinakar, D., Dixon, K.C., Durier, V., Durrant, S., Fennell, C., Ferguson, B., Ferry, A., Fikkert, P., Flanagan, T., Floccia, C., Foley, M., Fritzsche, T., Frost, R.L.A., Gampe, A., Gervain, J., Gonzalez-Gomez, N., Gupta, A., Hahn, L.E., Kiley Hamlin, J., Hannon, E.E., Havron, N., Hay, J., Hernik, M., Höhle, B., Houston, D.M., Howard, L.H., Ishikawa, M., Itakura, S., Jackson, I., Jakobsen, K.V., Jarto, M., Johnson, S.P., Junge, C., Karadag, D., Kartushina, N., Kellier, D.J., Keren-Portnoy, T., Klassen, K., Kline, M., Ko, E.-S., Kominsky, J.F., Kosie, J.E., Kragness, H.E., Krieger, A.A.R., Krieger, F., Lany, J., Lazo, R.J., Lee, M., Leservoiesier, C., Levelt, C., Lew-Williams, C., Lippold, M., Liszkowski, U., Liu, L., Luke, S.G., Lundwall, R.A., Macchi Cassia, V., Mani, N., Marino, C., Martin, A., Mastroberardino, M., Mateu, V., Mayor, J., Menn, K., Michel, C., Moriguchi, Y., Morris, B., Nave, K.M., Nazzi, T., Noble, C., Novack, M.A., Olesen, N.M., John Orena, A., Ota, M., Panneton, R., Esfahani, S.P., Paulus, M., Pletti, C., Polka, L., Potter, C., Rabagliati, H., Ramachandran, S., Rennels, J.L., Reynolds, G.D., Roth, K.C., Rothwell, C., Rubez, D., Ryjova, Y., Saffran, J., Sato, A., Savelkoul, S., Schachner, A., Schafer, G., Schreiner, M.S., Seidl, A., Shukla, M., Simpson, E.A., Singh, L., Skarabela, B., Soley, G., Sundara, M., Theakston, A., Thompson, A., Trainor, L.J., Trehub, S.E., Trøan, A.S., Tsui, A.S.-M., Twomey, K., Von Holzen, K., Wang, Y., Waxman, S., Werker, J.F., Wermelinger, S., Woolard, A., Yurovsky, D., Zahner, K., Zettersten, M., Soderstrom, M., 2020. Quantifying Sources of Variability in Infancy Research Using the Infant-Directed-Speech Preference. *Adv. Methods Pract. Psychol. Sci.* 3, 24–52. <https://doi.org/10.1177/2515245919900809>
- Zambrana, I.M., Ystrom, E., Pons, F., 2012. Impact of Gender, Maternal Education, and Birth Order on the Development of Language Comprehension: A Longitudinal Study from 18 to 36 Months of Age. *J. Dev. Behav. Pediatr.* 33, 146–155. <https://doi.org/10.1097/DBP.0b013e31823d4f83>

Article 2 :

Often overlooked, experience at daycare influences the perception of conversational turn-taking in pre-verbal infants.

Bastien Meunier¹, Maéva Thomazeau¹, Justine Thobois¹, Alban Lemasson^{1,2}, Virginie Durier^{1*}

¹ Univ Rennes, CNRS, Normandie Univ, EthoS (Éthologie animale et humaine) - UMR 6552, Rennes, France

² Institut Universitaire de France, 1 rue Descartes, 75231 Paris, Cedex 05, France

* co-last authors

Synthèse de l'article 2

Contexte

Avant de savoir parler, les nourrissons préverbaux interagissent vocalement avec leurs parents dans ce qu'on appelle des proto-conversations. Au cours de ces dernières, les nourrissons doivent apprendre à coordonner leurs vocalisations avec les productions de leurs parents pour respecter un tour de rôle. Cela nécessite des compétences de perception de patterns conversationnels qui sont encore mal connues par rapport aux compétences liées à la production. Ces compétences s'acquièrent en fonction de l'exposition du nourrisson, exposition qui varie d'un foyer à l'autre en fonction de facteurs socio-démographiques.

Méthode

Nous avons donc étudié la perception du tour de parole chez des nourrissons préverbaux pendant la présentation vidéo de trois situations conversationnelles qui différaient par leur transition de tour : une situation où la locutrice A donne explicitement la parole à la locutrice B (passage de parole), une situation où la locutrice B prend elle-même la parole juste après la locutrice A (tour de parole), et une situation où la locutrice B coupe la parole de la locutrice A (coupure de parole). Nous avons mesuré les regards selon ces situations conversationnelles chez 103 nourrissons de 3 à 11 mois pour étudier l'influence de l'âge sur ces compétences de perception. Nous avons aussi étudié l'influence de plusieurs facteurs connus comme le statut socio-économique, la présence d'une fratrie, le sexe du nourrisson testé ainsi qu'un facteur moins connu qui est le mode de garde du nourrisson.

Résultats

Nos résultats ont montré que les nourrissons de 3 à 11 mois étaient sensibles aux différences de transition de tours. Cette sensibilité aux transitions de tours dépendait du mode de garde du nourrisson : seuls les nourrissons gardés par leurs parents ont réagi différemment aux situations conversationnelles présentées. De plus, les nourrissons gardés en crèche ont plus regardé la locutrice après la situation de tour de parole que les nourrissons gardés par leurs parents. Cette différence pourrait s'expliquer par l'exposition différentes de ces nourrissons, les nourrissons en crèche étant confrontés à des interactions très différentes des nourrissons chez leurs parents. Nous n'avons pas trouvé d'influence des autres facteurs socio-démographiques étudiés, ni même de l'âge.

Often overlooked, experience at daycare influences the perception of conversational turn-taking in pre-verbal infants.

Bastien Meunier¹, Maéva Thomazeau¹, Justine Thobois¹, Alban Lemasson^{1,2*}, Virginie Durier^{1*}

¹ Univ Rennes, CNRS, Normandie Univ, EthoS (Éthologie animale et humaine) - UMR 6552, Rennes, France

² Institut Universitaire de France, 1 rue Descartes, 75231 Paris, Cedex 05, France

* co-last authors

Abstract

Before they are able to speak, preverbal infants interact vocally with their parents in so-called protoconversations, during which they learn to respect a turn-taking pattern. While the skills involved in producing turn-taking vocalizations are well known, the development of skills in turn-taking perception in preverbal infants is less understood. We therefore studied the looking times of 103 infants aged between 3 and 11 months when presented with a video of three conversational situations in which turn transitions were different (turn-giving, turn-taking, overlapping). We considered socio-demographic factors such as the family's socio-economic status, the presence of siblings, daycare and the infant's sex, in addition to the infant's age. Our results revealed that infants' sensitivity to turn transitions depended on their daycare. Indeed, only infants cared for by their parents reacted differently to different conversational situations. Moreover, infants cared for in day nurseries showed greater visual interest after the turn-taking situation than infants cared for by their parents, probably because they are exposed to very different interactions on a daily basis. We didn't find any influence of age or other socio-demographic factors.

Introduction

Language development has always fascinated by the apparent simplicity with which children learn the mother language "from scratch", as well as by the robustness of the process. The most important stages take place on a very short timescale in the first year, when infants must learn to combine perceptual, computational and social skills. Infants begin by learning to categorize the sounds they hear in their environment on the basis of regularities (called statistical learning), and these first learned patterns shape future perception and learning (Kuhl, 2004). The first year is therefore of paramount importance in language development, with exposure playing a key role (Hoff, 2006).

Not every exposure is relevant to an infant's language acquisition. The speech must come from a person physically present in live interaction with the infant (Kuhl et al., 2003), and must respond contingently with his or her own vocalizations (Goldstein et al., 2003). The infant-directed speech generally adapted to get the infant's attention (shorter utterances, lots of repetition, a higher-pitched voice, better articulated words) is preferred over adult-directed speech (The ManyBabies Consortium, 2020), and exposure to the former predicts vocabulary better than the latter (Shneidman et al., 2013). This exposure differs from one home to another, and this diversity can be embraced by looking at several socio-demographic factors. One of the most well-known factors is the family socio-economic status: children from the highest socio-economic status families master language better than same-aged children from less privileged families (Fernald et al., 2013; Hart and Risley, 2003; Hoff, 2003; Huttenlocher et al., 2010; Rowe, 2008). Another well-known factor is the presence of siblings. Later-borns have indeed less native language proficiency at the same age than first-borns (Berglund et al., 2005; Dunn, 2023; Hoff-Ginsberg, 1998; Pine, 1995; Zambrana et al., 2012). Yet, some factors related to the infants' social environment have been little studied, and so their influence on language development is largely unknown. Child care, for instance, could have a significant impact, since the infants' social environment of reference can vary depending on whether they spend a large part of their day with a multitude of peers and few adults in a day nursery, or alone with their parents or a child minder. However, the few studies that have been carried out present have found no strong differences in terms of language mastery between children cared for in day nurseries and those cared for by their parents (National Institute of Child Health, 2000; Peisner-Feinberg et al., 2001).

These early influences of the social environment on language development are highlighted after the first year, when children already have some proficiency in language. But infants develop

the conversational skills that precede and enable language acquisition long before. One such ability is turn-taking, a major feature of adult conversations: a rapid exchange of turns between different interlocutors, who themselves choose when to intervene, avoiding overlapping words or long gaps in the speech flow (Sacks et al., 1974). This pattern of interaction is present in all cultures (Stivers et al., 2009) and appears very early on: infants indeed respond vocally in alternation with their parents from their very first months (Bloom, 1975; Ginsburg and Kilbourne, 1988; Murray, 1985; Stern et al., 1975). Studies even highlighted this pattern of interaction in full-term and premature newborns (Caskey et al., 2014; Dominguez et al., 2016). In addition, infants early improve their innate ability to vocalize in turn. Whereas at 1 month they cannot distinguish between contingent and non-contingent interactions, at 3 months they respond more positively to contingent interactions (Striano et al., 2005). After 3 months, this increase in contingency sensitivity is confirmed by the observed decline in the amount of overlap during interactions (Ginsburg and Kilbourne, 1988; Hilbrink et al., 2015). However, the performance of infants remains below that of adults, as their average response time is longer (Hilbrink et al., 2015). Indeed, while adult and child speakers can rely on a combination of lexicosyntactic and prosodic cues to predict turn endings and intervene quickly (Casillas and Frank, 2017), infants may struggle to do this with non-verbal cues alone (silences, gazes, postures, gestures, prosody). It is therefore crucial to study the developmental dynamics of preverbal infants' perception of turn-taking, in particular by testing several transition styles to understand their representation of appropriate interaction. This perception may depend on the infants' experience: Bigelow (1998) indeed showed that 5-month-old infants already preferred their mother's level of temporal contingency to that of other people, even when the former was lower than the norm.

We aimed to investigate whether preverbal infants were sensitive to differences in turn transitions, to study the developmental dynamics of this perception, and to assess the influences of sociodemographic factors that could impact exposure to conversations. We have therefore set up an experiment consisting in presenting three conversational situations on video to infants from 3 to 11 months of age, and measuring their looking times to revealed perception differences. The conversational situations differed by the pattern of turn transition: a turn-giving pattern where the current speaker gives explicitly the turn to the next one, a turn-taking pattern where the second speaker starts speaking after the first one stops, and an overlapping pattern where the second speaker interrupts the first one. We studied the infants' perception to these situations in relation to age as well as several socio-demographic factors: the family SES, the presence of sibling, the infants' daycare and the infant sex.

Methods

Participants

Infants between 3 and 11 months-old participated in our study (N=120). Participants were recruited by letter and could carry out the experiment either in the laboratory or at home via a computer and a webcam using the LabVanced platform. Infants were excluded when: (1) they were born prematurely or with developmental, visual and/or hearing disorders; (2) technical problems impaired the quality of the video; (3) infants did not look sufficiently at two or three sketches (>50% of looking time looking away from the screen during the sketches) or did not look at the photos (before or after each sketch). If an infant looked at the screen for less than 50% of the sketch or did not look at the photo for only one situation, only the data for this situation were removed. In total, we excluded N=17 infants. Our final data set therefore includes 103 subjects, 75 of whom performed the experiment at home.

Sociodemographic characteristics considered for the analysis were: the infant's sex, the presence of sibling in the family (presence or absence), the socioeconomic status (SES) of the family and the infant's daycare. The family SES were estimated with parent's education: both parents with high school level or less (low SES); at least one parent with bachelor degree (undergraduate) (middle SES); one parent with a postgraduate (middle-high SES); both parents with a postgraduate degree (high SES). Due to a strong recruitment bias, we had to combine the low and middle SES categories into a single one (middle-low SES) to have similar numbers to the other two classes. Concerning daycare, we define three types: the infant spends 3 days per week or more in day nursery (day nursery), the infant spends 3 days per week or more with a childminder (childminder), the infant spends at least 4 days per week with their parents (parents).

Infants were categorized into 3 categories according to their age: 3-5 months, 6-8 months, 9-11 months.

Procedure

Whether in the laboratory or at home online, parents were asked to sit comfortably with their infant on their lap, facing the screen. They were asked not to intervene during the video sequence other than to hold their infant. In the lab, infants were in front of a screen (24'', DELL U2410) at distance of 60cm. A computer DELL Latitude (5400-NJVXR) was connected to the screen to broadcast the experiment through the LabVanced platform (so whether they were at home or in the lab, infants and parents went through the same process). The camera (SONY

HDR-SR12) was placed above the screen to record the infant's behavior during the video and photos presentation.

The test was the same as the one performed in a previous study (Meunier et al, in submission): infants watched a video sequence of three sketches played by two women (a different pair of women per sketch), each preceded and followed by photos of the two women for 10 sec on the same side of the screen that they appeared during the sketch. The two women told short stories while watching the camera, smiling and speaking in infant directed speech. Each sketch represented a conversational situation. In the turn-giving situation (S1), the first speaker stopped speaking after she told her story, turned towards the second women and said “And you?” then the second woman turn back toward the camera and told her story. In the turn-taking situation (S2), the first woman told her story, stopped at the end, and then second woman started to speak. In the overlapping situation (S3), the second woman started telling her story during the first woman story, overlapping her for 1.4 sec on average (3 to 5 words). We created 36 different video sequences to randomize the order of the situation within the sequence, women's pairs in the situations, the role and the places of women (first or second speaker, on the right- or left-hand of the screen).

Data processing

We measured the time each infant spent looking on each woman's photo before and after each sketch with BORIS software (Friard and Gamba, 2016). M.T. coded all the videos without knowing the situation the infant was watching. J.T. coded 10 random videos already coded by M.T. (being also unaware of the situation) to calculate the inter-rater reliability with Cohen's kappa implemented in BORIS. On these 10 random videos coded by both coders, we obtained a mean $k=0,75$.

We calculated the ratio of time spent looking at the second speaker over the time spent looking at any of the speakers ($\text{ratio_LT} = \text{LT speaker 2} / (\text{LT speaker 1} + \text{LT speaker 2})$) for each display. After, we calculated the difference of these ratio between the picture displayed after the sketch and the picture displayed before (Looking Time index = $\text{ratio_LT AFTER} - \text{ratio_LT BEFORE}$). This index allowed us to measure the growing or decreasing interest of infants towards the second speaker in relation to the situation and therefore to the turn-transition pattern she used.

Statistical analysis

We ran a Linear Mixed Model (LMM) to assess the impact of several factors on the Looking Time index. Our fixed effects were the situation displayed (S1, S2, or S3), the position of the second speaker to control laterality bias (right or left), the infant's sex (girl or boy), the infant's age class (3 modalities), the family SES (3 modalities), the presence of sibling(s) (yes or no), and the infant's daycare (3 modalities). We added the interactions between the situation and the other factors. The infants' identity was the random factor. We ran our model in R software (version 4.0.5; R Core Team, 2018) with lmerTest package (Kuznetsova et al., 2017), and we performed our post-hoc tests with emmeans package (Lenth, 2022).

Results

The linear model revealed that the infants tested looked differently according to the turn-transition pattern just observed (LMM, situation: $\chi^2 = 8.170$, $df = 2$, $p = 0.017$). Indeed, the infants looked more at the second speaker after the turn-giving situation than after the turn-taking situation (S1 vs S2: $t.ratio = 2.797$, $df = 368$, $p = 0.015$). On the other hand, there were no significant differences in looking times between the overlap situation and the other two situations (S1 vs S3: $t.ratio = 1.818$, $df = 368$, $p = 0.165$; S2 vs S3: $t.ratio = -0.964$, $df = 368$, $p = 0.600$).

Of all the sociodemographic factors tested, only infant's daycare significantly influenced infants' behavior in conversational situations (LMM, situation x infant's daycare: $\chi^2 = 9.885$, $df = 4$, $p = 0.042$). Age class, infant's sex, family SES and the presence of siblings did not interact with the situation variable. Post-hoc comparisons revealed that only infants cared for by their parents reacted differently to the three sketches: they looked more at the second speaker after the turn-giving situation than after the turn-taking situation (parents, S1 vs S2: $t.ratio = 3.068$, $df = 277$, $p = 0.043$). There were no significant differences between the latter two situations and the overlapping situation (see Fig. 1). Also, for the turn-taking situation only, infants cared for in nursery looked more at the second interlocutor than infants cared for by their parents (S2, nursery vs parents: $t.ratio = 2.843$, $df = 341$, $p = 0.043$). The other comparisons were not significant.

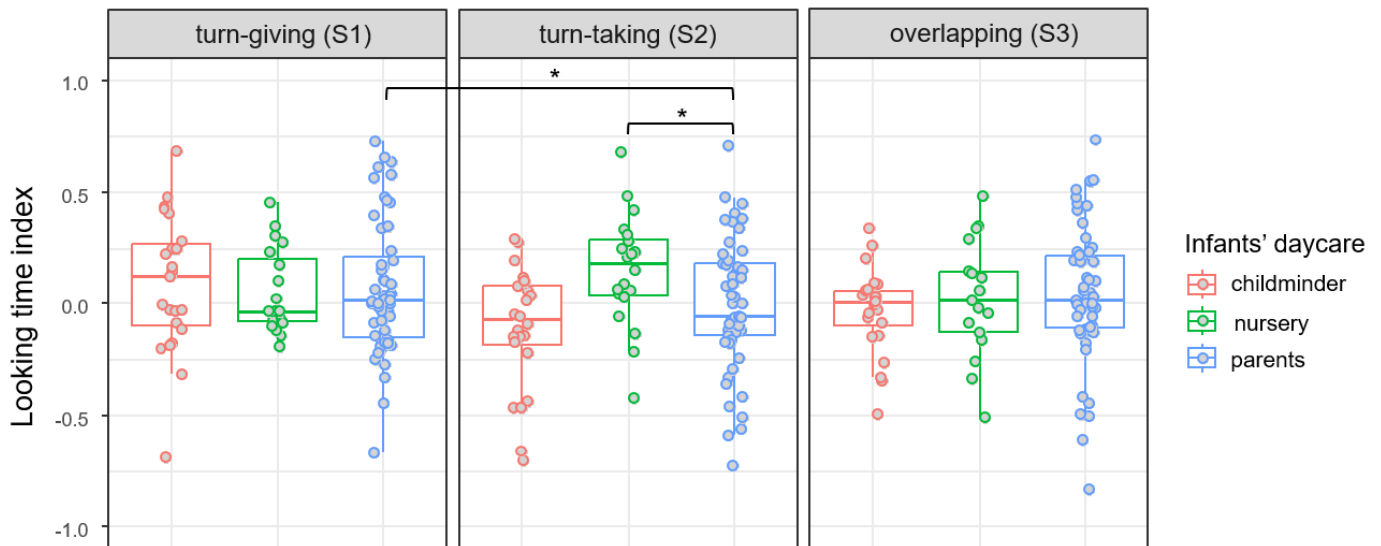


Fig 1: Difference of looking time on the 2nd speaker before and after the sketch in relation to the situation and infant's daycare. Post-hoc comparisons revealed significant differences in looking times: between turn-giving and turn-taking situations in infants cared for by parents, and between turn-taking situations in infants cared for by parents and infants cared for in nursery (* $p < 0.05$). In other post-hoc comparisons, $p > 0.05$.

Table 1: Effect of sociodemographic factors on the difference of looking time on the 2nd speaker after and before a sketch.

Fixed factors and interactions	Chi ²	df	P value
Situation	8.170	2	0.017*
Position of the 2 nd speaker	3.120	1	0.078
Infant's sex	0.000	1	0.990
Infant's age class	1.044	2	0.593
Family SES	1.048	2	0.592
Presence of sibling	1.612	1	0.204
Infant's daycare	3.156	2	0.172
Situation:Position of the 2 nd speaker	0.438	2	0.804
Situation:Infant's sex	2.651	2	0.266
Situation:Infant's age class	3.419	4	0.490
Situation:Family SES	1.925	4	0.750
Situation:Presence of sibling	0.339	2	0.844
Situation:Infant's daycare	9.885	4	0.042*

Discussion

Our study demonstrated that 3 to 11-month-olds infants are sensitive to differences in turn transitions, but not all were equally sensitive. In fact, our results showed that the determining factor in this perception was the infant's daycare, since only infants cared for by parents reacted differently to turn-giving and turn-taking situations. Moreover, infants cared for in day nursery looked more at the second speaker after the turn-taking situation than infants cared for by their parents. We found no influence of the other tested factors, such as infant sex, family SES, or the presence of siblings. We also didn't find an influence of infant's age on the perception of transitions.

Because infants cared for by their parents looked significantly more at the second speaker after the turn-giving situation than after the turn-taking situation, this means that they at least distinguished these two turn-transition patterns. Being able to distinguish turn transition patterns is of course an important step before being able to attribute a positive or negative valence to them. Indeed, studies have shown in adult that turn transition patterns can be associated with personality types and therefore drive preference in others: interrupting speech, for example, is often associated with a strong, less sociable personality (Robinson and Reis, 1989; ter Maat et al., 2010). These differences in interaction styles can have an early impact on a child's life, since children who take lengthy turns, frequently interrupt their peers and respond in a non-contingent manner are rejected by their peers (Hazen and Black, 1989). The acquisition of these important skills depends on infants' early exposure, and the link our results revealed between infants' perception of conversations and daycare suggests that the latter significantly influences this exposure. Indeed, infants cared for in a day nursery or with a childminder are usually with several of their peers, and may have fewer focused interactions with an adult than infants cared for by their parents. Yet it is precisely these interactions with adults that infants prefer, and which provide the input they need to learn language (Shneidman et al., 2013; Shneidman and Goldin-Meadow, 2012; The ManyBabies Consortium et al., 2020). The interactions an infant may have with peers in a day nursery or from the same childminder are therefore no alternative to interactions with an adult in IDS. But the adult-infant interactions in day nurseries are also different from these useful interactions. Torr and Pham (2016), by observing interactions between 9 to 20-month-olds children and educators in child care centers, found that educators interacted more with several children simultaneously than with a single one. The resulting conversations are less useful for language development, since they involve less joint attention and fewer successive turns on a single topic (Torr and Pham, 2016). This

can explain why our results showed that infants in day nurseries looked more at the second interlocutor after the turn-taking situation than infants cared for by their parents did during the same situation. Turn-taking situation with smooth transition are probably rarer in day nursery rather than with parents and home, and so infants who spend several days a week in day nurseries might show a visual preference for these more atypical transitions for them. Moreover, since it is the motivational factors underlying an interaction that play a key role in acquisition, whether these interactions are with a parent or another adult also has an influence, interactions with parents being more valuable than interactions with other adults. The instability of being transported to a different place from home every day can also affect an infant's motivation to interact and so language learning (Tran and Weinraub, 2006).

These results are somewhat surprising because they are quite different from those we obtained in a previous study with the same experimental design (Meunier et al, submitted). Indeed, we then evidenced an influence of family SES and the presence of sibling on the perception of 6-month-olds: only infants from high-SES families and singletons reacted differently to the different turn-transition patterns. Here, we didn't find such influences. The lack of a family SES effect is understandable, as we had to pool data from the previously defined low SES and middle SES categories due to a very important recruitment bias in favor of the high SES. This bias towards "convenience samples" unrepresentative of the larger population (i.e. infants from higher-SES families) is recurrent in studies on developmental psychology and is one reason for the replicability problems encountered in this field (Fernald et al., 2013). We compared the results of the 6-month-old infants in this study with those of the previous one, and found similar patterns of gaze times as a function of SES and sibling factors. This indicates that these results are reliable, but that it is the diversification of our sample over a wide age range that is confounding. Age does have an effect, even if it's not as clear as we expected with older infants being better at distinguishing turn transitions than younger infants. Age does indeed play a role in learning, and not just in maturation, with older infants having been more exposed to interactions than younger infants. As the developmental agenda for skills related to the perception of turn transitions is not yet well understood, different socio-demographic factors may play a role, but at different times. Further studies on the interaction between age and factors such as SES, the presence of siblings and childcare arrangements are therefore needed.

References

- Berglund, E., Eriksson, M., Westerlund, M., 2005. Communicative skills in relation to gender, birth order, childcare and socioeconomic status in 18-month-old children. *Scand. J. Psychol.* 46, 485–491. <https://doi.org/10.1111/j.1467-9450.2005.00480.x>
- Bigelow, A.E., 1998. Infants' sensitivity to familiar imperfect contingencies in social interaction. *Infant Behav. Dev.* 21, 149–162. [https://doi.org/10.1016/S0163-6383\(98\)90060-1](https://doi.org/10.1016/S0163-6383(98)90060-1)
- Bloom, K., 1975. Social elicitation of infant vocal behavior. *J. Exp. Child Psychol.* 20, 51–58. [https://doi.org/10.1016/0022-0965\(75\)90025-9](https://doi.org/10.1016/0022-0965(75)90025-9)
- Casillas, M., Frank, M.C., 2017. The development of children's ability to track and predict turn structure in conversation. *J. Mem. Lang.* 92, 234–253. <https://doi.org/10.1016/j.jml.2016.06.013>
- Caskey, M., Stephens, B., Tucker, R., Vohr, B., 2014. Adult Talk in the NICU With Preterm Infants and Developmental Outcomes. *Pediatrics* 133, e578–e584. <https://doi.org/10.1542/peds.2013-0104>
- Dominguez, S., Devouche, E., Apter, G., Gratier, M., 2016. The Roots of Turn-Taking in the Neonatal Period. *Infant Child Dev.* 25, 240–255. <https://doi.org/10.1002/icd.1976>
- Dunn, J., 2023. Becoming a Conversationalist despite (Or Because of) Having an Older Sibling.
- Fernald, A., Marchman, V.A., Weisleder, A., 2013. SES differences in language processing skill and vocabulary are evident at 18 months. *Dev. Sci.* 16, 234–248. <https://doi.org/10.1111/desc.12019>
- Friard, O., Gamba, M., 2016. BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. *Methods Ecol. Evol.* 7, 1325–1330. <https://doi.org/10.1111/2041-210X.12584>
- Ginsburg, G.P., Kilbourne, B.K., 1988. Emergence of vocal alternation in mother-infant interchanges. *J. Child Lang.* 15, 221–235. <https://doi.org/10.1017/S0305000900012344>
- Goldstein, M.H., King, A.P., West, M.J., 2003. Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proc. Natl. Acad. Sci.* 100, 8030–8035. <https://doi.org/10.1073/pnas.1332441100>
- Hart, B., Risley, T.R., 2003. The Early Catastrophe. The 30 Million Word Gap by Age 3. *Am. Educ.* 27, 4–9.
- Hazen, N.L., Black, B., 1989. Preschool Peer Communication Skills: The Role of Social Status and Interaction Context. *Child Dev.* 60, 867–876. <https://doi.org/10.2307/1131028>
- Hilbrink, E.E., Gattis, M., Levinson, S.C., 2015. Early developmental changes in the timing of turn-taking: a longitudinal study of mother–infant interaction. *Front. Psychol.* 6. <https://doi.org/10.3389/fpsyg.2015.01492>
- Hoff, E., 2006. How social contexts support and shape language development☆. *Dev. Rev.* 26, 55–88. <https://doi.org/10.1016/j.dr.2005.11.002>
- Hoff, E., 2003. The Specificity of Environmental Influence: Socioeconomic Status Affects Early Vocabulary Development Via Maternal Speech. *Child Dev.* 74, 1368–1378. <https://doi.org/10.1111/1467-8624.00612>
- Hoff-Ginsberg, E., 1998. The relation of birth order and socioeconomic status to children's language experience and language development. *Appl. Psycholinguist.* 19, 603–629. <https://doi.org/10.1017/S0142716400010389>
- Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., Hedges, L.V., 2010. Sources of Variability in Children's Language Growth. *Cognit. Psychol.* 61, 343–365. <https://doi.org/10.1016/j.cogpsych.2010.08.002>
- Kuhl, P.K., 2004. Early language acquisition: cracking the speech code. *Nat. Rev. Neurosci.* 5, 831–843. <https://doi.org/10.1038/nrn1533>

- Kuhl, P.K., Tsao, F.-M., Liu, H.-M., 2003. Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proc. Natl. Acad. Sci.* 100, 9096–9101. <https://doi.org/10.1073/pnas.1532872100>
- Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B., 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *J. Stat. Softw.* 82. <https://doi.org/10.18637/jss.v082.i13>
- Lenth, R.V., 2022. Estimated Marginal Means, aka Least-Squares Means. R package emmeans version 1.7.2. [WWW Document]. URL <https://CRAN.R-project.org/package=emmeans> (accessed 7.19.22).
- Murray, L., 1985. Emotional regulations of interactions between two-month-olds and their mothers. *Soc. Percept. Infants* 177–197.
- National Institute of Child Health, 2000. The Relation of Child Care to Cognitive and Language Development. *Child Dev.* 71, 960–980. <https://doi.org/10.1111/1467-8624.00202>
- Peisner-Feinberg, E.S., Burchinal, M.R., Clifford, R.M., Culkin, M.L., Howes, C., Kagan, S.L., Yazejian, N., 2001. The Relation of Preschool Child-Care Quality to Children’s Cognitive and Social Developmental Trajectories through Second Grade. *Child Dev.* 72, 1534–1553. <https://doi.org/10.1111/1467-8624.00364>
- Pine, J.M., 1995. Variation in Vocabulary Development as a Function of Birth Order. *Child Dev.* 66, 272. <https://doi.org/10.2307/1131205>
- R Core Team, 2018. R: A language and environment for statistical computing. R Found. Stat. Comput. Vienna Austria.
- Robinson, L.F., Reis, H.T., 1989. The effects of interruption, gender, and status on interpersonal perceptions. *J. Nonverbal Behav.* 13, 141–153. <https://doi.org/10.1007/BF00987046>
- Rowe, M.L., 2008. Child-directed speech: relation to socioeconomic status, knowledge of child development and child vocabulary skill*. *J. Child Lang.* 35, 185–205. <https://doi.org/10.1017/S0305000907008343>
- Sacks, H., Schegloff, E.A., Jefferson, G., 1974. A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language* 50, 696–735. <https://doi.org/10.2307/412243>
- Shneidman, L.A., Arroyo, M.E., Levine, S.C., Goldin-Meadow, S., 2013. What counts as effective input for word learning? *J. Child Lang.* 40, 672–686. <https://doi.org/10.1017/S0305000912000141>
- Shneidman, L.A., Goldin-Meadow, S., 2012. Language input and acquisition in a Mayan village: how important is directed speech? *Dev. Sci.* 15, 659–673. <https://doi.org/10.1111/j.1467-7687.2012.01168.x>
- Stern, D.N., Jaffe, J., Beebe, B., Bennett, S.L., 1975. Vocalizing in unison and in alternation: Two modes of communication within the mother- infant dyad. *Ann. N. Y. Acad. Sci.* 263, 89–100. <https://doi.org/10.1111/j.1749-6632.1975.tb41574.x>
- Stivers, T., Enfield, N.J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., Ruitter, J.P. de, Yoon, K.-E., Levinson, S.C., 2009. Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci.* 106, 10587–10592. <https://doi.org/10.1073/pnas.0903616106>
- Striano, T., Henning, A., Stahl, D., 2005. Sensitivity to social contingencies between 1 and 3 months of age. *Dev. Sci.* 8, 509–518. <https://doi.org/10.1111/j.1467-7687.2005.00442.x>
- ter Maat, M., Truong, K.P., Heylen, D., 2010. How Turn-Taking Strategies Influence Users’ Impressions of an Agent, in: Allbeck, J., Badler, N., Bickmore, T., Pelachaud, C., Safonova, A. (Eds.), *Intelligent Virtual Agents, Lecture Notes in Computer Science*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 441–453. https://doi.org/10.1007/978-3-642-15892-6_48
- The ManyBabies Consortium, 2020. Quantifying Sources of Variability in Infancy Research Using the Infant-Directed-Speech Preference: *Adv. Methods Pract. Psychol. Sci.* 3, 24–52. <https://doi.org/10.1177/2515245919900809>

- The ManyBabies Consortium, Frank, M.C., Alcock, K.J., Arias-Trejo, N., Aschersleben, G., Baldwin, D., Barbu, S., Bergelson, E., Bergmann, C., Black, A.K., Blything, R., Böhlend, M.P., Bolitho, P., Borovsky, A., Brady, S.M., Braun, B., Brown, A., Byers-Heinlein, K., Campbell, L.E., Cashon, C., Choi, M., Christodoulou, J., Cirelli, L.K., Conte, S., Cordes, S., Cox, C., Cristia, A., Cusack, R., Davies, C., de Klerk, M., Delle Luche, C., Ruitter, L. de, Dinakar, D., Dixon, K.C., Durier, V., Durrant, S., Fennell, C., Ferguson, B., Ferry, A., Fikkert, P., Flanagan, T., Floccia, C., Foley, M., Fritzsche, T., Frost, R.L.A., Gampe, A., Gervain, J., Gonzalez-Gomez, N., Gupta, A., Hahn, L.E., Kiley Hamlin, J., Hannon, E.E., Havron, N., Hay, J., Hernik, M., Höhle, B., Houston, D.M., Howard, L.H., Ishikawa, M., Itakura, S., Jackson, I., Jakobsen, K.V., Jarto, M., Johnson, S.P., Junge, C., Karadag, D., Kartushina, N., Kellier, D.J., Keren-Portnoy, T., Klassen, K., Kline, M., Ko, E.-S., Kominsky, J.F., Kosie, J.E., Kragness, H.E., Krieger, A.A.R., Krieger, F., Lany, J., Lazo, R.J., Lee, M., Leservoiesier, C., Levelt, C., Lew-Williams, C., Lippold, M., Liszkowski, U., Liu, L., Luke, S.G., Lundwall, R.A., Macchi Cassia, V., Mani, N., Marino, C., Martin, A., Mastroberardino, M., Mateu, V., Mayor, J., Menn, K., Michel, C., Moriguchi, Y., Morris, B., Nave, K.M., Nazzi, T., Noble, C., Novack, M.A., Olesen, N.M., John Orena, A., Ota, M., Panneton, R., Esfahani, S.P., Paulus, M., Pletti, C., Polka, L., Potter, C., Rabagliati, H., Ramachandran, S., Rennels, J.L., Reynolds, G.D., Roth, K.C., Rothwell, C., Rubez, D., Ryjova, Y., Saffran, J., Sato, A., Savelkouls, S., Schachner, A., Schafer, G., Schreiner, M.S., Seidl, A., Shukla, M., Simpson, E.A., Singh, L., Skarabela, B., Soley, G., Sundara, M., Theakston, A., Thompson, A., Trainor, L.J., Trehub, S.E., Trøan, A.S., Tsui, A.S.-M., Twomey, K., Von Holzen, K., Wang, Y., Waxman, S., Werker, J.F., Wermelinger, S., Woolard, A., Yurovsky, D., Zahner, K., Zettersten, M., Soderstrom, M., 2020. Quantifying Sources of Variability in Infancy Research Using the Infant-Directed-Speech Preference. *Adv. Methods Pract. Psychol. Sci.* 3, 24–52. <https://doi.org/10.1177/2515245919900809>
- Torr, J., Pham, L., 2016. Educator Talk in Long Day Care Nurseries: How Context Shapes Meaning. *Early Child. Educ. J.* 44, 245–254. <https://doi.org/10.1007/s10643-015-0705-6>
- Tran, H., Weinraub, M., 2006. Child care effects in context: Quality, stability, and multiplicity in nonmaternal child care arrangements during the first 15 months of life. *Dev. Psychol.* 42, 566–582. <https://doi.org/10.1037/0012-1649.42.3.566>
- Zambrana, I.M., Ystrom, E., Pons, F., 2012. Impact of Gender, Maternal Education, and Birth Order on the Development of Language Comprehension: A Longitudinal Study from 18 to 36 Months of Age. *J. Dev. Behav. Pediatr.* 33, 146–155. <https://doi.org/10.1097/DBP.0b013e31823d4f83>

Article 3 :

Cross- and same-sex analysis of parent-infant protoconversations: fathers also matters

**Bastien Meunier¹, Maéva Thomazeau¹, Alban Lemasson^{1,2*},
Virginie Durier^{1*}**

¹ Univ Rennes, CNRS, Normandie Univ, EthoS (Éthologie animale et humaine) - UMR 6552, Rennes, France

² Institut Universitaire de France, 1 rue Descartes, 75231 Paris, Cedex 05, France

* co-last authors

Synthèse de l'article 3

Contexte

Dès la naissance, les nourrissons préverbaux sont capables d'interagir vocalement avec leurs parents dans des proto-conversations. Les nourrissons améliorent leur capacité à coordonner leurs productions vocales avec leurs parents, tout comme les parents apprennent à s'adapter aux vocalisations de leurs nourrissons. La structure temporelle de ces proto-conversations dépend donc du comportement du parent et du nourrisson, lequel peut varier en fonction du sexe du nourrisson ou de son interlocuteur. Si les proto-conversations mères-nourrissons ont beaucoup été étudiées, les études incluant les dyades pères-nourrissons sont beaucoup moins nombreuses et les éventuelles différences sont donc moins connues.

Méthode

Nous avons observé des interactions dyadiques parents-nourrisson de 10 min. Nous avons pu observer 32 nourrissons au total (17 garçons et 15 filles), certains en interaction avec leur mère uniquement, d'autres avec leur père, et enfin d'autres avec leur mère et leur père (mais pas en même temps). Nous avons demandé aux parents d'interagir naturellement avec leur nourrisson pour susciter des vocalisations. Nous avons analysé les émissions vocales des parents et des nourrissons séparément en définissant trois patterns temporels (chevauchement, tour de parole, et isolé) afin de voir s'il y avait une influence du sexe des nourrissons et des parents sur la fréquence de ces patterns.

Résultats

Nos résultats ont ainsi révélé des différences liées au sexe dans les proto-conversations parent-nourrisson. Nous n'avons pas trouvé de différences entre les pères et les mères dans leur comportement vocal. Les deux parents ont plus parlé aux garçons qu'aux filles lors des interactions dyadiques. Concernant les nourrissons, les garçons ont plus produit de vocalisations en chevauchement que les filles mais uniquement avec leur père.

Cross- and same-sex analysis of parent-infant proto-conversations: fathers also matters

Bastien Meunier¹, Maéva Thomazeau¹, Alban Lemasson^{1,2*}, Virginie Durier^{1*}

¹ Univ Rennes, CNRS, Normandie Univ, EthoS (Éthologie animale et humaine) - UMR 6552, Rennes, France

² Institut Universitaire de France, 1 rue Descartes, 75231 Paris, Cedex 05, France

* co-last authors

Abstract

Infants interact vocally with their parents long before they can speak. Parents and infants coordinate their vocal emissions in a turn-taking pattern, a pattern found in conversations between human adults. Most studies of proto-conversations have focused on mothers and their infants, but little is known about the temporal coordination between fathers and their infants. To investigate whether vocal temporal coordination during proto-conversations differs according to the sex of the interlocutors, we conducted an observational study on dyadic interactions of 32 infants aged 3 to 11 months (N=15 girls). Some had to interact with their father, others with their mother, still others with both (but not at the same time) in 10-min interactions. We analyzed the vocal emissions of parents and infants separately, defining three temporal patterns (overlapping, turn-taking, and isolated). Our results revealed no differences between the vocal behavior of fathers and mothers. Both parents spoke more to boys than to girls. Finally, regarding infants, boys overlapped more than girls, but only with fathers.

Introduction

While it is generally accepted that infancy ends with the emergence of language after the first year of life, infants interact vocally with their parents extensively throughout this period. Infants produce vocalizations very early and in a flexible way (André et al. 2020). The first month, they are already able to imitate heard vowel sounds (Kuhl and Meltzoff, 1996), and to use different types of vocalizations to express positive, negative and neutral emotions at 3-4 months (Oller et al., 2013). These vocal utterances are elicited and encouraged by parents' own utterances (Bloom, 1975; Bloom et al., 1987), especially when infants' vocalizations are speech-like (Gros-Louis et al., 2006; Yoo et al., 2018). This mechanism leads to parental and infant vocalizations being produced contingently in alternation increasingly during development (Bateson, 1975; Ginsburg and Kilbourne, 1988; Stern et al., 1975), a reminder of the turn-taking pattern observed in adult conversations (Sacks et al., 1974; Stivers et al., 2009). Infants have an active role in these proto-conversations, since they are able to initiate and anticipate turns at 2 months (Gratier et al., 2015). Studies suggest that infants are prepared to interact in this way since this type of contingent vocal response has been demonstrated in neonates and preterm infants (Caskey et al., 2014; Dominguez et al., 2016). In contrast, the decrease in the number of overlaps with development show a role of learning in turn-taking skills (Ginsburg and Kilbourne, 1988; Hilbrink et al., 2015).

Interactions with infants have an impact on their development and shape their future interactions. Infants aged 4-5 months can thus have different expectations on the basis of one single interaction: faced with two strangers interacting contingently with them, they paid more attention to the one who had interacted contingently with them a few days earlier than to the one who had interacted non-contingently (Bigelow and Birch, 1999). Nevertheless, the interaction style and preference of infants are mainly influenced by their social relationships, defined by the repetition of interactions with the same person, and so with the members of their household. For example, a 5-month-old infant already prefers his mother's level of temporal contingency, even when this is rather low compared with that of other mothers (Bigelow, 1998). Parents' style of interaction with their child can even have repercussions on the child's future production and social integration: children rejected by their peers are those who perform longer turns, interrupt frequently and fail to respond contingently to requests, a pattern that may be linked to the large number of interactions with their parents where the latter leave no time to respond (Black and Logan, 1995; Hazen and Black, 1989). The composition of a household is

not limited to the parents, and the presence of siblings also influences the quality and quantity of interactions that infants will have with their parents (Jones and Adamson, 1987; Oshima-Takane and Robbins, 2003), leading to differences in children's language development depending on their birth order (Hoff-Ginsberg, 1998).

While it is acknowledged that turn-taking interactions with parents are crucial for infants, most studies of these proto-conversations focus on mother-infant interactions and little is known about interactions with fathers. Indeed, mothers have traditionally assumed the bulk of infant care, leading researchers to focus on mothering rather than fathering (Bornstein and Arterberry, 2022). Nevertheless, fathers are also found coordinating their vocal productions contingently with newborns (Boiteau et al., 2021), and practicing infant-directed-speech (IDS), a way of talking to infants that is essential for language learning (Ramírez, 2022). Despite this, parents differ in certain ways: in the quantity of interactions, with fathers generally interacting less with their infants than mothers (Lamb et al., 1985; Lang et al., 2014), but also in quality, with father-infant interactions more characterized by physical play than mother-infant interactions are (Lamb, 1977; Lamb et al., 1985), and mothers being more responsive to infants' attention signals (Power, 1985; Power and Parke, 1983). Parents may also behave differently according to the infants' sex, with mothers, for example, responding more to their daughters than to their sons, and fathers responding more to their sons than to their daughters (Johnson et al., 2014). These differences in interactions according to parents' gender thus lead infants to have different experiences and therefore possibly behavioral differences.

Studying sex differences in parent-infant interactions can be done in a number of ways. Firstly, studies can be conducted in a naturalistic context, with researchers moving into the subjects' homes (Lamb, 1977; Power and Parke, 1986), or in a more controlled setting by inviting the subjects in laboratory (Clearfield and Nelson, 2006). Similarly, the contexts can be free, as when a microphone is set up for a day to record all the subjects' vocal productions (Johnson et al., 2014), or parents can be explicitly asked to interact in a certain way, by asking them to play for example (Feldman, 2003). Finally, measurements differ from one study to another, where researchers may choose to record attitudes, contacts, gazes or even the content of vocal productions. Here, we wanted to study parent-infant vocal interactions to see if there were differences in their temporal coordination according to the sex of the interactants. We therefore decided to observe strictly dyadic parent-infant interactions, by explicitly asking parents to interact with their infant in order to elicit vocalizations. We conducted these observations on

infants aged between 3 and 11 months, to cover the whole period of emergence of these proto-conversations.

Methods

Participants

Parents of infants aged from 3 to 11 months living in Rennes (France) were contacted by mail to present the study. Parents could participate from home by filming themselves or in the laboratory (N_{lab}=28, one infant participated with the mother in the lab and with the father at home). A total of 32 infants took part in the study (N=17 boys, N=15 girls), some with their mothers only (N = 15), some with their fathers only (N = 4), and some with their fathers and mothers (N = 13). Our exclusion criteria were premature birth and developmental disorders.

Procedure

Parents were asked to interact face-to-face with their infants for 10 min. Parents were instructed to talk to their infant as naturally as possible in an attempt to elicit vocalizations (no singing or rhyming). They were allowed to use a few toys if necessary (as long as the toys did not represent a character in anyway, i.e. no doll or stuffed animal...). In the laboratory, the parent was left alone with the infant in a baby-friendly experimental room for ten minutes. If both parents were participating in the study, they had to take part one after the other. Interactions in the laboratory were recorded using a SONY HDR-CX625 video camera and a Kodak playtouch backup video camera to cover a different angle. The parents used their own cameras at home (they were simply asked to place the camera/smartphone on a support to avoid interfering with the interaction).

Data collection

Data collection on the vocal productions of parents and infants was carried out semi-automatically using the "Annotate To TextGrid" function in Praat software (v.6.1.52). The criteria used were those present by default (minimum pitch = 100 Hz, time step = 0.0s, silence threshold = -25.0 dB, minimum silent interval = 0.1s, minimum sounding interval = 0.1s). Laughter, crying, vegetative sounds such as sneezing and coughing, and kissing were not considered as vocal productions in the analysis, but rather as silences in order to keep only clear

voiced production (hereafter vocal utterances). When all vocal utterances were coded, we extracted silences durations between two vocal utterances from two different individuals (parent and infant) automatically using a praat script. Our praat script printed a line for each silence interval between two individuals, with a duration value measured from the textgrid boundaries (corresponding to the difference between the end time of a vocal utterance from one subject and the start time of the directly following vocal utterance from the other subject). Silences durations between two voiced utterances from the same subject were thus ignored.

Statistical analysis

In order to determine objectively whether a vocal utterance from one subject responds to a preceding vocal utterance from another, we plotted inter-utterances intervals durations obtained by our praat script on a frequency distribution graph to established a duration threshold (method used for nonhuman primates: Lemasson et al., 2018; Meunier et al., 2023). As this threshold could differ between adults and infants, we split the analysis in two (and therefore used two graphs): on the one hand we analyzed the parent-infant intervals to obtain the infant threshold, on the other hand we analyzed the infant-parent intervals to obtain the parent threshold. We defined three temporal patterns: an overlapping pattern (for the negative intervals, below 0s), a turn-taking pattern (for intervals between 0s and the threshold), and an isolated pattern (for intervals longer than the threshold).

We then statistically analyzed the vocal behavior of parents and infants using R software (R Core Team, 2021), with a significance threshold set at 0.05.

To find out whether parents and infants interacted differently according to their sex and to the sex of their partner of interaction, we built a generalized linear mixed model with the number of vocal utterances as the dependent variable, and as fixed factors the sex of the parent (mother or father), the sex of the infant (girl or boy), and the temporal pattern of the vocal utterance (overlapping, turn-taking, isolated). Fixed factors were studied independently and in interaction with each other. The random factor was the infant's identity, and the model's distribution law was binomial. We adapted this model into two parts: one that studied the infants' vocal behavior, with the infants' vocal utterances as the dependent variable, and one that studied the parents' vocal behavior, with the parents' vocal utterances as the dependent variable. We analyzed these GLMM with type II ANOVAs with a Wald chi-squared. We used packages: "MASS" version 7.3.54 (Venables and Ripley, 2002), "rstatix" version 0.7.0 (Kassambara, 2021), and "emmeans" version 1.7.2 (Lenth, 2022).

Results

Vocal utterances temporal patterns

By studying the distribution of parent-infant intervals, we graphically determine the thresholds (Fig 1). Thus, for an infant's vocal utterance to be considered as a response to the parent, the interval between them should be less than 3300ms. In the other way, for a parent's vocal utterance to be considered as a response to the infant, the interval between them should be less than 2300ms. An infant's vocal utterance made more than 3300 ms after a parent's utterance was considered to be isolated, and the same applied to a parent's vocal utterance made more than 2300 ms after an infant's vocal utterance. Among the vocal utterances in response to another, we distinguished between those that were overlapping (negative response time) and those that were not (response time between 0 and the threshold).

Parents' vocal behavior

As expected, most of the vocal utterances of parents could be considered as turn-taking exchanges. Our model revealed an effect of pattern and infant sex on parents' vocal behavior (GLMM: pattern, $\chi^2=183.17$, $df=2$, $p<0.001$; infants' sex, $\chi^2=5.49$, $df=1$, $p=0.02$), but not of parents' sex ($\chi^2=2.56$, $df=1$, $p=0.11$). Regardless of their sex, parents talked more to boys than girls (GLMM: girls vs boys, estimate=-0.511, SE=0.219, z-ratio=-2.329, $p=0.02$) (Fig 2). The model did not reveal any significant interactions between the factors studied (GLMM: parents' sex x infants' sex, $\chi^2<0.001$, $df=1$, $p=0.98$; parents' sex x pattern, $\chi^2=1.46$, $df=2$, $p=0.49$; infants' sex x pattern, $\chi^2=1$, $df=2$, $p=0.61$; parents' sex x infants' sex x pattern, $\chi^2=1.14$, $df=2$, $p=0.57$).

Infants' vocal behavior

Our model revealed an effect of pattern, infants' sex, and a trend of parents' sex on infants' vocal behavior (GLMM: pattern, $\chi^2=246.16$, $df=2$, $p<0.001$; infants' sex, $\chi^2=5.87$, $df=1$, $p=0.02$; parents' sex, $\chi^2=3.05$, $df=1$, $p=0.08$). Above all, the interaction between these three factors was significant (GLMM: infants' sex x parent's sex x pattern, $\chi^2=6.7$, $df=2$, $p=0.04$). In fact, boys overlapped more with their fathers than girls (GLMM: for overlapping utterances, girls-fathers vs boys-fathers, estimate=-1.11, SE=0.37, z-ratio=-3, $p=0.01$) (Fig 3).

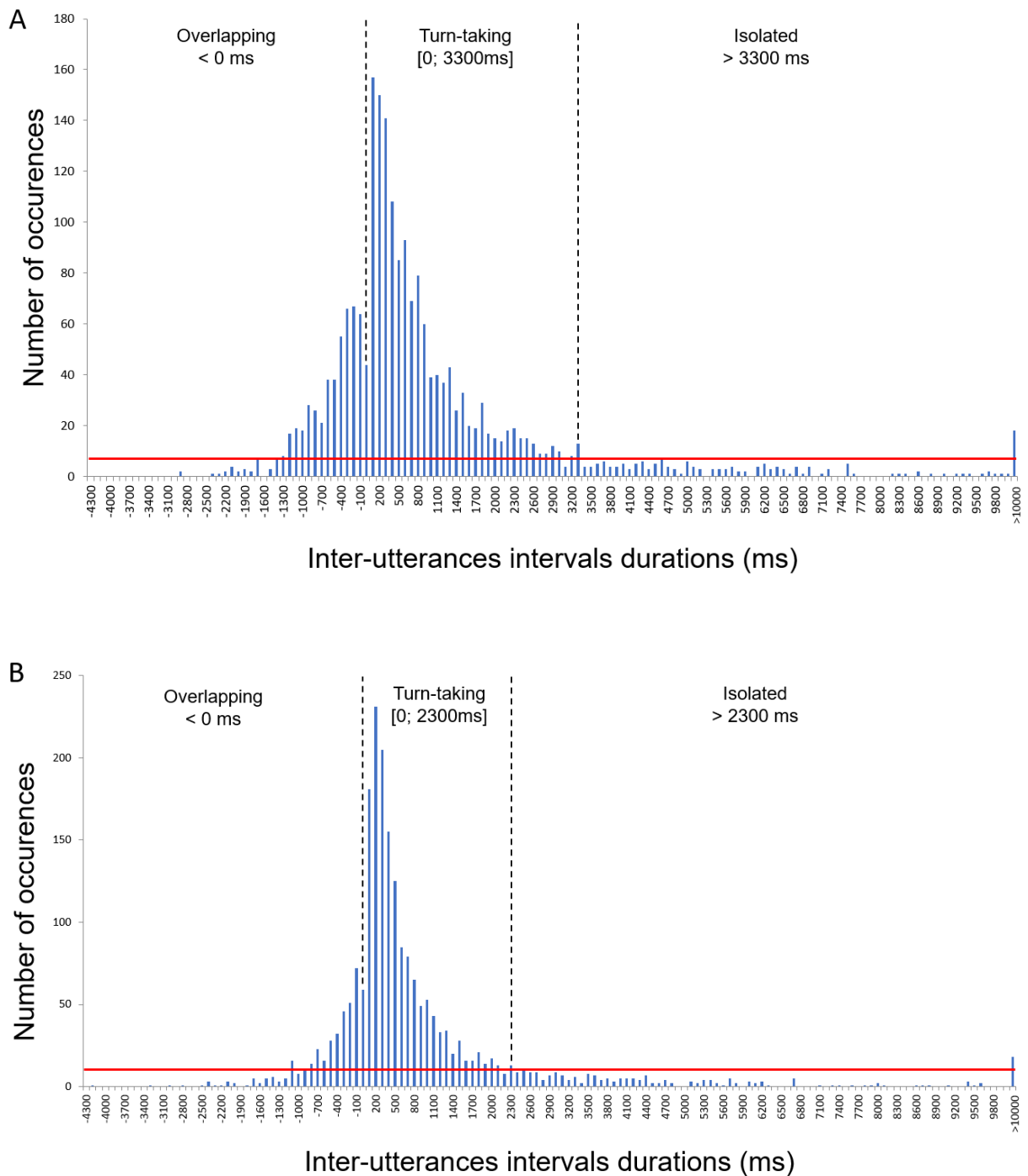


Fig 1. Distributions of inter-utterances intervals durations in milliseconds. Graph A represent the parent-infant intervals and so the infants' response time. Graph B represent the infant-parent intervals and so the parents' response time. The horizontal red line represents the baseline of the occurrence of intervals durations. The vertical dotted lines delimited the three-temporal pattern. The first ones from the left are placed at 0, delimiting overlapping utterances from turn-taking ones. The second ones are placed where intervals start declining below the baseline, thus delimiting turn-taking utterances from isolated ones.

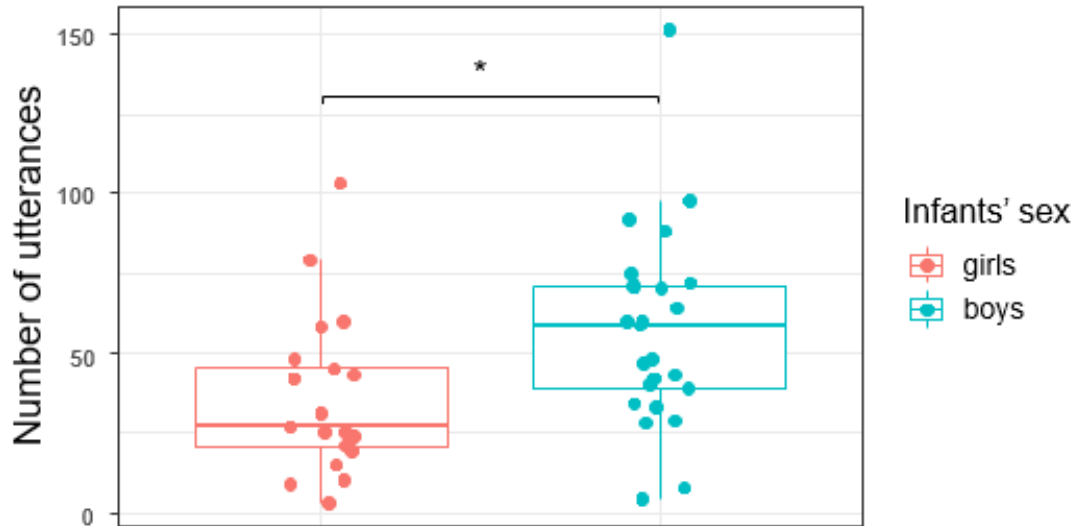


Fig 2. Number of parents' utterances in relation to infants' sex. * p=0.02

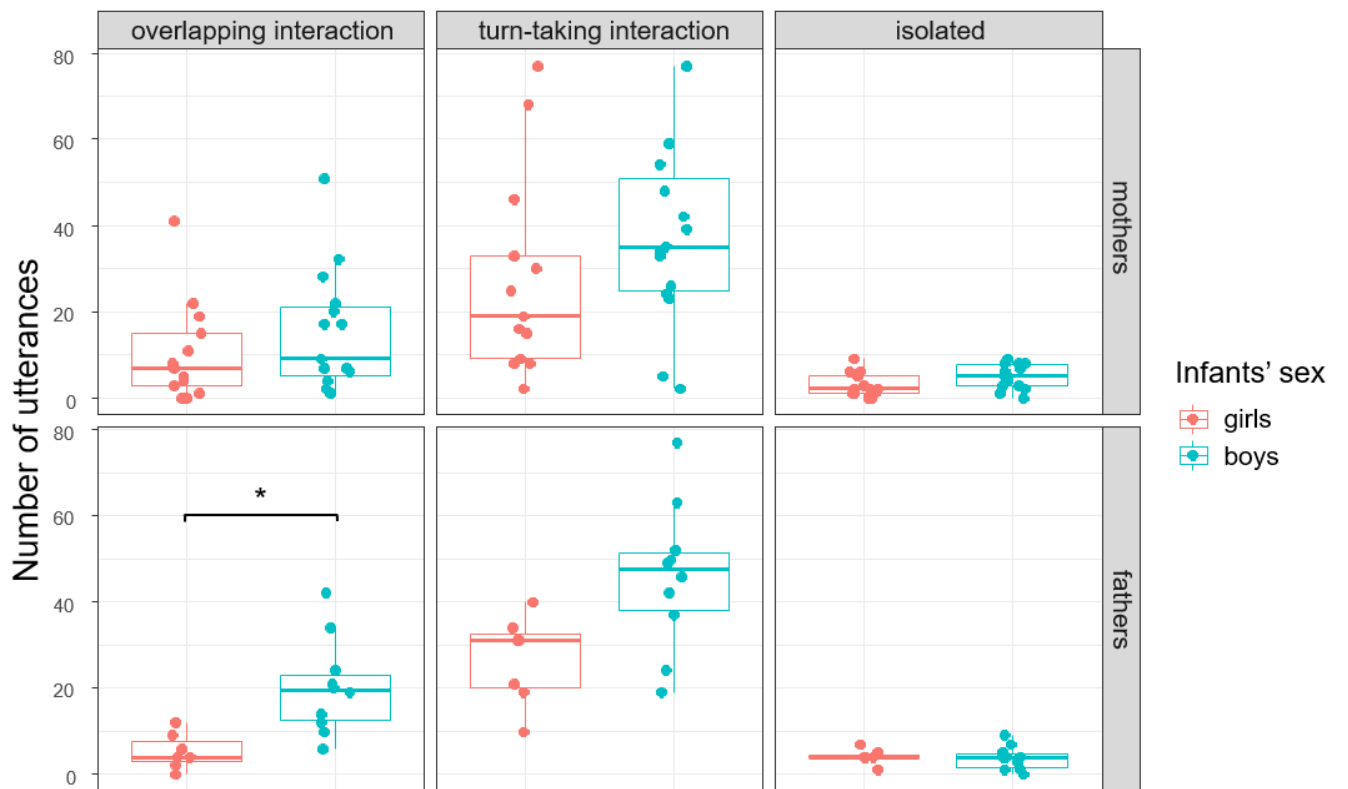


Fig 3. Number of infants' utterances from different patterns in relation to parents' sex. * p=0.01

Discussion

Our study revealed sex-related differences in parent-infant proto-conversations. Although there were no behavioral differences between mothers and fathers, both produced more vocal utterances during interactions with boys than with girls. For their part, the infants' vocal behavior depended both on their sex and on the parent with whom they were interacting, since boys overlapped more than girls, but only with fathers.

When we compared the parents' vocal behavior, we found no significant differences. However, some studies have highlighted such differences. This is the case of Feldman (2003), who showed that father-infant interactions were characterized by peaks of excitement and intensively positive attentional states, whereas mother-infant interactions were more graded, with one or no peaks of excitement and neutral or low attentional states on the part of the infants. The study by Johnson et al. (2014) further highlighted that mothers responded more to infant vocalizations than fathers, confirming the findings of other authors that mothers are more responsive to their infants' attentional cues than fathers (Power, 1985; Power and Parke, 1983). One possible reason for the lack of difference in our study may be that parents were strongly urged to use their speech to elicit vocal responses from their infants. As fathers and mothers were given the same precise instructions, they did not differ in terms of the amount of language input they provided to their infants.

Although they did not differ in their vocal behavior, fathers and mothers talked differently to their infants depending on their sex: they produced more utterances in the presence of boys than girls. This result may be surprising, as studies have shown that, conversely, parents talk more to girls than to boys, which would explain their better language skills compared to boys (Clearfield and Nelson, 2006; Leaper et al., 1998; Lovas, 2011; Sung et al., 2013). But an absence of mother-father differences during the 10 min we observed does not mean that there are no differences in everyday interactions. For example, mothers of boys generally make more requests for attention than mothers of girls (Clearfield and Nelson, 2006). Yet, in a context like our study where parents were asked to elicit vocalizations from their infants, the proportion of such speech acts may be greater than in others, causing this difference in language input in favor of boys. Also, Mothers of boys make more comments than mothers of girls, who make more interpretations (Clearfield and Nelson, 2006). Then, our instruction prohibiting figurative toys may have played a role, since it is not present in every other study. We know that some toys, such as dolls can elicit more language than others (O'Brien and Nagle, 1987). This highly stereotyped toy may be one of the reasons why parents talk more to girls during interactions,

because when they do, they are more likely to use dolls than with boys. This difference may therefore have been eliminated by the instruction not to use figurative toys.

Moreover, infants differed in vocal behavior only in the context of interactions with fathers: boys overlapped more with their fathers' speech than girls. These overlaps can be interpreted as a particular motivation on the part of infants to interact with their fathers. Interactions with fathers are indeed more attractive to infants because of the importance attached to physical play and the excitement it arouses (Feldman, 2003; Lamb et al., 1985), but this applies to infants of both sexes. Here, since the sex of the parent had no effect on the infants' behavior, it suggests an adjustment on the part of the boys towards the parent of the same sex. Osofsky and O'Connell (1977) have indeed revealed that male newborns already showed greater reactivity to peaks of excitement, irritability and lability of emotional states than their female counterparts. This endogenous factor would therefore match better with the fathers' interaction style, explaining the boys' increased motivation to interact. Infants also show greater distress when the same-sex parent leaves the room (Weinraub and Frankel, 1977), and same-sex parent-infant interactions show better synchrony with shorter reaction times (Feldman, 2003). An alternative hypothesis for interpreting these overlaps is that boys are less skilled at doing turn-taking than girls. The avoidance of overlaps in conversation is a universal principle of turn-taking (Stivers et al., 2009), which is acquired very early on, since decreases in the number of overlaps are observed in early infancy (Ginsburg and Kilbourne, 1988; Hilbrink et al., 2015). Thus, because they overlap less with their fathers than boys, girls could be more advanced in mastering turn-taking than their counterparts, an advance to be found later in verbal communication. As interactions with fathers are different from those with mothers, the difference between girls and boys was only visible during the early stages. A more detailed analysis of speech, mimics and gestures during these interactions could help determine what differed between them.

Although preliminary, this study provides a new perspective on the effect of interactant sex on proto-conversations. Indeed, while behavioral differences related to the parents' sex are well known, behavioral differences related to the infants' sex from their perspective are much rarer, and invite questioning. More detailed analyses, considering speech, facial mimics and gestures, are needed to decide between our two interpretative hypotheses: same-sex matching or differences in turn-taking skills. Also, looking several parent-infant interactions at different times (e.g. through a longitudinal study) could help us to get a better idea of the influence of infants' social experience on these interactions.

References

- Bateson, M.C., 1975. Mother-infant exchanges: the epigenesis of conversational interaction. *Ann. N. Y. Acad. Sci.* 263, 101–113. <https://doi.org/10.1111/j.1749-6632.1975.tb41575.x>
- Bigelow, A.E., 1998. Infants' sensitivity to familiar imperfect contingencies in social interaction. *Infant Behav. Dev.* 21, 149–162. [https://doi.org/10.1016/S0163-6383\(98\)90060-1](https://doi.org/10.1016/S0163-6383(98)90060-1)
- Bigelow, A.E., Birch, S.A.J., 1999. The effects of contingency in previous interactions on infants' preference for social partners. *Infant Behav. Dev.* 22, 367–382. [https://doi.org/10.1016/S0163-6383\(99\)00016-8](https://doi.org/10.1016/S0163-6383(99)00016-8)
- Black, B., Logan, A., 1995. Links between Communication Patterns in IVIothei Child, Father-Child, and Child-Peer Interactions and Children's Social Status. *Child Dev.* 66, 255–274.
- Bloom, K., 1975. Social elicitation of infant vocal behavior. *J. Exp. Child Psychol.* 20, 51–58. [https://doi.org/10.1016/0022-0965\(75\)90025-9](https://doi.org/10.1016/0022-0965(75)90025-9)
- Bloom, K., Russell, A., Wassenberg, K., 1987. Turn taking affects the quality of infant vocalizations. *J. Child Lang.* 14, 211. <https://doi.org/10.1017/S0305000900012897>
- Boiteau, C., Kokkinaki, T., Sankey, C., Buil, A., Gratier, M., Devouche, E., 2021. Father–newborn vocal interaction: A contribution to the theory of innate intersubjectivity. *Infant Child Dev.* 30, e2259. <https://doi.org/10.1002/icd.2259>
- Bornstein, M.H., Arterberry, M.E., 2022. *Infancy: The Basics*. Taylor & Francis.
- Caskey, M., Stephens, B., Tucker, R., Vohr, B., 2014. Adult Talk in the NICU With Preterm Infants and Developmental Outcomes. *PEDIATRICS* 133, e578–e584. <https://doi.org/10.1542/peds.2013-0104>
- Clearfield, M.W., Nelson, N.M., 2006. Sex Differences in Mothers' Speech and Play Behavior with 6-, 9-, and 14-Month-Old Infants. *Sex Roles* 54, 127–137. <https://doi.org/10.1007/s11199-005-8874-1>
- Dominguez, S., Devouche, E., Apter, G., Gratier, M., 2016. The Roots of Turn-Taking in the Neonatal Period. *Infant Child Dev.* 25, 240–255. <https://doi.org/10.1002/icd.1976>
- Feldman, R., 2003. Infant-mother and infant-father synchrony: The coregulation of positive arousal. *Infant Ment. Health J.* 24, 1–23. <https://doi.org/10.1002/imhj.10041>
- Ginsburg, G.P., Kilbourne, B.K., 1988. Emergence of vocal alternation in mother-infant interchanges. *J. Child Lang.* 15, 221–235. <https://doi.org/10.1017/S0305000900012344>
- Gratier, M., Devouche, E., Guellai, B., Infanti, R., Yilmaz, E., Parlato, E., 2015. Early development of turn-taking in vocal interaction between mothers and infants. *Front. Psychol.* 6. <https://doi.org/10.3389/fpsyg.2015.01167>
- Gros-Louis, J., West, M.J., Goldstein, M.H., King, A.P., 2006. Mothers provide differential feedback to infants' prelinguistic sounds. *Int. J. Behav. Dev.* 30, 509–516. <https://doi.org/10.1177/0165025406071914>
- Hazen, N.L., Black, B., 1989. Preschool Peer Communication Skills: The Role of Social Status and Interaction Context. *Child Dev.* 60, 867–876. <https://doi.org/10.2307/1131028>
- Hilbrink, E.E., Gattis, M., Levinson, S.C., 2015. Early developmental changes in the timing of turn-taking: a longitudinal study of mother–infant interaction. *Front. Psychol.* 6. <https://doi.org/10.3389/fpsyg.2015.01492>
- Hoff-Ginsberg, E., 1998. The relation of birth order and socioeconomic status to children's language experience and language development. *Appl. Psycholinguist.* 19, 603–629. <https://doi.org/10.1017/S0142716400010389>
- Johnson, K., Caskey, M., Rand, K., Tucker, R., Vohr, B., 2014. Gender Differences in Adult-Infant Communication in the First Months of Life. *Pediatrics* 134, e1603–e1610. <https://doi.org/10.1542/peds.2013-4289>

- Jones, C.P., Adamson, L.B., 1987. Language Use in Mother-Child and Mother-Child-Sibling Interactions 12.
- Kassambara, A., 2021. Pipe-Friendly Framework for Basic Statistical Tests. R package version 0.7.0 [WWW Document]. URL <https://CRAN.R-project.org/package=rstatix> (accessed 7.19.22).
- Kuhl, P.K., Meltzoff, A.N., 1996. Infant vocalizations in response to speech: Vocal imitation and developmental change. *J. Acoust. Soc. Am.* 100, 2425–2438.
- Lamb, M.E., 1977. Father-Infant and Mother-Infant Interaction in the First Year of Life. *Child Dev.* 48, 167–181. <https://doi.org/10.2307/1128896>
- Lamb, M.E., Pleck, J.H., Charnov, E.L., Levine, J.A., 1985. Paternal Behavior in Humans. *Am. Zool.* 25, 883–894.
- Lang, S.N., Schoppe-Sullivan, S.J., Kotila, L.E., Feng, X., Dush, C.M.K., Johnson, S.C., 2014. Relations between Fathers’ and Mothers’ Infant Engagement Patterns in Dual-Earner Families and Toddler Competence. *J. Fam. Issues* 35, 1107–1127. <https://doi.org/10.1177/0192513X14522243>
- Leaper, C., Anderson, K.J., Sanders, P., 1998. Moderators of gender effects on parents’ talk to their children: A meta-analysis. *Dev. Psychol.* 34, 3–27. <https://doi.org/10.1037/0012-1649.34.1.3>
- Lemasson, A., Pereira, H., Levréro, F., 2018. Social basis of vocal interactions in western lowland gorillas (*Gorilla g. gorilla*). *J. Comp. Psychol.* 132, 141–151. <https://doi.org/10.1037/com0000105>
- Lenth, R.V., 2022. Estimated Marginal Means, aka Least-Squares Means. R package emmeans version 1.7.2. [WWW Document]. URL <https://CRAN.R-project.org/package=emmeans> (accessed 7.19.22).
- Lovas, G.S., 2011. Gender and patterns of language development in mother-toddler and father-toddler dyads. *First Lang.* 31, 83–108. <https://doi.org/10.1177/0142723709359241>
- Meunier, B., Durier, V., Giacalone, A., Coye, C., Lemasson, A., 2023. Social factors drive vocal exchanges in red-capped mangabeys (*Cercocebus torquatus*). *Sci. Nat.* 110, 2. <https://doi.org/10.1007/s00114-023-01830-3>
- O’Brien, M., Nagle, K.J., 1987. Parents’ speech to toddlers: the effect of play context. *J. Child Lang.* 14, 269–279. <https://doi.org/10.1017/S0305000900012927>
- 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. *Proc. Natl. Acad. Sci.* 110, 6318–6323. <https://doi.org/10.1073/pnas.1300337110>
- Oshima-Takane, Y., Robbins, M., 2003. Linguistic Environment of Secondborn Children. *First Lang.* 23, 21–40. <https://doi.org/10.1177/0142723703023001002>
- Osofsky, J.D., O’Connell, E.J., 1977. Patterning of Newborn Behavior in an Urban Population. *Child Dev.* 48, 532–536. <https://doi.org/10.2307/1128650>
- Power, T.G., 1985. Mother- and Father-Infant Play: A Developmental Analysis. *Child Dev.* 56, 1514–1524. <https://doi.org/10.2307/1130470>
- Power, T.G., Parke, R.D., 1986. Patterns of Early Socialization: Mother-and Father-Infant Interaction in the Home. *Int. J. Behav. Dev.* 9, 331–341. <https://doi.org/10.1177/016502548600900305>
- Power, T.G., Parke, R.D., 1983. Patterns of mother and father play with their 8-month-old infant: A multiple analyses approach. *Infant Behav. Dev.* 6, 453–459. [https://doi.org/10.1016/S0163-6383\(83\)90256-4](https://doi.org/10.1016/S0163-6383(83)90256-4)
- Ramírez, N.F., 2022. Fathers’ infant-directed speech and its effects on child language development. *Lang. Linguist. Compass* 16, e12448. <https://doi.org/10.1111/lnc3.12448>
- Sacks, H., Schegloff, E.A., Jefferson, G., 1974. A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language* 50, 696–735. <https://doi.org/10.2307/412243>

- Stern, D.N., Jaffe, J., Beebe, B., Bennett, S.L., 1975. Vocalizing in unison and in alternation: Two modes of communication within the mother- infant dyad. *Ann. N. Y. Acad. Sci.* 263, 89–100. <https://doi.org/10.1111/j.1749-6632.1975.tb41574.x>
- Stivers, T., Enfield, N.J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., Ruitter, J.P. de, Yoon, K.-E., Levinson, S.C., 2009. Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci.* 106, 10587–10592. <https://doi.org/10.1073/pnas.0903616106>
- Sung, J., Fausto-Sterling, A., Coll, C.G., Seifer, R., 2013. The Dynamics of Age and Sex in the Development of Mother–Infant Vocal Communication Between 3 and 11 Months. *Infancy* 18, 1135–1158. <https://doi.org/10.1111/infa.12019>
- Venables, W.N., Ripley, B.D., 2002. *Modern Applied Statistics with S*. Fourth Edition. Springer, New York. ISBN 0-387-95457-0.
- Weinraub, M., Frankel, J., 1977. Sex Differences in Parent-Infant Interaction during Free Play, Departure, and Separation. *Child Dev.* 48, 1240–1249. <https://doi.org/10.2307/1128481>
- Yoo, H., Bowman, D.A., Oller, D.K., 2018. The Origin of Protoconversation: An Examination of Caregiver Responses to Cry and Speech-Like Vocalizations. *Front. Psychol.* 9. <https://doi.org/10.3389/fpsyg.2018.01510>

Chapitre III :

Proto-conversations chez les mangabeys à collier (*Cercocebus torquatus*)

Article 4 :

Social factors drive vocal exchanges in red-capped mangabeys (*Cercocebus torquatus*)

**Bastien Meunier¹, Virginie Durier¹, Aline Giacalone¹,
Camille Coye^{1,2}, Alban Lemasson^{1,3}**

¹ Univ Rennes, CNRS, Normandie Univ, EthoS (Éthologie animale et humaine) - UMR 6552, Rennes, France

² Institut Jean Nicod, Département d'études cognitives, ENS, EHESS, CNRS, PSL University, 75005 Paris, France

³ Institut Universitaire de France 1 rue Descartes, Cedex 05, 75231 Paris, France

DOI : <https://doi.org/10.1007/s00114-023-01830-3>

Synthèse de l'article 4

Contexte

Parmi tous les cris émis par les primates non humains, la fonction des cris de contact à courte distance reste largement à déterminer. Ces cris sont les plus fréquents du répertoire et sont le plus souvent échangés entre individus de manière non aléatoire. A notre connaissance, aucune étude n'a jamais examiné la structure des échanges vocaux chez le mangabey à collier (*Cercocebus torquatus*), un singe semi-terrestre vivant dans les forêts africaines avec un système social complexe semi-tolérant/semi-despotique.

Méthode

Afin d'étudier les échanges vocaux de cette espèce et leurs liens avec des facteurs sociaux, nous avons observé 20 mangabeys captifs répartis en 4 groupes (N=13 mâles) avec des focus de 15min. Durant ces observations, nous avons relevé les vocalisations ainsi que les comportements sociaux de l'individu focal (toilette sociale, proximité spatiale, agressions et soumissions). Nous avons défini un échange vocal chez cette espèce comme une série de cris de contact effectués par deux individus ou plus avec des délais de réponse n'excédant pas 2s. Nous avons conduit une analyse au niveau individuel pour voir si la participation des individus aux échanges vocaux dépend de leur statut social, et au niveau dyadique pour voir si les échanges vocaux dépendaient de la nature des liens sociaux.

Résultats

Nos résultats ont révélé qu'au niveau individuel, plus le rang hiérarchique de l'individu est élevé, moins il initie d'échanges vocaux. En outre, les individus les plus intégrés socialement (ceux qui étaient au centre des toilettes sociales) avaient un temps de réponse moyen plus long que les moins intégrés. Au niveau dyadique, les partenaires d'échange préférés étaient les individus souvent observés à proximité l'un de l'autre ou les individus les plus éloignés en âge. En outre, plus les deux individus étaient éloignés dans la hiérarchie de dominance, plus le temps de réponse était court. Cette modulation des échanges vocaux par la dominance et l'affinité est cohérente avec le système social semi-despotique/semi-tolérant de l'espèce.



Social factors drive vocal exchanges in red-capped mangabeys (*Cercocebus torquatus*)

Bastien Meunier¹ · Virginie Durier¹ · Aline Giacalone¹ · Camille Coye^{1,2} · Alban Lemasson^{1,3} Received: 30 September 2022 / Revised: 4 January 2023 / Accepted: 6 January 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

Of all the calls made by non-human primates, the function of short-distance contact calls has largely remained to be determined. These calls are the most frequent in the repertoire and are most often exchanged between individuals in a non-random way. To our knowledge, no study has ever examined how vocal exchanges are structured in red-capped mangabeys (*Cercocebus torquatus*), a semi-terrestrial monkey living in the African forest with a complex semi-tolerant/semi-despotic social system. Our goal was to assess the organization of contact call exchanges in this species and their relationship with individual and social factors such as age, affinity, and hierarchy. Therefore, we observed several captive groups of red-capped mangabeys and collected data on vocal behavior, as well as grooming, agonistic behavior, and spatial proximity. We defined a vocal exchange in this species as a series of contact calls made by two or more individuals within a maximum inter-caller interval of 2 s. At the individual level, the higher the individual's hierarchical rank, the less they initiated exchanges. Furthermore, the most socially integrated individuals had a longer average response time than the less integrated ones. At the dyadic level, preferred exchange partners were individuals often observed near one other or individuals most distant in age. Also, the further apart two individuals were in the dominance hierarchy, the shorter the response time. Our results support both the social bonding hypothesis and a modulating key role of the dominance hierarchy on the social use of contact calls, which is in line with the social style of this species.

Keywords Turn-taking · Vocal response · Social bonding · Hierarchy · Old World monkeys

Introduction

Vocal communication in many animal species is achieved through several species-specific calls. Most of these calls are emitted in specific contexts, and the study of their function is therefore facilitated: anti-predation alarm calls (Macedonia and Evans, 1993), distress calls (Gouzoules et al., 1984), food calls (Slocombe and Zuberbühler, 2006), and

copulation calls (Maestriperi and Roney, 2005; Zanoli et al., 2021). However, the function of so-called contact calls, emitted in a broad range of contexts (traveling, feeding, resting, grooming...), has long been a source of debate. The diversity of hypotheses is partially due to the “contact call” category covering a great diversity of calls ranging from loud long-distance calls to soft short-distance calls and the flexible use of contact calls depending logically on the social needs of the species. Several functions, all related to socio-spatial coordination within a group and not mutually exclusive, have been proposed. For some authors, these calls would have a determinant role in the initiation and/or regulation of group movements (Neumann and Zuberbühler, 2016; Sperber et al., 2017), in the coordination of subgroup fusions (Briseño-Jaramillo et al., 2022), or would allow for maintained contact between isolated or distant individuals by giving spatial indications (Digweed et al., 2007; Snowdon and Hodun, 1981). Therefore, a proposed key function of all contact calls is to inform about individual identity (reviewed in Kondo and Watanabe, 2009). Indeed, contact calls are

Communicated by: Matthias Waltert

✉ Bastien Meunier
bastien.meunier@univ-rennes1.fr¹ Université de Rennes, Université de Normandie, CNRS, EthoS (Ethologie animale et humaine) - U.M.R 6552, F-35000 Rennes, France² Institut Jean Nicod, Département d'études cognitives, ENS, EHESS, CNRS, PSL University, 75005 Paris, France³ Institut Universitaire de France 1 rue Descartes, Cedex 05, 75231 Paris, France

typically individually variable (Bouchet et al., 2013), and the degree of acoustic identity coded varies according to the context: calls diverge more acoustically between callers when they move in visually closed areas, probably to highlight identity (Candiotti et al., 2012). This may allow receivers to anticipate a social encounter and to adapt their behaviors accordingly (Ramos-Fernández, 2005; Rendall and Owren, 2002).

However, in most species, contact calls are also frequently emitted during the day in the wild and captivity, even when individuals are not separated. Thus, this also occurs in contexts where there is no need to coordinate group movements or a lesser need to identify oneself since individuals can see each other (Bouchet et al., 2010; Ramos-Fernández, 2005; Range and Fischer, 2004; Byrne and Teixidor, 1999; Kudo, 1987). A growing number of studies propose an additional function for contact calls: social bonding. Beyond identity coding, contact calls are sometimes described as social badges advertising affinities through shared acoustic patterns (i.e., individuals with a particular social bond produce similar acoustic structures; Levréro et al., 2019; Lemasson and Hausberger, 2004; Snowden and Elowson, 1999). These calls could consequently be used to create new social bonds within a group and maintain existing bonds (Lemasson et al., 2011a). This hypothesis is notably based on the fact that (1) these calls are more often uttered within vocal interactions than in isolation and are, in fact, the most frequently exchanged call types in the vocal repertoire; (2) the temporal structure of the vocal exchanges is better explained by the nature of the social bond between interlocutors than by the inter-individual distances (Pougnault et al., 2020a). Parallels between contact call exchanges in non-human primates and human conversations have even been made (Pougnault et al., 2020a). These exchanges often occur in relaxed and peaceful contexts with visible partners in non-human primates. Moreover, exchange partners are not randomly chosen also in non-human primates. In several species, the social affinity between individuals within a group (i.e., inter-individual affiliative relationships are based on repeated positive and neutral interactions) is related to the frequency of exchanges between them (Arlet et al., 2015; Briseño-Jaramillo et al., 2018; Kulahci et al., 2015; Levréro et al., 2019). In some species, the age of the caller is related to the response rates, with, for example, older Campbell's monkey females being the most popular interlocutors (Lemasson et al., 2010a). In other species, the caller's hierarchical status matters more, such as in western lowland gorillas, where higher ranked group members exchange the most contact calls (Lemasson et al., 2018). These results are consistent with Dunbar's (1998) hypothesis that vocal exchanges fulfil the same bonding function as physical grooming, with the advantage that they can be performed at a distance, more frequently and with a greater number of partners.

Parallels between contact call exchanges and human conversations rely on the temporal rules structuring these interactions. For example, speech overlap avoidance and turn-taking are universal temporal rules in humans (Stivers et al., 2009). Pioneering studies on new world monkeys have shown that monkeys also take turns and that the individual's social status determines the order of vocalizations within a vocal exchange (Snowdon and Cleveland, 1984). In marmosets, turn-taking respects the characteristics of coupled oscillators found in human conversations (Takahashi et al., 2013). To avoid call overlaps, individuals respect a species-specific response time, but it can also vary according to social factors. In squirrel monkeys, response times are shorter between individuals with high affinity than between individuals with low affinity (Masataka and Biben, 1987). In Japanese macaques, response times vary with the physical distance between exchange partners (Sugiura, 2007) and between populations, suggesting a cultural transmission (Sugiura and Masataka, 1995) as found in humans (Stivers et al., 2009). As these examples show, vocal turn-taking is present in many animal species (reviewed in Pika et al., 2018).

While turn-taking skills may have an evolutionary inheritance component (turn-taking being widely distributed in primates), the importance of this coordinated behavior in daily life varies greatly between species. Turn-taking and call overlap are vocal interacting patterns found in non-human primates, and they can be used strategically by each species. In howler monkeys, while males take turns, females prefer to call simultaneously, likely compensating for their less intense voice (Briseño-Jaramillo et al., 2021). In apes, vocal interactions are more frequent in social species than in more solitary ones (i.e., orangutans), and turn-taking is more frequent in tolerant societies (i.e., bonobos and gorillas). In contrast, overlap occurs more often in despotic ones (i.e., chimpanzees) (Pougnault et al., 2022). Also, even in species using turn-taking regularly, some individuals fail to respect this rule more often than others (juveniles in Campbell's monkeys: Lemasson et al., 2011b; subordinate males in Japanese macaques: Lemasson et al., 2013). Using playback experiments of vocal exchanges violating the rule or respecting the overlap avoidance (Gorillas: Pougnault et al., 2020b) and the alternation between interlocutors (Campbell's monkeys: Lemasson et al., 2011a), authors confirmed the relevance of turn-taking for animals by their different reactions to the stimuli presented (head orientation, gazes towards the loudspeaker).

The studies mentioned above on vocal turn-taking in non-human primates and the possible social factors responsible for the observed diversity of use encourage researchers to carry out more comparative studies on the subject, focusing on a broader range of species. Here, we conducted a study on several captive groups

of red-capped mangabeys (*Cercocebus torquatus*). To our knowledge, no study has ever investigated the structure and potential function of the vocal exchanges of this species. We wanted to assess how temporally structured vocal exchanges may reflect an individual’s social position in the group. Red-capped mangabeys are interesting models for this topic for several reasons. First, they live in the wild in forested areas making vocal communication important for the species. Second, although they show a noticeable sexual dimorphism and sex differences in social and vocal behaviors, both males and females produce contact calls (Bouchet et al., 2010). In the wild, red-capped mangabeys are often seen in multi-male, multi-female groups with seasonal fission-fusion dynamics (Dolado et al., 2016), making the species socially flexible. Third, they have a semi-despotic/semi-tolerant social organization with both some of the characteristics of despotic societies (i.e., steep gradient of hierarchy, unidirectionality of aggression) and tolerant societies (i.e., high proximity between individuals, reciprocal grooming behavior and dominant individuals occupying peripheral positions) (Dolado and Beltran, 2012). Fourth, turn-taking-based contact call exchanges were found in closely related species (guenons, macaques, and baboons). Our first objective was to confirm the existence of turn-taking-based vocal exchanges in red-capped mangabeys and that it principally or exclusively concerns the contact call type. Investigations on more socially varied species are needed to better understand the evolution of “conversational rules” in the primate lineage. Our second objective was to test which, if any, age-related, affinity-related, and hierarchy-related parameters best shape these exchanges in that species and thus to understand more about their function.

Methods

Ethical note

This purely observational study was conducted without modifying the living conditions of the captive mangabeys housed at the Station Biologique de Paimpont (University Rennes I, France), where animal facilities and animal care procedures are regularly monitored by the responsible local authorities (Housing agreement for research D35-211-18, delivered by the “Direction Départementale de la Cohésion Sociale et de la Protection des Populations” (DDCSPP)). This research protocol has been approved by the CREEA Ethic committee (“Comité Rennais d’Ethique en matière d’expérimentation animale”) under the reference APAFIS# 2021081711259974.

Subjects and housing conditions

Twenty captive adult red-capped mangabeys (*Cercocebus torquatus*) were observed (N=13 males, N = 7 females, Table 1). These individuals, all born in captivity, were housed at the Station Biologique de Paimpont. Individuals were divided into four groups (i.e., three only-male groups and one one-male multi-female group, Table 1), group compositions have been stable for several months or years. The enclosures for each group included an indoor portion (from 8 to 27m²) and an outdoor portion (from 15 to 37m²), with a height ranging from 2.5 to 4m. Individuals were free to come and go between different sections via tunnels, except during our observations when the monkeys were kept outside. The indoor enclosure temperature was maintained at 22°C, and the floor was covered with straw and wood shavings, while the outdoor floor was made of cement or covered with bark. All spaces were enriched with wood and metal perches. The mangabeys were fed twice daily, with fresh fruits and vegetables in the morning and monkey chows in the afternoon. Water was available *ad libitum*.

Data collection

A.G. observed groups II, III, and IV between September and October 2015, and B.M. observed group I between February and April 2021. As the data collections were made 6 years apart, it was impossible to calculate inter-observer

Table 1 Characteristics of the studied red-capped mangabeys

Social group	Subject	Sex	Date of birth
I.	Chipie	Female	28/06/1992
	Bell	Female	31/03/2002
	Julie	Female	08/05/2004
	Kargi	Male	19/05/2005
	Chipse	Female	03/01/2006
	Many	Female	14/08/2008
	Maillette	Female	29/12/2009
	Triskelle	Female	21/04/2015
II.	Isba	Male	20/04/2004
	Marti	Male	16/10/1998
	Lenni	Male	07/10/2006
III.	Pirate	Male	17/10/1992
	Carillon	Male	02/04/2007
	Georges	Male	05/06/2006
IV.	Elky	Male	06/11/2006
	Lupa	Male	27/09/2006
	Kamel	Male	07/09/2010
	Roby	Male	18/11/2010
	Tips	Male	10/07/2011
	Coët	Male	31/08/2011

agreement. Therefore, we decided to sample only non-confusing vocal and non-vocal behaviors in the same way. We applied a 15-min focal sampling method (Altmann, 1974), with 10 (groups II, III, IV) or 15 (group I) sessions per individual for a total of 60 h of observation. Observations were done between 9 and 17h. Mangabeys were observed in a pseudo-randomized order so that each individual was observed at different times of the day.

Vocal data collection and processing

We recorded vocalizations using a Sony ECM-672 semi-directional microphone connected to a Marantz PMD660 digital stereo recorder (sampling rate: 44.1kHz, amplitude resolution: 16 bits, WAV format). To compare the temporal emission pattern of contact calls with those of other call types (described below), we sampled all vocalizations emitted by the focal individual and vocalizations from others emitted in the same sequence. We defined a sequence as a series of vocalizations preceded and followed by a silence of at least 7 s. This choice of a long delay was a precaution to be sure to collect all the potential vocal exchanges of interest for our study, knowing that all the studies carried out previously on exchanges of contact calls in primates defined a response delay threshold between 1 and 3 s (Levréro et al., 2019; Briseno-Jaramillo et al., 2018; Candiotti et al., 2012; Lemasson et al., 2018, 2010a; Sugiura, 1993; Masataka and Biben, 1987). For each sequence, we noted all inter-call durations (time between the end of one call and the beginning of the following call), caller identities and call types. A negative duration signals an overlap of calls. Call types were determined using the red-capped mangabey's vocal repertoire published by Bouchet et al. (2010): *Oe+* (female sexual call), *Wi* (female distress call), *Ti+(Uh)* (female food excitement call), *A+(Uh)* (male food competition call), *Un+(Uh)* (female and male threat call), *Ro+(Uh)* (female and male contact call), *WaHoo* (female and male alert call), and *Whoop-Gobble* (loud male call). The identification of callers was relatively simple as follows: (i) there were few individuals per group (maximum 8); (ii) red-capped mangabeys open their mouths when they produce a call; and (iii) their vocal repertoire is relatively discrete, with no ambiguity between call types.

Social data collection and processing

We reported the duration (in seconds) of grooming received or emitted, the duration (in seconds) other individuals spent within arm's length, and the number of agonistic interactions received or emitted (in occurrences) for each focal individual. This latter included aggressive (e.g., grab the tail roughly, chase, hit, push, supplant) and submissive (e.g., flee, withdraw) behaviors.

Four indices were calculated at the individual level to study whether the exchange pattern varied with the social status of individuals. We calculated a dominance index for each individual following the method initially proposed by Zumpe and Michael (1986) and used by Aychet et al. (2022) in red-capped mangabeys. To calculate the dominance index of individual A according to this method: (1) we reported aggressive behaviors given by individual A to individual B as a proportion of all the aggressive behaviors sampled in the dyad AB, and we did the same for submissive behaviors received by individual A from individual B; (2) we then averaged these two measures (i.e., for dyad AB: (% aggressive behaviors given by A + % submissive behaviors received by A)/2); (3) a dominance index was obtained by averaging the scores of all within-group dyads involving A (e.g., AB, AC, AD...).

We also calculated an individual index of spatial integration (cumulative time during which individual A had a conspecific within arm's reach), as well as an individual index of social integration (cumulative time individual A was grooming or being groomed by a conspecific). All durations were weighted to consider differences in the total observation time per subject. We also reported the age of each subject in months.

Because the inherent nature of vocal exchanges is dyadic, we also calculated these four indices at the dyadic level to study if there are any variations in the exchanges based on the partners involved. For each possible within-group dyad (e.g., AB), we calculated the following: the hierarchical difference (difference between the dominance indices of individuals A and B, bidirectional index: $A \leftrightarrow B$), the spatial proximity (cumulative duration of observations when A and B were within each other's arm's reach, bidirectional index), the social grooming (cumulative time within A and B observations when A groomed B, unidirectional index: $A \rightarrow B$), and the age difference (in months, bidirectional index). We always subtracted the lowest score from the highest for bidirectional indices to maintain positive values.

Data analysis

Statistics analyses were conducted using R software version 4.1.2 (R Core Team, 2021) with a significance threshold set at 0.05.

First, we defined vocal interactions for mangabeys. Therefore, we plotted a frequency distribution graph with all occurrences of time intervals between two calls (i.e., two consecutive calls emitted by two different individuals) within sequences and between the last and the first calls of two consecutive sequences. We then graphically determined (by visual inspection of the graph) the vocal response threshold time characterizing a call interaction and defined three vocal patterns: isolated calling, vocal interactions with call

overlap, and vocal interactions based on turn-taking, i.e., without call overlap (methods notably previously used by Briseno-Jaramillo et al., 2018; Lemasson et al., 2018; Masataka and Biben, 1987). We then statistically tested whether the contact calls were principally involved in these temporally-coordinated vocal interactions compared to others call types. To test for the homogeneity in call type contributing to calling patterns, we ran a generalized linear mixed model fit by maximum likelihood with a negative binomial distribution (model with Poisson distribution was overdispersed) and a log link function using `glmer.nb` function of “MASS” package version 7.3.54 (Venables and Ripley, 2002): call type and pattern as fixed effects, number of calls as the dependent variable, and caller identity as a random factor. We analyzed this by running a type II ANOVA with a Wald chi-square test (using the `Anova` function of “rstatix” package version 0.7.0; (Kassambara, 2021). We did post hoc tests with a contrast table to explore a few chosen comparisons—which will be detailed in the results section—with FDR adjustments for p-value using the contrast function of the “emmeans” package version 1.7.2 (Lenth, 2022).

Second, once the vocal interaction was defined, we could extract all dyadic interactions of contact calls (named here vocal exchange). For example, a vocal exchange with three consecutive callers, A and B and then C, was broken down into two dyadic exchanges AB and BC. It is possible that C also responds to A in this case, but since it is difficult to know, we ruled out this possibility. However, a previous study showed that the results were quite similar when considering all possible dyads (AB, BC, and AC) than when focusing only on direct successions (only AB and BC) (Lemasson et al., 2018). These dyadic exchanges were used to assess the individual contributions to such vocal interactions (i.e., the number of dyadic exchanges to which a given individual appeared/contributed) and their positions in the interaction. For this latter measure, we counted the number of times that individual A was in first position (i.e., AB + AC + AD...) minus the number of times individual A was in second position (i.e., BA + CA + DA...) divided by the total number of dyadic exchanges involving A. This calculation gave an index that was negative when the individual was more likely to respond to a call and positive when he was more likely to trigger a response. Then, we measured the average response time to each one of the other group members for each individual.

Third, using correlation tests, we investigated the possible relationship between these three variables (i.e., contribution to exchanges, position in exchanging dyads and response time) and the aforementioned social individual and dyadic indices. To test for the influence of individual characteristics on vocal behavior, the number of calls emitted was divided by the number of focal sessions since sampling efforts differ between groups. Non-parametric

tests were used when the conditions for using parametric ones were not met. Before that, we confirmed that individual indices were not correlated, and we did the same for dyadic indices.

Results

Temporal calling patterns and definition of a vocal exchange

The graphical analysis of the distribution of inter-individual inter-call delays allowed us to set a maximum response time threshold defining a vocal interaction at 2 s (Fig. 1). Beyond 2 s, two consecutive calls could no longer be considered part of the same interaction. We could also distinguish interactions with call overlap (negative response time) and without call overlap ($0 < \text{response time} < 2\text{s}$): the latter being an interaction based on the turn-taking principle. The contribution of the different call types to the three calling patterns was not random (Table 2). We tested the different contributions of four call types, Ro+(Uh), Ti+(Uh), Un+(Uh), and Wahoo (others being produced too infrequently for statistical analysis), to the different calling patterns. We found a significant interaction between the two factors (GLMM: call types \times calling pattern interaction, $\chi^2=14.767$, $p=0.02$). To test our prediction about contact calls being more frequently used in interactions, we focused on the Ro+(Uh) call type to do our post hoc comparisons using a contrast table. We showed that (i) there is no significant difference between the occurrence of Ro+(Uh) or other call types in interactions with overlap (GLMM: estimate= -0.988, SE= 1.899, $z\text{-ratio}=-0.520$, $p=0.6$); (ii) there is significantly more Ro+(Uh) than other call types in turn-taking based interactions (GLMM: estimate= 6.048, SE= 1.284, $z\text{-ratio}= 4.709$, $p<0.0001$) and in isolated calls (GLMM: estimate= 3.925, SE= 1.648, $z\text{-ratio}= 2.383$, $p=0.02$), and (iii) there are significantly more Ro+(Uh) in turn-taking based interactions than in other temporal patterns (GLMM: estimate= 6.803, SE= 0.991, $z\text{-ratio}= 3.864$, $p<0.0001$). We thus defined a turn-taking-based vocal exchange as a sequence of Ro+(Uh) calls (i.e., contact calls) emitted by two or more individual callers within a 0-to-2-s inter-call interval.

According to this definition, we could identify 62 vocal exchanges. They most frequently involved only 2 individuals (only 5 exchanges involved more than 2 individuals, none involved more than 4 individuals) at a time. The number of calls in these exchanges ranged from 2 to 16, with a median of 3 calls per exchange. These vocal sequences could then be broken down into 115 dyadic exchanges to examine the possible effect of social factors.

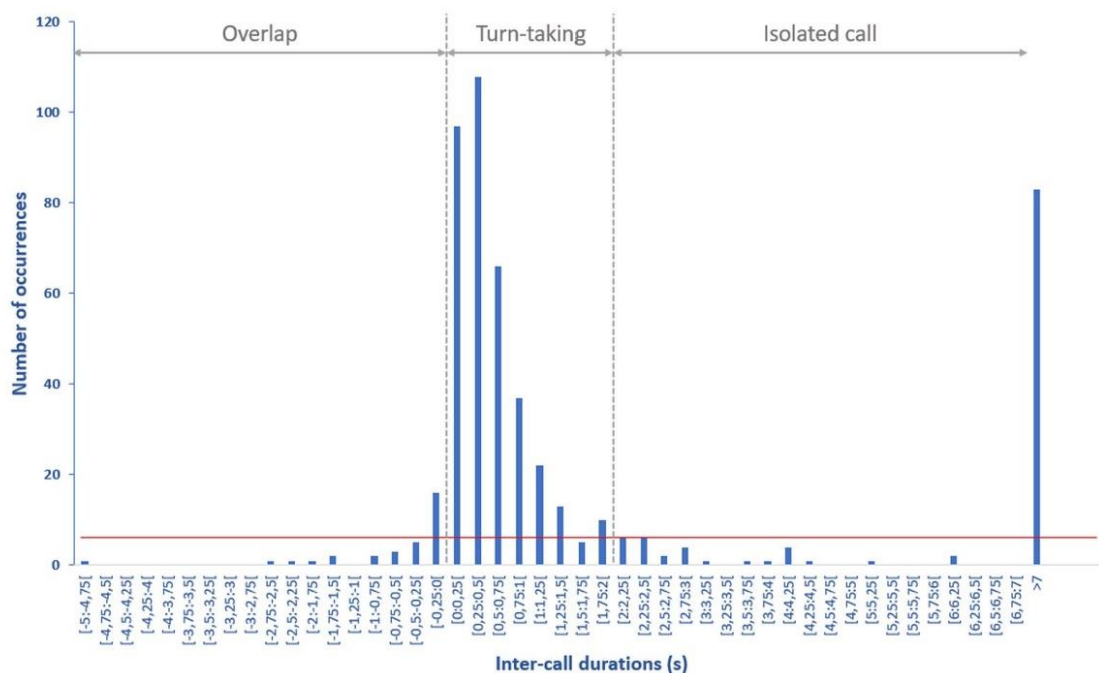


Fig. 1 Distribution of inter-call intervals. The red line shows the baseline when two calls are emitted independently from one another. The dotted lines mark the three temporal patterns

Influence of the caller's characteristics on exchange rate, exchange position, and response delay

Individual characteristics did not clearly relate to their propensity to contribute to contact call exchanges. We did not find any correlation between the dominance, spatial integration, or social integration status of individuals with the number of dyadic exchanges in which they were involved (dominance: spearman test, $rs=-0.05, p=0.85$; spatial integration: Spearman test, $rs=0.05, p=0.85$; social integration:

spearman test, $rs=0.32, p=0.17$). We found a trend for age, with older individuals tending to exchange more (Spearman test, $rs=0.43, p=0.06$).

However, when we considered individual's positions within exchanges rather than their general involvement, we found a clear link of dominance (Spearman test, $rs=-0.52, p=0.02$), with higher ranked individuals being more frequently found in response positions (Fig. 2). There was no link of other positional characteristics of an individual to the exchanges (age: Pearson test, $rs=-0.26, p=0.27$; spatial integration: Pearson test, $rs=0.01, p=0.96$; social integration: Spearman test, $rs=-0.02, p=0.93$).

Regarding the average response time of each individual, the more socially integrated an individual was, the slower its response (Spearman test, $rs=0.722, p=0.001$; Fig. 2). Age, dominance, and spatial integration, however, did not influence this parameter (age: Pearson test, $rs=0.38, p=0.13$; dominance: Pearson test, $rs=-0.11, p=0.69$; spatial integration: Pearson test, $rs=0.16, p=0.54$).

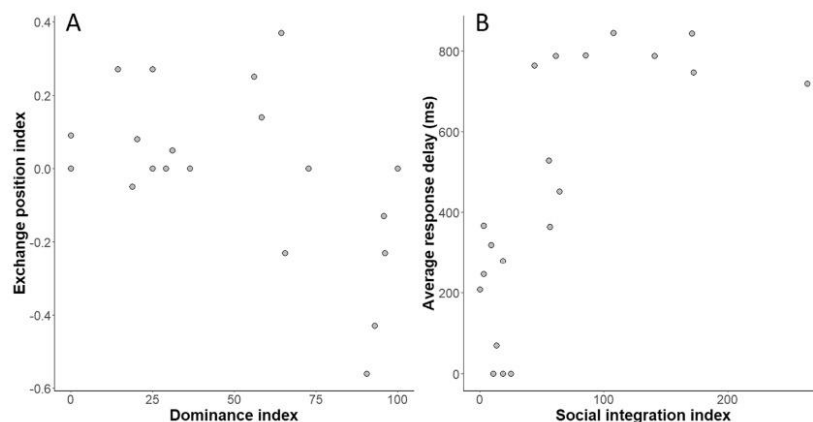
Table 2 Distribution of call types per calling pattern

Call type	Overlapped calls	Calling pattern Turn-taking based interactions	Isolated calls
Ro+(Uh)	6	188	55
Ti+(Uh)	12	94	27
Un+(Uh)	9	54	10
WaHoo	4	19	11
A+(Uh)	0	1	1
Oe+	1	0	1
Whoop-Gobble	0	1	0
Wi	0	0	0

Influence of the dyads' characteristics on exchange rate and response delay

Age difference and spatial proximity were positively correlated with the number of exchanges between two individuals

Fig. 2 Relationship between individual characteristics (dominance index or social integration index, **A** and **B** respectively) and its implication in exchanges (position index or response delay, **A** and **B** respectively). In **A**, when the index has a positive value, the individual was more often in the first position within exchanges (i.e., initiated the exchange); when the index is negative, the individual responded more often to a previous vocalization. Each dot represents an individual



(age difference: Spearman test, $r_s=0.31$, $p=0.003$; spatial proximity: Spearman test, $r_s=0.31$, $p=0.003$). The greater the age difference between two individuals, or the more time they spent close to one another (Fig. 3), the more they vocally exchanged. However, there was no correlation between hierarchical difference or social affinity with the number of exchanges made between two group members (hierarchical difference: Spearman test, $r_s=-0.02$, $p=0.84$; social affinity: Spearman test, $r_s=0.13$, $p=0.21$).

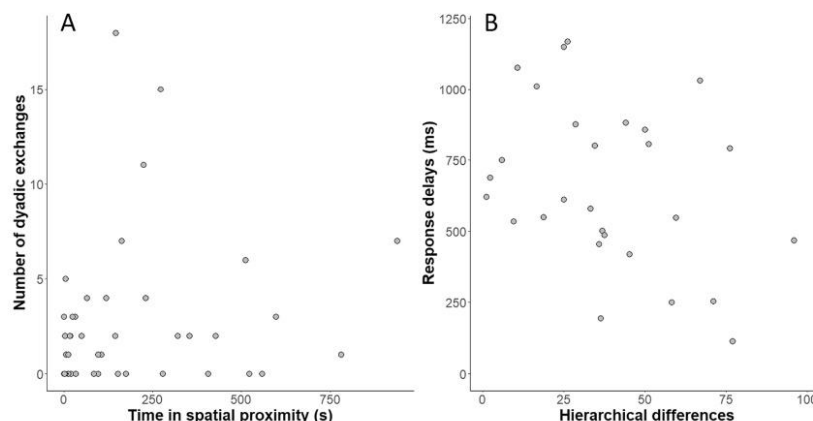
The hierarchical difference was negatively correlated with response delays (Spearman test, $r_s=-0.43$, $p=0.003$). The further away a member was from another one in the group hierarchy, the shorter the response time (Fig. 3). However, there was no correlation between age difference, spatial proximity, and social affinity with the response delays (age difference: Spearman test, $r_s=0.11$, $p=0.45$; spatial proximity: Spearman test, $r_s=-0.15$, $p=0.34$; social affinity: Spearman test, $r_s=-0.006$, $p=0.97$).

Discussion

Red-capped mangabeys, a semi-tolerant species, exchange, as predicted, turn-taking-based contact calls in a non-random and socially predictive way. Indeed, the red-capped mangabeys respond in a time window that does not exceed 2 s and seem to avoid overlapping as illustrated by the gap in the frequencies between inter-call interval $[-0.25s;0s]$ (i.e., overlapping vocalizations) and inter-call interval $[0s;0.25s]$ (i.e., turn-taking based vocalization) (Fig. 1). The social status of callers, consisting of parameters reflecting both affinity (cf. tolerant society) and hierarchy (cf. despotic society), affects the response rate, the position in the exchange (initiator or respondent), and the response delay.

In our study, individuals exchanged more with the close-proximity partners with whom they spent the most time rather than their preferred grooming partners. Although these two behaviors (i.e., spatial proximity and grooming)

Fig. 3 Relationship between dyadic characteristics (spatial proximity or hierarchical differences, **A** and **B** respectively) and its implication in exchanges (number of exchanges or response delay, **A** and **B** respectively). Each dot represents a dyad



may both indicate an affiliative social relationship, they are not necessarily equivalent (Dunbar and Shultz, 2010; Silk et al., 2013). Indeed, in some studies, grooming behaviors are closely related to vocal exchanges (Arlet et al., 2015; Kulaheci et al., 2015; Ramos-Fernández, 2005), while in other studies, spatial proximity is a better predictor of the exchange rate between two given individuals (Lemasson et al., 2018; Levréro et al., 2019). Rather than being due to a methodological issue, we believe that social grooming and spatial proximity have different affinity values according to the species. Even in studies where both proximity and grooming indices have been used, they correlate to vocal exchange rates differently in different species (Arlet et al., 2015; Lemasson et al., 2018). In the same vein, several studies on other mammalian species highlighted no correlation between the networks of spatial proximity and socio-positive behaviors (e.g., pigs: Camerlink et al., 2022; horses: Wolter et al., 2018). Treves and Baguma (2002) showed that spatial proximity better predicted social organizations in several non-human primate species than social behaviors such as social grooming. Compared to many rarer but more visible behaviors, spatial proximity would thus be a subtle but more reliable behavioral measure of social bonding for many taxa because it is common and continuous (Dunbar and Shultz, 2010). Social grooming is, moreover, subject to ambiguity for several primate species. Indeed, Seyfarth (1977) proposed a model of social grooming in non-human female primates that was later confirmed by a meta-analysis (Schino, 2001); the “grooming up the hierarchy model” is based on the fact that there is competition between members of the same group for the grooming of higher ranked individuals, as well as a preference for grooming related individuals. Other authors report that social grooming can be exchanged as a commodity whose value depends on the local food context (whether there is competition or not), which they, therefore, compare to market prices (Barrett et al., 1999). This evidence shows that motivations other than affinity may underlie this behavior.

We also found an influence of age in this study, with individuals exchanging more grunts with partners of different ages. Age is an important social marker in non-human primates, and many studies have shown its influence on vocal exchanges. For instance, in both Campbell’s monkeys and common marmosets, older individuals elicit more vocal responses, showing that they are preferred exchange partners for the group as a whole (Chen et al., 2009; Lemasson et al., 2010a).

Our study also highlighted that the order of callers within the exchange was not random and was influenced by the hierarchical status of the individuals involved. A collaborative study using data on 26 non-human primate species showed that the influence of hierarchy on vocal behavior is well established. Dominant individuals who were tolerant

with their subordinate partners vocalized at a higher rate than dominant despotic individuals, showing that more tolerant individuals have a greater need to communicate to manage their dominance in a group (Kavanagh et al., 2021). In our study, the higher an individual was in the hierarchy, the more he or she was in a position to respond rather than to initiate an exchange. This similar exchange pattern was found in western lowland gorillas (Lemasson et al., 2018). This similarity between mangabeys and gorillas is consistent with the results of the collaborative study mentioned above, which established an equivalent dominance index for these two species directly in the middle of the tolerant/despotic spectrum (see fig 4 in Kavanagh et al., 2021). Therefore, we can infer that this pattern is expected in species living in semi-tolerant/semi-despotic social systems.

Finally, our results reveal that the response delay was influenced by social factors at both the individual and dyadic levels. On the one hand, the more socially integrated an individual, the longer its average response time; conversely, the more significant the hierarchical gap between the two callers, the shorter the response time between their calls. This variation can be explained in several non-mutually exclusive ways, the first being stress. Indeed, stress is an internal factor strongly related to vocal production, as evidenced by the study of Lemasson et al. (2010b), which established that the more a primate is confronted with a stressful situation, the higher the number of calls in a sequence. In a society with despotic characteristics, such as the red-capped mangabeys, the least socially integrated individuals are those with a higher stress baseline (Brent et al., 2011; Crockford et al., 2008). A higher stress baseline may induce a shorter reaction time to various contextual factors, including contact calls. This fact could better explain why they have a shorter response time on average and why the response time is shorter when a subordinate answers to a dominant. However, this does not explain why a dominant would answer more quickly to a subordinate than to another dominant. Another hypothesis to explain variations in response time is because responding quickly prevents another individual from responding. In songbirds, for instance, Henry et al. (2015) showed that individuals adjust the spacing between their consecutive songs to allow more or less room for others to take turns. In our study, a shorter response time for less socially integrated individuals and hierarchically distant callers could thus be viewed as individuals rushing to respond to a vocalization before someone else does in order to gain a social advantage in establishing, securing a relationship or to appease the partner before an interaction. A less integrated and subordinate individual could be highly motivated to answer quickly, in general, to jump at the chance to social bond with others, especially with highly dominant ones. From a dominant’s perspective, rushing to respond to a vocalization may be a way for the dominant

respondent to monopolize the floor to prevent one individual from linking up with another and avoid long-term coalition building. A final explanation, especially for cases of response latency close to zero, can be that responding quickly may be an attempt to cut off the initiator's voice. Overlapping is associated with aggressiveness and strong personality in humans (ter Maat et al., 2010), as well as dominance in songbirds (Baker et al., 2012). Consequently, responding quickly to a subordinate when at the top of the hierarchy is a less costly way to demonstrate dominance than aggression or intimidation.

We acknowledge the somewhat exploratory nature of our study, and we are aware that the small size of groups in captivity, as well as the artificial composition (more or less recent) of groups, does not allow us to describe all the social uses of contact calls in a species. Nevertheless, the significant number of individuals represented in our analysis, from which interesting correlations arise, opens doors to new research perspectives on the social function of vocal exchange in primates. Studies on free-ranging and large multi-male, multi-female groups are still necessary to confirm and extend these first findings. This would allow us to investigate potential sex-related roles during the exchange of contact calls. Further analyses for polyadic exchanges may also be interesting, but a larger sample size is needed. Finally, studies using playbacks of exchanges broadcast in captivity and/or in the wild are also crucial to better understand the extent to which red-capped mangabeys are aware of these exchange patterns and may use the potential information coded in these interactions in a strategic manner. Beyond these limitations, our study offers a valuable contribution to the literature, given the limited number of studies on primates' "proto-conversations" and the limited knowledge about vocal communication in mangabey species (Snowdon and Cleveland, 1984).

Beyond the debate on the function of contact calls, the study of turn-taking exchanges allows us to understand the social strategies of non-human primates better. There is a theoretical convergence between the study of call exchanges in non-human primates and the conversation analysis conducted by sociologists. These two fields seek to explain how group-level communicative phenomena emerge from individual-level behavioral mechanisms (Logue and Stivers, 2012). Turn-taking exchanges are thus at the foundation of all human institutions (Goodwin and Heritage, 1990) since it forms the core of all human interactions, verbal or not. Language, however, offers unparalleled flexibility in transmitting information and especially intentions, making parallels between the functioning of human societies and societies of non-human primates difficult. Nevertheless, the latter is embedded in an interactive engine defined by several elements that are all present as primordia in non-human primates: (i) in the multimodality of signals (joint use of

gestures, vocalizations, gazes, and facial displays), (ii) in the control of timing (short delay of responses), (iii) in the contingencies between initiations and responses, and (iv) in the sharing of intentions via communicative acts (the theory of mind) (Levinson, 2022). The results of this study thus illustrate how animals without language can instead exploit a simple aspect of their communication to elicit a plurality of meanings and functions. They offer interesting insights into the ability of red-capped mangabeys to control the timing of a vocalization to manage their social position in their group.

Acknowledgements Our main thanks go to the animal caretakers in charge of the animals studied (Arnaud Rossard, Philippe Bec, Céline Nicolle, Marine Mergen).

Funding This research received funds from Rennes 1 University, National Center for Scientific Research (C.N.R.S.), French University Institute (I.U.F.) and from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 788077, Orisem, PI: Schlenker). Research was conducted in part at Institut d'Études Cognitives, Ecole Normale Supérieure – PSL Research University, supported by grants ANR-10-IDEX-0001-02 and FrontCog ANR-17-EURE-0017.

Declarations

Conflict of interest The authors declare no competing interests.

References

- Altmann J (1974) Observational study of behavior: sampling methods. *Behaviour* 49:227–266. <https://doi.org/10.1163/156853974X00534>
- Arlet M, Jubin R, Masataka N, Lemasson A (2015) Grooming-at-a-distance by exchanging calls in non-human primates. *Biol Lett* 11:20150711–20150714. <https://doi.org/10.1098/rsbl.2015.0711>
- Aychet J, Monchy N, Blois-Heulin C, Lemasson A (2022) Context-dependent gestural laterality: a multifactorial analysis in captive red-capped mangabeys. *Animals* 12:186. <https://doi.org/10.3390/ani12020186>
- Baker TM, Wilson DR, Mennill DJ (2012) Vocal signals predict attack during aggressive interactions in black-capped chickadees. *Anim Behav* 84:965–974. <https://doi.org/10.1016/j.anbehav.2012.07.022>
- Barrett L, Henzi SP, Weingrill T, Lycett JE, Hill RA (1999) Market forces predict grooming reciprocity in female baboons. *Proc R Soc Lond B* 266:665–670. <https://doi.org/10.1098/rspb.1999.0687>
- Bouchet H, Pellier A-S, Blois-Heulin C, Lemasson A (2010) Sex differences in the vocal repertoire of adult red-capped mangabeys (*Cercocebus torquatus*): a multi-level acoustic analysis. *Am J Primatol* 72:360–375. <https://doi.org/10.1002/ajp.20791>
- Brent L, Semple S, Dubuc C, Heistermann M, MacLarnon A (2011) Social capital and physiological stress levels in free-ranging adult female rhesus macaques. *Physiol Behav* 102:76–83. <https://doi.org/10.1016/j.physbeh.2010.09.022>
- Briseno-Jaramillo M, Ramos-Fernández G, Palacios-Romo TM, Sosa-López JR, Lemasson A, Sosa-López JR, Lemasson A (2018) Age and social affinity effects on contact call interactions in free-ranging spider monkeys. *Behav Ecol Sociobiol* 72:192. <https://doi.org/10.1007/s00265-018-2615-2>

- Briseño-Jaramillo M, Berthet M, Estrada A, Biquand V, Lemasson A (2021) Socially mediated overlap in vocal interactions between free-ranging black howler monkeys. *Am J Primatol* 83:e23297. <https://doi.org/10.1002/ajp.23297>
- Briseño-Jaramillo M, Sosa-López José R, Ramos-Fernández G, Lemasson A (2022) Flexible use of contact calls in a species with high fission-fusion dynamics. <https://doi.org/10.1101/2022.05.25.492876>
- Byrne R, Teixidor P (1999) The “whinny” of spider monkeys: individual recognition before situational meaning. *Behaviour* 136:279–308. <https://doi.org/10.1163/156853999501333>
- Camerlink I, Scheck K, Cadman T, Rault J-L (2022) Lying in spatial proximity and active social behaviours capture different information when analysed at group level in indoor-housed pigs. *Appl Anim Behav Sci* 246:105540. <https://doi.org/10.1016/j.applanim.2021.105540>
- Candiotti A, Zuberbühler K, Lemasson A (2012) Convergence and divergence in Diana monkey vocalizations. *Biol Lett* 8:382–385. <https://doi.org/10.1098/rsbl.2011.1182>
- Chen HC, Kaplan G, Rogers LJ (2009) Contact calls of common marmosets (*Callithrix jacchus*): influence of age of caller on antiphonal calling and other vocal responses. *Am J Primatol* 71:165–170. <https://doi.org/10.1002/ajp.20636>
- Crockford C, Wittig RM, Whitten PL, Seyfarth RM, Cheney DL (2008) Social stressors and coping mechanisms in wild female baboons (*Papio hamadryas ursinus*). *Horm Behav* 53:254–265. <https://doi.org/10.1016/j.yhbeh.2007.10.007>
- Digweed SM, Fedigan LM, Rendall D (2007) Who cares who calls? Selective responses to the lost calls of socially dominant group members in the white-faced capuchin (*Cebus capucinus*). *Am J Primatol* 69:829–835. <https://doi.org/10.1002/ajp.20398>
- Dolado R, Beltran FS (2012) Emergent patterns of social organization in captive *Cercocebus torquatus*: testing the GrooFiWorld agent-based model. *J Biosci* 37:777–784. <https://doi.org/10.1007/s12038-012-9231-5>
- Dolado R, Cooke C, Beltran FS (2016) How many for lunch today? Seasonal fission-fusion dynamics as a feeding strategy in wild red-capped mangabeys (*Cercocebus torquatus*). *Folia Primatol* 87:197–212. <https://doi.org/10.1159/000449220>
- Dunbar RIM (1998) The social brain hypothesis. *Evol Anthropol* 6:178–190
- Dunbar RIM, Shultz S (2010) Bondedness and sociality. *Behaviour* 147:775–803
- Goodwin C, Heritage J (1990) Conversation Analysis. *Annu Rev Anthropol* 19:283–307
- Gouzoules S, Gouzoules H, Marler P (1984) Rhesus monkey (*Macaca mulata*) screams: representational signalling in the recruitment of agonistic aid. *Anim Behav* 32:182–193
- Henry L, Craig AJFK, Lemasson A, Hausberger M (2015) Social coordination in animal vocal interactions. Is there any evidence of turn-taking? The starling as an animal model. *Front Psychol* 6:1416. <https://doi.org/10.3389/fpsyg.2015.01924>
- Kavanagh E, Street SE, Angwela FO, Bergman TJ, Blaszczyk MB, Bolt LM, Briseño-Jaramillo M, Brown M, Chen-Kraus C, Clay Z, Coye C, Thompson ME, Estrada A, Fichtel C, Fruth B, Gamba M, Giacomini C, Graham KE, Green S et al (2021) Dominance style is a key predictor of vocal use and evolution across nonhuman primates. *R Soc Open Sci* 8:210873. <https://doi.org/10.1098/rsos.210873>
- Kondo N, Watanabe S (2009) Contact calls: Information and social function. *Jpn Psychol Res* 51:197–208. <https://doi.org/10.1111/j.1468-5884.2009.00399.x>
- Kudo H (1987) The study of vocal communication of wild mandrills in Cameroon in relation to their social structure. *Primates* 28:289–308. <https://doi.org/10.1007/BF02381013>
- Kulahci IG, Rubenstein DI, Ghazanfar AA (2015) Lemurs groom-at-a-distance through vocal networks. *Anim Behav* 110:179–185. <https://doi.org/10.1016/j.anbehav.2015.09.016>
- Lemasson A, Hausberger M (2004) Patterns of vocal sharing and social dynamics in a captive group of Campbell’s monkeys (*Cercocebus campbelli campbelli*). *J Comp Psychol* 118:347–359. <https://doi.org/10.1037/0735-7036.118.3.347>
- Lemasson A, Gandon E, Hausberger M (2010a) Attention to elders’ voice in non-human primates. *Biol. Lett.* 6:325–328. <https://doi.org/10.1098/rsbl.2009.0875>
- Lemasson A, Ouattara K, Bouchet H, Zuberbühler K (2010b) Speed of call delivery is related to context and caller identity in Campbell’s monkey males. *Naturwissenschaften* 97:1023–1027. <https://doi.org/10.1007/s00114-010-0715-6>
- Lemasson A, Glas L, Barbu S, Lacroix A, Guilloux M, Remeuf K, Koda H (2011a) Youngsters do not pay attention to conversational rules: is this so for nonhuman primates? *Sci Rep* 1:12–15. <https://doi.org/10.1038/srep00022>
- Lemasson A, Ouattara K, Petit EJ, Zuberbühler K (2011b) Social learning of vocal structure in a nonhuman primate? *BMC Evol Biol* 11:1–7. <https://doi.org/10.1186/1471-2148-11-362>
- Lemasson A, Guilloux M, Barbu S, Lacroix A, Koda H (2013) Age- and sex-dependent contact call usage in Japanese macaques. *Primates* 54:283–291. <https://doi.org/10.1007/s10329-013-0347-5>
- Lemasson A, Pereira H, Levréro F (2018) Social basis of vocal interactions in western lowland gorillas (*Gorilla g. gorilla*). *J Comp Psychol* 132:141–151. <https://doi.org/10.1037/com0000105>
- Levinson SC (2022) The interaction engine: cuteness selection and the evolution of the interactional base for language. *Philos Trans R Soc* 377:20210108. <https://doi.org/10.1098/rstb.2021.0108>
- Levréro F, Touitou S, Frédet J, Nairaud B, Guéry J-P, Lemasson A (2019) Social bonding drives vocal exchanges in Bonobos. *Sci Rep* 9:711. <https://doi.org/10.1038/s41598-018-36024-9>
- Logue DM, Stivers T (2012) Squawk in interaction: A primer of conversation analysis for students of animal communication. *Behaviour* 149:1283–1298. <https://doi.org/10.1163/1568539X-00003031>
- Macedonia JM, Evans CS (1993) Essay on contemporary issues in ethology: variation among mammalian alarm call systems and the problem of meaning in animal signals. *Ethology* 93:177–197. <https://doi.org/10.1111/j.1439-0310.1993.tb00988.x>
- Maestriperi D, Roney JR (2005) Primate copulation calls and post-copulatory female choice. *Behav Ecol* 16:106–113. <https://doi.org/10.1093/beheco/arl120>
- Masataka N, Biben M (1987) Temporal rules regulating affiliative vocal exchanges of squirrel monkeys. *Behaviour* 101:311–319
- Neumann C, Zuberbühler K (2016) Vocal correlates of individual sooty mangabey travel speed and direction. *PeerJ* 4:e2298. <https://doi.org/10.7717/peerj.2298>
- Pika S, Wilkinson R, Kendrick KH, Vernes SC (2018) Taking turns: bridging the gap between human and animal communication. *Proc Royal Soc B* 285:20180598. <https://doi.org/10.1098/rspb.2018.0598>
- Pougnault L, Levréro F, Lemasson A (2020a) Conversation among primate species. In: Masataka N (ed) *The origins of language revisited: differentiation from music and the emergence of neurodiversity and autism*. Springer, Singapore, pp 73–96. https://doi.org/10.1007/978-981-15-4250-3_4
- Pougnault L, Levréro F, Mulot B, Lemasson A (2020b) Breaking conversational rules matters to captive gorillas: a play-back experiment. *Sci Rep* 10:6947. <https://doi.org/10.1038/s41598-020-63923-7>
- Pougnault L, Levréro F, Leroux M, Paulet J, Bombani P, Dentressangle F, Deruti L, Mulot B, Lemasson A (2022) Social pressure drives “conversational rules” in great apes. *Biol Rev Camb Philos Soc* 97:749–765. <https://doi.org/10.1111/brv.12821>

- R Core Team (2021) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria <http://www.R-project.org/>
- Ramos-Fernández G (2005) Vocal communication in a fission-fusion society: do spider monkeys stay in touch with close associates? *Int J Primatol* 26:1077–1092. <https://doi.org/10.1007/s10764-005-6459-z>
- Range F, Fischer J (2004) Vocal repertoire of sooty mangabeys (*Cercocebus torquatus atys*) in the Tai National Park. *Ethology* 110:301–321. <https://doi.org/10.1111/j.1439-0310.2004.00973.x>
- Rendall D, Owren MJ (2002) Animal vocal communication: say what?, in: The cognitive animal: empirical and theoretical perspectives on animal cognition. MIT Press, Cambridge, MA, US, pp 307–313. <https://doi.org/10.7551/mitpress/1885.001.0001>
- Schino G (2001) Grooming, competition and social rank among female primates: a meta-analysis. *Anim Behav* 62:265–271. <https://doi.org/10.1006/anbe.2001.1750>
- Seyfarth RM (1977) A model of social grooming among adult female monkeys. *J Theor Biol* 65:671–698. [https://doi.org/10.1016/0022-5193\(77\)90015-7](https://doi.org/10.1016/0022-5193(77)90015-7)
- Silk J, Cheney D, Seyfarth R (2013) A practical guide to the study of social relationships. *Evol Anthropol* 22:213–225. <https://doi.org/10.1002/evan.21367>
- Slocombe KE, Zuberbühler K (2006) Food-associated calls in chimpanzees: responses to food types or food preferences? *Anim Behav* 72:989–999. <https://doi.org/10.1016/j.anbehav.2006.01.030>
- Snowdon CT, Cleveland J (1984) “Conversations” among pygmy marmosets. *Am J Primatol* 7:15–20. <https://doi.org/10.1002/ajp.1350070104>
- Snowdon CT, Elowson AM (1999) Pygmy marmosets modify call structure when paired. *Ethology* 105:893–908. <https://doi.org/10.1046/j.1439-0310.1999.00483.x>
- Snowdon CT, Hodun A (1981) Acoustic adaptations in pygmy marmoset contact calls: locational cues vary with distances between conspecifics. *Behav Ecol Sociobiol* 7:295–300
- Sperber AL, Werner LM, Kappeler PM, Fichtel C (2017) Grunt to go—vocal coordination of group movements in redfronted lemurs. *Ethology* 123:894–905. <https://doi.org/10.1111/eth.12663>
- Stivers T, Enfield NJ, Brown P, Englert C, Hayashi M, Heinemann T, Hoymann G, Rossano F, de Ruiter JP, Yoon K-E, Levinson SC (2009) Universals and cultural variation in turn-taking in conversation. *Proc Natl Acad Sci USA* 106:10587–10592. <https://doi.org/10.1073/pnas.0903616106>
- Sugiura H (1993) Temporal and acoustic correlates in vocal exchange of coo calls in Japanese macaques. *Behaviour* 124:3–4. <https://doi.org/10.1163/156853993X00588>
- Sugiura H (2007) Adjustment of temporal call usage during vocal exchange coo calls in Japanese macaques. *Ethology* 113:528–533. <https://doi.org/10.1111/j.1439-0310.2007.01349.x>
- Sugiura H, Masataka N (1995) Temporal and acoustic flexibility in vocal exchanges of coo calls in Japanese macaques (*Macaca fuscata*). In: Zimmermann E, Newman JD, Jürgens U (eds) Current topics in primate vocal communication. Springer US, Boston, MA, pp 121–140. https://doi.org/10.1007/978-1-4757-9930-9_6
- Takahashi DY, Narayanan DZ, Ghazanfar AA (2013) Coupled oscillator dynamics of vocal turn-taking in monkeys. *Curr Biol* 23:2162–2168. <https://doi.org/10.1016/j.cub.2013.09.005>
- ter Maat M, Truong KP, Heylen D (2010) How turn-taking strategies influence users’ impressions of an agent. In: Allbeck J, Badler N, Bickmore T, Pelachaud C, Safonova A (eds) Intelligent virtual agents, lecture notes in computer science. Springer, Berlin Heidelberg, Berlin, Heidelberg, pp 441–453. https://doi.org/10.1007/978-3-642-15892-6_48
- Treves A, Baguma P (2002) Interindividual proximity and surveillance of associates in comparative perspective. In: Glenn ME, Cords M (eds) The guenons: diversity and adaptation in african monkeys, developments in primatology: progress and prospects. Springer US, Boston, MA, pp 161–172. https://doi.org/10.1007/0-306-48417-X_12
- Venables WN, Ripley BD (2002) Modern applied statistics with S, 4th ed. Springer, New York
- Wolter R, Stefanski V, Krueger K (2018) Parameters for the analysis of social bonds in horses. *Animals* 8:191. <https://doi.org/10.3390/ani8110191>
- Zumpe D, Michael RP (1986) Dominance index: a simple measure of relative dominance status in primates. *Am J Primatol* 10:291–300. <https://doi.org/10.1002/ajp.1350100402>
- Bouchet H, Blois-Heulin C, Lemasson A (2013) Social complexity parallels vocal complexity: a comparison of three non-human primate species. *Front Psychol* 4. <https://doi.org/10.3389/fpsyg.2013.00390>
- Kassambara, A., 2021. Pipe-friendly framework for basic statistical tests. R package version 0.7.0 [WWW Document]. URL <https://CRAN.R-project.org/package=rstatix> (accessed 7.19.22).
- Lenth, R.V., 2022. Estimated Marginal Means, aka Least-squares means. R package emmeans version 1.7.2. [WWW Document]. URL <https://CRAN.R-project.org/package=emmeans> (accessed 7.19.22).
- Zanoli A, Gamba M, Lemasson A, Norscia I, Palagi E (2021) Intersexual multimodal communication during mating in wild geladas: the leading role of females. *Curr Zool:zoab093*. <https://doi.org/10.1093/cz/zoab093>

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Article 5 :

Turn-taking in red-capped mangabey (*Cercocebus torquatus*): an experimental investigation of a conversation rule in a non-human primate species.

**Bastien Meunier¹, Virginie Durier¹, Arnaud Rossard¹, Alban
Lemasson^{1,2}**

¹ Univ Rennes, CNRS, Normandie Univ, EthoS (Éthologie animale et humaine) - UMR 6552, Rennes, France

² Institut Universitaire de France, 1 rue Descartes, 75231 Paris, Cedex 05, France

Synthèse de l'article 5

Contexte

Plusieurs espèces de primates non-humains font des échanges vocaux de cris de contact. Ces échanges vocaux suivent un pattern de tour de rôle similaire à celui que l'on retrouve dans les conversations humaines, qui consistent en l'alternance de tours entre plusieurs individus qui évitent les chevauchements et minimisent les silences. Si plusieurs espèces pratiquent ce pattern, la question de savoir s'il fait sens pour eux comme une règle reste à explorer.

Méthode

Nous avons réalisé une expérience de repasse chez 12 mangabeys à collier mâles captifs, des mangabeys connus pour échanger vocalement selon un tour de rôle. Nous avons repassé via haut-parleurs des échanges vocaux avec un temps de réponse habituel pour cette espèce (0.5s), des échanges vocaux avec un temps de réponse plus long (1.5s), et des échanges vocaux avec des cris qui se chevauchent. Nous avons mesuré la réaction des mangabeys en observant leurs regards 60s après la diffusion de chaque stimuli.

Résultats

Les individus testés n'ont pas réagi différemment aux différents stimuli repassés. Ils ont en outre manifesté un intérêt très fort pour chaque repasse, en regardant longtemps le haut-parleur après chacune d'elles. Néanmoins, les individus les plus vieux ont montré moins d'intérêt pour les échanges vocaux avec chevauchement que les jeunes. Cela suggère que si le tour de rôle n'est pas une règle chez cette espèce, l'expérience peut amener certains individus à être moins attentifs aux stimuli les moins pertinents socialement.

Turn-taking in red-capped mangabey (*Cercocebus torquatus*): an experimental investigation of a conversation rule in a non-human primate species.

Bastien Meunier¹, Virginie Durier¹, Arnaud Rossard¹, Alban Lemasson^{1,2}

¹ Univ Rennes, CNRS, Normandie Univ, EthoS (Éthologie animale et humaine) - UMR 6552, Rennes, France

² Institut Universitaire de France, 1 rue Descartes, 75231 Paris, Cedex 05, France

Abstract

Several non-human primate species engage in vocal exchanges of contact calls, throughout the day in peaceful contexts. These vocal exchanges have been compared to human conversations because vocalizations uttered according to a turn-taking: a temporal pattern where interlocutors minimize silences and avoid overlaps. Here, we conducted a playback experiment among red-capped mangabeys (*Cercocebus torquatus*) to investigate whether turn-taking was a known rule for the individuals who practiced it. We played back vocal exchanges with usual minimized response time (0.5 sec), vocal exchange with longer response time (1.5 sec) and vocal exchanges with call overlap. We measured and compared reaction of 12 captive males 60 sec after each playback conditions. The individuals tested did not react differently to different playback situations and were generally very interested in all playbacks. However, older individuals showed less interest in vocal exchanges with overlap than younger one. While these results did not reveal that red-capped mangabeys are aware of the turn-taking pattern underlying their vocal exchanges, they suggest that experience shape attention towards more socially relevant situations.

Introduction

Several patterns of vocal interactions exist in animals, the most frequent being choruses, duets, and antiphony, questioning their evolutionary trajectories. While choruses and duets have most of the times well-identified functions (e.g. mate attraction, territory defense) and involve one or few specific recurrent interlocutors (e.g. alliance between males for chimpanzee choruses, male-female pair-bonds for gibbon duets), antiphony is distinguished by a larger number of partners involved, more diverse emission contexts and an apparently flexible temporal structure (Yoshida and Okanoya, 2005). Among these antiphonal structures, we find in particular in non-human primates what several authors qualify as conversation-like vocal exchanges (Snowdon and Cleveland, 1984; Symmes and Biben, 1988; Pournault et al., 2020a). Indeed, these vocal exchanges typically concern only one specific call type of the species' vocal repertoire, the so-called "contact call", they do not occur at predictable times of the day or in the presence of a particular event but are emitted in several peaceful and affiliative contexts encountered throughout the day (Pournault et al., 2020a). Furthermore, even in species forming large social groups, a given exchange rarely involves more than two or three partners at a time, chosen for their long-term affinity, suggesting a "grooming-at-distance" function (Arlet et al., 2015; Kulaheci et al., 2015; Levréro et al., 2019). These contact call exchanges are therefore often described as possible precursors to human conversations (Dunbar, 1998; Pournault et al., 2020a), and the parallel is even more striking when it comes to their temporal structure qualified as "turn-taking" (Levinson, 2016).

Turn-taking in human conversations consists in the smooth exchange of short utterances between several speakers that self-select themselves to intervene while avoiding overlaps and minimizing gaps between turns (Sacks et al., 1974). This overlap avoidance and gap minimization mechanism is a universal feature of human language despite some cultural variation in the response time due to languages differences (Stivers et al., 2009), suggesting a biological basis. Turn-taking has thus been assessed in several non-human primates where the vast majority of individuals' vocal responses has been found to occur within one to three seconds (according to the species) after a third party's vocalization and to rarely overlap with it (Masataka and Biben, 1987; Briseno-Jaramillo et al., 2018; Lemasson et al., 2010, 2018; Levréro et al., 2019; Meunier et al., 2023). Furthermore, this temporal structure can vary within a species. For example, in squirrel monkeys, response times are shorter between two individuals with high affinity compared to individuals with low affinity (Biben et al., 1986; Masataka and Biben, 1987). In Japanese macaques, response times can vary at two levels: within a group

depending on the physical distance between vocal partners (Sugiura, 2007), and between different groups suggesting a cultural transmission (Sugiura and Masataka, 1995).

Beyond a temporal pattern of vocal interaction, turn-taking could be a rule as in humans, implicitly known by the interactants and socially learned. Indeed, human turn-taking is learned early on during ontogeny by infants through the daily interactions they have with their parents (Bateson, 1975; Hilbrink et al., 2015; Casillas et al., 2016). This learning is fundamental because individuals who break this rule are perceived negatively by their peers (Robinson and Reis, 1989; ter Maat et al., 2010), which could impair their social relationships (Hazen and Black, 1989). This may also be true for nonhuman primates. For example in marmosets, young individuals fail to use the appropriate type of vocalization for affiliative exchanges and interrupt others more than adults do (Chow et al., 2015). This tendency decreased after a few months, with parent-guided social learning as they responded more to appropriately timed and chosen offspring vocalizations. A "violation-of-expectation" experimental paradigm can help demonstrate the relevance of a rule for the animal and thus the knowledge of this rule at a certain age: when an animal is aware of a rule, it should present different levels of attention between a situation of violation and a situation of respect for the rule. In this line, two studies already showed that important features of turn-taking pattern act as a rule. Indeed, Lemasson et al (2011) highlighted that adults Campbell's monkeys were more interested in exchanges with alternation (i.e. two interlocutors A and B call in turn to form ABA exchanges) than in exchanges with failed alternation (i.e. artificially rearranged BAA, consecutive calling from the same individuals being rare). Moreover, youngsters did not present such interest difference during the same experiment, and were observed to interrupt more than adults during observations of spontaneous calling, suggesting that this rule need social learning to be mastered. Using the same paradigm, Pougnault et al. (2020b) highlighted that overlap avoidance matters in adults' gorillas as they paid less attention to the play-back of a vocal exchange with an overlap than to one that respect the species-specific response time. The intensity of the subject response also varied with subjects' age, supporting again a possible role of learning.

These few studies, although interesting, remain rare and more comparative studies on a larger diversity of species is now needed. Red-capped mangabeys (*Cercocebus torquatus*) are interesting in this perspective for several reasons. This species has a well-known vocal repertoire and produce conversation-like contact call exchanges (Bouchet et al., 2010). A recent study specifically described the observed delay in vocal response within these exchanges (Meunier et al., 2023). It showed firstly that response time varies between 0 and 2 seconds with

a clear peak around 0.5 second; secondly that response time is flexible and socially-guided (i.e. varying according to the social integration and hierarchical positions of the individuals). Moreover, this species forms multi-female multi-male groups with both sexes producing contact call exchanges (Bouchet et al., 2010). Therefore, the temporal structure of these exchanges might be relevant as a conversational rule whatever the sex of the individuals involved. Here, we conducted an experimental study in which we tested the reaction of male mangabeys of various ages to playback of vocal exchanges between females while varying the temporal structure of this exchange. Three temporal structures were compared: a vocal exchange not violating any expectation with a usual response time (0.5 sec), a vocal exchange with turn-taking violation (call overlap) and a vocal exchange with an abnormally long delay (1.5 sec) to test for the first time the sensitivity to longer response times than usual. Our expectations were threesome as we expected subjects' attention to vary according to the presence or not of an overlap, to the presence or not of a long delay between calls, and to be age-dependent.

Methods

Animals and housing conditions

Twelve captive adult male red-capped mangabeys (*Cercocebus torquatus*) were tested during this experiment. These individuals were all born in captivity and housed at the Station Biologique de Paimpont. The animals were distributed in several enclosures during the experiments (Table 1): two individuals were alone (II, VI), one was living with two individuals from another species (*Cercopithecus campbelli*) (V), and the others were living with conspecifics in small groups (I, III, IV). Along the individuals' life, group compositions changed several times. However, all individuals experienced living within social groups composed of both males and females for several years. The enclosures for each group included an indoor part (from 8 to 27 m²) and an outdoor part (from 15 to 37m²), with a height ranging from 2.5 to 4m. Each indoor part was composed of three large adjacent cages. All subjects were in auditory and visual contact with one another outdoor as well as with females in nearby enclosures. Individuals were free to come and go in/out via tunnels, except during our experiment where the subject was isolated in the indoor part while all the other individuals were kept outside. All selected subjects were habituated to be temporarily isolated for cognitive experiments (Maille et al., 2013; Rivière et al., 2017; Aychet et al., 2020). The indoor enclosure temperature was maintained at 22°C, and the floor was covered with straw and wood shavings,

while the outdoor floor was made of cement or covered with bark. All spaces were enriched with wood and metal perches. The mangabeys were fed twice a day, with fresh fruits and vegetables in the morning and monkey chows in the afternoon. Water was available ad libitum.

Table 1: Characteristics of the adult males studied

Enclosure	Subject	Age (years)
I.	Pirate	30
	George	16
	Carillon	15
	Elky	13
II.	Kamel	12
III.	Isba	18
	Roby	12
IV.	Tips	11
	Pouët	7
	Litchi	7
V.	Lupa	16
VI.	Coët	11

Stimuli preparation

We created three different vocal exchange stimuli:

- 1st condition “standard response time” (later named STANDARD): a likely vocal exchange between 2 individuals that respected the 500ms average inter-call duration of the species
- 2nd condition “long response time” (later named LONG): a rarer vocal exchange between 2 individuals that respected a long 1500ms inter-call duration
- 3rd condition “call overlap” (later named OVERLAP): an unlikely vocal exchange that violated the turn-taking rule with a 50% overlap between the 2 calls (as in Pougnault et al., 2020b).

Each stimulus was composed of two contact calls (named "Ro" in that species; Bouchet et al., 2010) from two different female callers. We did not use calls from mangabeys present in any of the enclosures to avoid other factors that could influence response time perception (i.e. dominance or affinity). We therefore selected calls from previous recordings of three individuals (A, B, C) no longer present on the site. We could thus build six different exchange dyads (AB, BA, BC, CB, AC, CA) and assigned them randomly to subjects. Each dyad was used to test two individuals. Calls were initially recorded with a Sony ECM-672 directional

microphone connected to a Marantz PMD660 (44.1 kHz sampling rate - 16 bits resolution). Only calls with good sound quality were selected (i.e. no echo, no background noise). A given call exemplar was never used twice for a given subject and when re-used for another subject, it was never placed in the same position (first or second) in the exchange. Vocal exchanges were created artificially stringing calls (with a silent gap between calls for STANDARD and LONG stimuli) using Audacity software. For each call, linear fade-in and fade-out effects were applied (i.e. 100ms pre and 100ms post). Stimuli were homogenized in intensity so that they reach 60 db at one meter from the loudspeaker (as in Pougnault et al., 2020b).

Playback protocol

Each type of vocal exchanges was played back once to each mangabey subject using a Bluetooth loudspeaker JBL GO connected to a DELL Latitude 5400-NJVXR laptop. For each subject, we randomized the order of presentation of the three types of stimuli and we waited at least 5 days between two consecutive tests playback for a given individual to limit habituation effects. In a given half day, only one test was done per pair of adjacent enclosures. Prior to a playback, all non-tested individuals were moved outside by the caretaker who then hid the connected loudspeaker on the ground under the straw, next to a closed entrance trapdoor of the cage adjacent to the cage where the tested individual was. During the speaker positioning, the experimenter diverted the attention of the subject who remained inside by feeding him with cashew nuts, a very appetizing food for them. Then, the caretaker left and the subject stayed with the experimenter. The experimenter was sitting and gave irregularly cashew nuts to keep the tested mangabey in front of him with the speaker in the subject's back so that attentive responses to playbacks consisted in similar head orientation efforts for all subjects. When the subject was calm, not eating, with its face towards the experimenter and the surrounding environment was quiet, the playback was launched by the experimenter. During the whole playback session, the experimenter did not look at the mangabey so as not to influence its behavior. The experimenter looked at the screen of his camera JVC GZ-RX615BE to record the mangabey reaction for 60 seconds after the end of the second calls. At the end of these 60 seconds, the subject was rewarded independently of his reaction with cashew nuts. For each individual, we did a mock session with no sound diffusion (between the 1st and the 2nd playback, or between the 2nd and the 3rd playback) yet again to limit habituation.

Data processing and analyses

Video recordings were analyzed using BORIS software. We first extracted the 60 seconds to be analyzed (just after the second call of the stimulus) from each video so that, in a second step,

the experimenter could code the samples blind to the condition being tested. We coded the head orientation towards the loudspeaker, defined as when the loudspeaker was in the visual field of the mangabey. We measured the number of occurrences, total duration, mean duration, as well as latency and duration of the first head orientation (as in Pougault et al., 2020b).

Statistics analyses were conducted using R software version 4.1.2 (R Core Team, 2021) with a significance threshold set at 0.05. We used non-parametric tests due to the limited number of individuals. We used Friedman test to assessed differences in the subject's reactions to the stimuli condition and Spearman correlations test to investigate link between mangabeys reactions to stimuli and age.

Results

The stimuli were generally of high interest to the subjects despite a great variability (total duration of head orientation mean=15.3 s, se=10.3 s, max=42.1 s, min=2.2 s). Analyses showed no differences in attention level toward the stimuli according to the condition tested. Individuals did not show differences of number of occurrences (Friedman test: $\chi^2(2) = 1.30$, $p=0.52$), total duration (Friedman test: $\chi^2(2) = 0.5$, $p= 0.78$), and mean duration of head orientation toward the loudspeakers (Friedman test: $\chi^2(2) = 2.17$, $p=0.34$). They also did not show differences of latency and duration of the first head orientation (Friedman test, latency: $\chi^2(2) = 6$, $p= 0.05$; duration: $\chi^2(2) = 0.17$, $p= 0.92$).

However, we found a correlation between age and total duration of head orientation during the vocal exchange with call overlap (Spearman test - OVERALP: $r_s = -0.63$, $p=0.03$), but we did not find any correlation in the two other conditions (Spearman test, STANDARD: $r_s = 0.13$, $p=0.69$; LONG: $r_s = -0.23$, $p=0.48$). Indeed, during the overlapping condition, the older an individual, the less interest paid to the loudspeaker area.

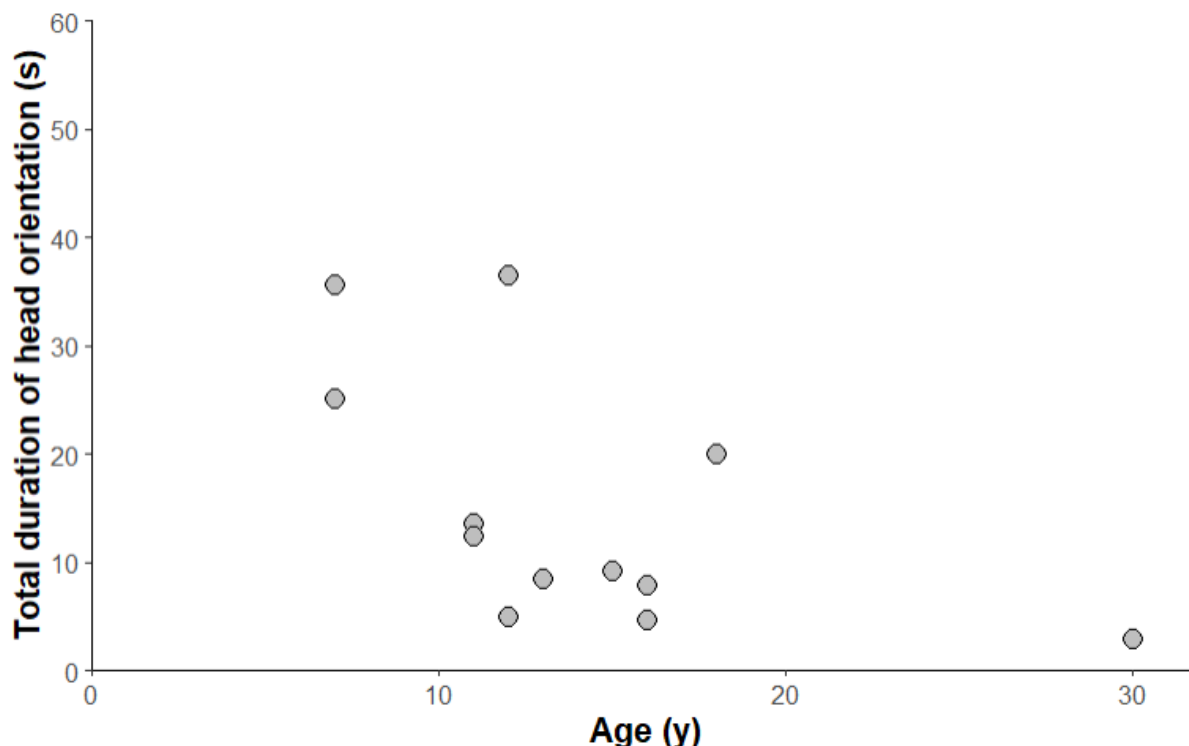


Fig. 1: Relationship between the total duration of head orientation towards the loudspeaker for the 60 seconds following the playback with overlap and the age of individuals.

Discussion

Contrary to our predictions, male mangabeys did not pay a differential attention to a standard contact call exchanges compared to exchanges where females were violating the turn-taking pattern, or using a long response time. Indeed, our subjects appeared to be systematically highly interested by the simulated female interactions with long-lasting head orientations. However, our prediction concerning age was validated, as attention was age-dependent in a context-specific way. While age did not influence males' responses in the two situations not violating the turn-taking expectation, we found that older individuals lost the most interest in the abnormal situation (i.e. the one breaking the turn-taking rule with overlapping calls).

We were thus unable to clearly demonstrate the relevance for mangabeys of the turn-taking conversational rule, since they did not show different reactions to the different interactional patterns presented. This is surprising because it goes against the results of comparable studies in several primate species and for several supposed conversational rules (caller alternation - Lemasson et al., 2011; overlap avoidance - Pournault et al., 2020b; acoustic matching - Bouchet et al., 2017; calling order - Cheney et al., 1995). It is therefore unlikely that mangabeys differ

so much from others, although we cannot exclude this interpretation at this stage. Another plausible interpretation would be the unsuitability of the experimental paradigm for our question. The violation-of-expectations paradigm is a well-known and widely used procedure for testing social cognition of non-speaking individuals, whether human infants or non-human primates (Aslin, 2007; Zuberbühler, 2014). While it is difficult to immediately compare our results with those of the other aforementioned contact call playback experiments, because different gaze measures are used, it clearly appears that our subjects showed a remarkably high overall interest in the played-back stimuli, with an average looking time of 15.3s, representing 26% of the analyzed time window. It is therefore likely that the stimuli we have chosen are not sufficiently differentiating and generally all interesting. A recent playback experiment on another topic showed a similar bias of attention (Treschnak et al., 2023). Indeed, authors also played-back female contact calls to male Guinea baboons (a species phylogenetically close to mangabeys) both consistently (the loudspeaker was spatially close to the played-back female) and inconsistently (the loudspeaker was in the opposite direction of the played-back female). The male baboons tested looked strongly at the females or loudspeakers in both conditions equally, in similar proportions to ours (median of 7.8s in 30s observation time; Treschnak et al., 2023). The fact that the callers were females may have been more important than their location (in baboons) or the temporal structure of the interaction (here in mangabeys). This may even have been amplified by the fact that, at the time of testing, our males had no access to females. This is consistent with several studies that showed the sensitivity of the gaze duration metric not only to the acoustic stimulus played back, but also to the identities of the simulated callers and to the broadcast context (Townsend et al., 2012; Wittig et al., 2007). Here, although the female calls were from old recordings of dead or departed individuals, most males had known them anyway and it was shown in previous studies that nonhuman mammals have a memory of past voices and that this typically triggers a strong interest (Bruck, 2013; Keenan et al., 2016; Lemasson et al., 2005). This is all the more plausible since contact calls are known to encode identity (Kondo and Watanabe, 2009), which has been demonstrated with the contact calls of mangabeys even allowing individuals to recognize their neighbors (non-group members) in captivity by ear (Bouchet et al., 2013; Candiotti et al., 2013).

Although our subjects appear to have been more interested in the caller's identity than in the temporal structure of the interaction, this does not mean that the latter is totally irrelevant for them. Indeed, another important result of this study is that the response of males depends on their age and that this is done in a context-dependent manner. Thus, our three situations are not totally equivalent in terms of influence on male interest. Firstly, only the situation that

completely violates the expectations of turn-taking makes males of different ages react differently. This confirms our hypothesis that the turn-taking rule could make sense for them. Here, older males are more disinterested in the incongruent situation than younger males. The fact that the incongruent situation is less interesting than the congruent situation is frequent in these experiments of vocal exchanges of contact calls (Lemasson et al., 2011; Pournault et al., 2020b). Secondly, the fact that an older age and thus experience condition this disinterest is consistent with the idea of an acquisition by social learning of the rule. Several conversational rules are suspected to be socially learned, such as the “acoustic matching” rule (i.e. the responder adjusts the acoustic structure of his call to copy the initiator of the exchange) in spider monkeys and Japanese macaques (Bouchet et al., 2017; Briseno-Jaramillo et al., 2018), and the “turn-taking” rule in Campbell’s monkeys and gorillas (Lemasson et al., 2011; Pournault et al., 2020b). These evidences are far from trivial due to the historical debate on vocal social learning in nonhuman primates, but also because social learning is an inherent feature of what we call a “rule”. Here, our results showed differences between young adults and older ones (as in gorillas; Pournault et al., 2020b), suggesting that social experience continues to play a role beyond youth throughout an individual’s life.

To summarize, the tested individuals do not seem to be aware of the turn-taking pattern underlying their vocal exchanges, but our results suggest that experience shape attention towards more socially relevant situations.

References

- Arlet, M., Jubin, R., Masataka, N., Lemasson, A., 2015. Grooming-at-a-distance by exchanging calls in non-human primates. *Biology Letters* 11, 20150711–20150714. <https://doi.org/10.1098/rsbl.2015.0711>
- Aslin, R.N., 2007. What’s in a look? *Developmental Science* 10, 48–53. <https://doi.org/10.1111/j.1467-7687.2007.00563.x>
- Aychet, J., Pezzino, P., Rossard, A., Bec, P., Blois-Heulin, C., Lemasson, A., 2020. Red-capped mangabeys (*Cercocebus torquatus*) adapt their interspecific gestural communication to the recipient’s behaviour. *Scientific Reports* 10, 12843. <https://doi.org/10.1038/s41598-020-69847-6>
- Bateson, M.C., 1975. Mother-infant exchanges: the epigenesis of conversational interaction. *Annals of the New York Academy of Sciences* 263, 101–113. <https://doi.org/10.1111/j.1749-6632.1975.tb41575.x>
- Biben, M., Symmes, D., Masataka, N., 1986. Temporal and Structural Analysis of Affiliative Vocal Exchanges in Squirrel Monkeys (*Saimiri sciureus*). *Behaviour* 98, 259–273.
- Bouchet, H., Blois-Heulin, C., Lemasson, A., 2013. Social complexity parallels vocal complexity: A comparison of three non-human primate species. *Frontiers in Psychology* 4, 1–15. <https://doi.org/10.3389/fpsyg.2013.00390>

- Bouchet, H., Koda, H., Lemasson, A., 2017. Age-dependent change in attention paid to vocal exchange rules in Japanese macaques. *Animal Behaviour* 129, 81–92. <https://doi.org/10.1016/j.anbehav.2017.05.012>
- Bouchet, H., Pellier, A.-S., Blois-Heulin, C., Lemasson, A., 2010. Sex differences in the vocal repertoire of adult red-capped mangabeys (*Cercocebus torquatus*): a multi-level acoustic analysis. *Am. J. Primatol.* 72, 360–375. <https://doi.org/10.1002/ajp.20791>
- Briseno-Jaramillo, M., Ramos-Fernández, G., Palacios-Romo, T.M., Sosa-López, & J.R., Lemasson, & A., Sosa-López, J.R., Lemasson, A., 2018. Age and social affinity effects on contact call interactions in free-ranging spider monkeys. *Behavioral Ecology and Sociobiology* 72, 192. <https://doi.org/10.1007/s00265-018-2615-2>
- Bruck, J.N., 2013. Decades-long social memory in bottlenose dolphins. *Proceedings of the Royal Society B: Biological Sciences* 280, 20131726. <https://doi.org/10.1098/rspb.2013.1726>
- Candiotti, A., Zuberbühler, K., Lemasson, A., 2013. Voice discrimination in four primates. *Behavioural Processes* 99, 67–72. <https://doi.org/10.1016/j.beproc.2013.06.010>
- Casillas, M., Bobb, S.C., Clark, E.V., 2016. Turn-taking, timing, and planning in early language acquisition*. *Journal of Child Language* 43, 1310–1337. <https://doi.org/10.1017/S0305000915000689>
- Cheney, D.L., Seyfarth, R.M., Silk, J.B., 1995. The Responses of Female Baboons (*Papio cynocephalus ursinus*) to Anomalous Social Interactions: Evidence for Causal Reasoning? *Journal of Comparative Psychology* 109, 134–141. <https://doi.org/10.1037/0735-7036.109.2.134>
- Chow, C.P., Mitchell, J.F., Miller, C.T., 2015. Vocal turn-taking in a non-human primate is learned during ontogeny. *Proceedings of the Royal Society B: Biological Sciences* 282, 20150069. <https://doi.org/10.1098/rspb.2015.0069>
- Dunbar, R.I.M., 1998. The Social Brain Hypothesis. *Evolutionary Anthropology* 6, 178_190.
- Hazen, N.L., Black, B., 1989. Preschool Peer Communication Skills: The Role of Social Status and Interaction Context. *Child Development* 60, 867–876. <https://doi.org/10.2307/1131028>
- Hilbrink, E.E., Gattis, M., Levinson, S.C., 2015. Early developmental changes in the timing of turn-taking: a longitudinal study of mother–infant interaction. *Front. Psychol.* 6. <https://doi.org/10.3389/fpsyg.2015.01492>
- Keenan, S., Mathevon, N., Stevens, J.M., Guéry, J.P., Zuberbühler, K., Levréro, F., 2016. Enduring voice recognition in bonobos. *Scientific Reports* 6, 4–11. <https://doi.org/10.1038/srep22046>
- Kondo, N., Watanabe, S., 2009. Contact calls: Information and social function. *Japanese Psychological Research* 51, 197–208. <https://doi.org/10.1111/j.1468-5884.2009.00399.x>
- Kulahci, I.G., Rubenstein, D.I., Ghazanfar, A.A., 2015. Lemurs groom-at-a-distance through vocal networks. *Animal Behaviour* 110, 179–185. <https://doi.org/10.1016/j.anbehav.2015.09.016>
- Lemasson, A., Gandon, E., Hausberger, M., 2010. Attention to elders' voice in non-human primates. *Biology Letters* 6, 325–328. <https://doi.org/10.1098/rsbl.2009.0875>
- Lemasson, A., Glas, L., Barbu, S., Lacroix, A., Guilloux, M., Remeuf, K., Koda, H., 2011. Youngsters do not pay attention to conversational rules: Is this so for nonhuman primates? *Scientific Reports* 1, 12–15. <https://doi.org/10.1038/srep00022>
- Lemasson, A., Hausberger, M., Zuberbühler, K., 2005. Socially meaningful vocal plasticity in adult Campbell's monkeys (*Cercopithecus campbelli*). *Journal of Comparative Psychology* 119, 220–229. <https://doi.org/10.1037/0735-7036.119.2.220>

- Lemasson, A., Pereira, H., Levréro, F., 2018. Social Basis of Vocal Interactions in Western Lowland Gorillas (*Gorilla g. gorilla*). *Journal of Comparative Psychology* 132, 141–151. <https://doi.org/10.1037/com0000105>
- Levinson, S.C., 2016. Turn-taking in Human Communication – Origins and Implications for Language Processing. *Trends in Cognitive Sciences* 20, 6–14. <https://doi.org/10.1016/j.tics.2015.10.010>
- Levréro, F., Touitou, S., Frédet, J., Nairaud, B., Guéry, J.-P., Lemasson, A., 2019. Social bonding drives vocal exchanges in Bonobos. *Scientific Reports* 9. <https://doi.org/10.1038/s41598-018-36024-9>
- Maille, A., Belbeoc'h, C., Rossard, A., Bec, P., Blois-Heulin, C., 2013. Which are the features of the TUBE task that make it so efficient in detecting manual asymmetries? An investigation in two Cercopithecine species (*Cercopithecus neglectus* and *Cercocebus torquatus*). *Journal of Comparative Psychology* 127, 436–444. <https://doi.org/10.1037/a0032227>
- Masataka, N., Biben, M., 1987. Temporal Rules Regulating Affiliative Vocal Exchanges of Squirrel Monkeys. *Behav* 101, 311–319. <https://doi.org/10.1163/156853987X00035>
- Meunier, B., Durier, V., Giacalone, A., Coye, C., Lemasson, A., 2023. Social factors drive vocal exchanges in red-capped mangabeys (*Cercocebus torquatus*). *Sci Nat* 110, 2. <https://doi.org/10.1007/s00114-023-01830-3>
- Pougnault, L., Levréro, F., Lemasson, A., 2020a. Conversation Among Primate Species, in: Masataka, N. (Ed.), *The Origins of Language Revisited: Differentiation from Music and the Emergence of Neurodiversity and Autism*. Springer, Singapore, pp. 73–96. https://doi.org/10.1007/978-981-15-4250-3_4
- Pougnault, L., Levréro, F., Mulot, B., Lemasson, A., 2020b. Breaking conversational rules matters to captive gorillas: A playback experiment. *Scientific Reports* 10, 6947. <https://doi.org/10.1038/s41598-020-63923-7>
- R Core Team, 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
- Rivière, J., Stomp, M., Augustin, E., Lemasson, A., Blois-Heulin, C., 2017. Decision-making under risk of gain in young children and mangabey monkeys. *Developmental Psychobiology*. <https://doi.org/10.1002/dev.21592>
- Robinson, L.F., Reis, H.T., 1989. The effects of interruption, gender, and status on interpersonal perceptions. *Journal of Nonverbal Behavior* 13, 141–153. <https://doi.org/10.1007/BF00987046>
- Sacks, H., Schegloff, E.A., Jefferson, G., 1974. A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language* 50, 696–735. <https://doi.org/10.2307/412243>
- Snowdon, C.T., Cleveland, J., 1984. “Conversations” among pygmy marmosets. *American Journal of Primatology* 7, 15–20. <https://doi.org/10.1002/ajp.1350070104>
- Stivers, T., Enfield, N.J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., de Ruiter, J.P., Yoon, K.-E., Levinson, S.C., 2009. Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences* 106, 10587–10592. <https://doi.org/10.1073/pnas.0903616106>
- Sugiura, H., 2007. Adjustment of temporal call usage during vocal exchange coo calls in Japanese macaques. *Ethology* 113, 528–533. <https://doi.org/10.1111/j.1439-0310.2007.01349.x>
- Sugiura, H., Masataka, N., 1995. Temporal and Acoustic Flexibility in Vocal Exchanges of Coo Calls in Japanese Macaques (*Macaca fuscata*), in: Zimmermann, E., Newman, J.D., Jürgens, U. (Eds.), *Current Topics in Primate Vocal Communication*. Springer US, Boston, MA, pp. 121–140. https://doi.org/10.1007/978-1-4757-9930-9_6

- Symmes, D., Biben, M., 1988. Conversational Vocal Exchanges in Squirrel Monkeys, in: Todt, D., Goedecking, P., Symmes, D. (Eds.), *Primate Vocal Communication*. Springer, Berlin, Heidelberg, pp. 123–132. https://doi.org/10.1007/978-3-642-73769-5_8
- ter Maat, M., Truong, K.P., Heylen, D., 2010. How Turn-Taking Strategies Influence Users' Impressions of an Agent, in: Allbeck, J., Badler, N., Bickmore, T., Pelachaud, C., Safonova, A. (Eds.), *Intelligent Virtual Agents, Lecture Notes in Computer Science*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 441–453. https://doi.org/10.1007/978-3-642-15892-6_48
- Townsend, S.W., Allen, C., Manser, M.B., 2012. A simple test of vocal individual recognition in wild meerkats. *Biol. Lett.* 8, 179–182. <https://doi.org/10.1098/rsbl.2011.0844>
- Treschnak, D., Zinner, D., Fischer, J., 2023. Male Guinea baboons may be oblivious to associated females' whereabouts. *Animal Behaviour* 201, 53–62. <https://doi.org/10.1016/j.anbehav.2023.04.010>
- Wittig, R.M., Crockford, C., Seyfarth, R.M., Cheney, D.L., 2007. Vocal alliances in Chacma baboons (*Papio hamadryas ursinus*). *Behav Ecol Sociobiol* 61, 899–909. <https://doi.org/10.1007/s00265-006-0319-5>
- Yoshida, S., Okanoya, K., 2005. Evolution of turntaking: a bio-cognitive perspective. *Cognitive Studies* 12, 153–165.
- Zuberbühler, K., 2014. Experimental field studies with non-human primates. *Current Opinion in Neurobiology* 28, 150–156. <https://doi.org/10.1016/j.conb.2014.07.012>



Chapitre IV :

Discussion

Au cours de ce travail, nous avons exploré les compétences conversationnelles de modèles biologiques ne disposant pas du langage : les nourrissons humains et les mangabeys à collier. Nous avons étudié ces compétences de manière systématique en menant, pour chacun de ces deux modèles, une étude observationnelle et une étude expérimentale. Nous avons intégré à chaque fois l'étude de l'impact de facteurs sociaux pertinents pour le modèle. Bien que nous ayons fait en sorte que chaque étude ait un intérêt pour le champ disciplinaire dans lequel elle s'intègre, elles ont un intérêt supplémentaire dans leur approche comparative. Ainsi, dans cette discussion générale je me concentrerai sur la comparaison des résultats déjà analysés individuellement dans les parties afférentes. Tout d'abord, je rappellerai les principaux résultats obtenus (1). Puis, je comparerai les deux modèles biologiques du point de vue de la structure temporelle des échanges observés de manière spontanée (2). Ensuite, je questionnerai la notion de norme pour interpréter les résultats obtenus lors de nos expériences (3). Aussi, j'évoquerai des pistes évolutives susceptibles d'expliquer l'apparition des conversations (4). Enfin, je discuterai des perspectives intéressantes pouvant suivre ce travail (5).

1. Analyse des proto-conversations

1.1 Chez les nourrissons humains

A notre connaissance, nos deux expériences sont les premières à avoir étudié aussi tôt chez des nourrissons préverbaux la perception des conversations en lien avec des facteurs socio-démographiques. Nous avons ainsi pu mettre en évidence que certains nourrissons étaient sensibles aux différences de transitions de tour au sein d'une conversation (passage de parole, tour de parole ou coupure de parole), mais que cette sensibilité était déterminée par des facteurs socio-démographiques. Dans la première expérience portant sur des nourrissons de 6 mois uniquement, le statut socio-économique de la famille et la présence d'une fratrie étaient déterminants dans les différences de perception observées : les nourrissons des familles au statut socio-économique élevé et les nourrissons qui n'avaient ni frère ni sœur étaient ceux qui ont réagi différemment aux situations présentées. Dans la seconde expérience portant sur des nourrissons de 3 à 11 mois, c'est le mode de garde qui était déterminant (facteur qui n'a pas pu être pris en compte dans l'expérience précédente). En effet, seuls les nourrissons gardés par leurs parents ont réagi différemment aux situations, contrairement à ceux gardés par un assistant maternel ou en crèche. De plus, les nourrissons gardés en crèche ont plus regardé la situation de tour de parole que les nourrissons gardés par leurs parents. Ces variations de perception en

fonction de facteurs socio-démographiques suggèrent un rôle de l'apprentissage social précoce à travers des expositions différentes.

Nos observations d'interactions dyadiques parent-nourrisson nous ont permis d'étudier les différences entre les interactions avec les mères et celles avec les pères, car les proto-conversations ont toujours été étudiées dans le cadre d'interactions mère-nourrisson. Nous n'avons pas observé de différences entre le comportement vocal des mères et celui des pères. Les deux ont plus parlé durant les interactions avec des garçons que durant les interactions avec des filles. Surtout, nous avons observé des différences entre le comportement vocal des garçons et celui des filles en lien avec les parents, puisque les garçons ont plus chevauché la parole que les filles mais seulement avec leurs pères. Les délais de réponses des parents de maximum 2300ms étaient plus courts que les délais de réponses des nourrissons qui allaient jusqu'à 3300ms.

1.2 Chez les mangabeys à collier

En ce qui concerne les mangabeys à collier, nous avons mis en évidence pour la première fois chez cette espèce qu'ils effectuaient des échanges vocaux assimilables à des proto-conversations. En effet, ils ont produit abondamment de cris de contact à courte distance en se répondant mutuellement avec des latences courtes qui ne dépassaient pas 2 secondes. La plupart de ces échanges vocaux étaient sous une forme simple composée d'une vocalisation initiatrice produite par un individu A suivie d'une vocalisation de réponse produite par un individu B. Les individus observés ne participaient pas aléatoirement à ces échanges vocaux : les individus les plus dominants étaient plus souvent en position de répondre à la vocalisation d'un autre individu que l'inverse, et le nombre d'échanges vocaux entre deux individus étaient faiblement corrélé à leur affinité (définie par le temps passé à proximité l'un de l'autre). Surtout, le délai de réponse variait de différentes manières en fonction du statut de l'individu et en fonction du lien entre les individus. Ainsi, les individus les plus intégrés socialement (au centre des toilettages sociaux) avaient des délais de réponse plus longs et les délais de réponse entre deux individus éloignés dans la hiérarchie de dominance étaient plus longs.

Nous avons aussi fait une expérience de repasse audio pour savoir si les mangabeys percevaient le pattern de tour de rôle qu'ils pratiquaient. Les individus testés n'ont manifesté aucune différence d'intérêt lors de la repasse des trois stimuli (un échange vocal avec chevauchement, un échange vocal avec un délai normal de 0.5s, et un échange vocal avec un long délai de

réponse de 1.5s). Les individus testés ont globalement été très intéressés par toutes les repasses, ce qui pourrait expliquer que nous n'avons pas trouvé de différences de réaction en fonction des conditions. Toutefois, l'âge des individus était corrélé à leurs réactions durant les repasses d'échanges vocaux présentant un chevauchement. Plus un individu était âgé, moins il a tourné la tête en direction du haut-parleur et donc manifesté de l'intérêt pour cette repasse. Le fait que la réponse soit liée à l'âge durant cette situation uniquement n'est probablement pas dû au hasard, puisque c'est la situation la moins fréquente et donc la moins pertinente socialement.

2. Une structure temporelle similaire

Dans son hypothèse d'un moteur interactionnel sous-jacent au langage humain, Levinson (2022, 2006) accorde une grande importance au contrôle du timing, qui se traduit par des latences de réponse courtes. Ce sont ces latences de réponse courtes que nous avons mises en évidence pour chacun de nos modèles, en mesurant la fréquence des intervalles séparant deux vocalisations de deux individus différents. Les courbes de distribution de nos modèles présentent en effet une allure similaire, et nous les avons utilisés pour établir un seuil de temps à partir duquel une vocalisation n'était pas considérée comme répondant à la vocalisation précédente. Mais au-delà de ce seuil, l'allure des courbes de distribution nous donne d'autres informations à comparer entre nos différents modèles.

Chez les mangabeys à collier, ce pic se situe entre 250 et 500ms, ce qui est légèrement plus élevé que le pic de réponses lors des conversations humaines, qui se situe entre 0 et 200ms (Levinson and Torreira, 2015). Ces latences de réponse très courtes ont été expliquées par les auteurs comme résultant du fait que les adultes humains planifient leur production vocale durant le tour de leurs interlocuteurs pour pouvoir intervenir immédiatement après, parfois même en léger chevauchement. Chez les mangabeys à collier, les tours n'étant constitués que d'une vocalisation de 100ms (200ms quand cette vocalisation est doublée), ils n'ont pas le temps de planifier leur tour et doivent donc se reposer uniquement sur leur temps de réaction. C'est probablement pour la même raison qu'il n'y a que très peu de chevauchements, le temps de réaction des mangabeys étant trop long pour qu'une vocalisation produite en réponse à une autre soit en chevauchement. Chez les singes hurleurs, Briseño-Jaramillo et al. (2021) ont repéré des patterns de chevauchement pour plusieurs vocalisations. Pourtant, lorsqu'on regarde quels sont les cris concernés et leurs patterns d'émission, les chevauchements ne concernent pas les vocalisations affiliatives, mais d'autres vocalisations qui sont émises en longues suites de

plusieurs unités, permettant aux autres individus de chevaucher ces dernières aisément en produisant simultanément une même suite de cette vocalisation (Briseño-Jaramillo et al., 2021). Chez les mones de Campbell, les jeunes répondent plus de manière inappropriée que les adultes en répétant plusieurs fois à la suite des cris de contact qui sont censés être émis une seule fois, pour laisser les autres membres répondre. Dans ce cas-là, des chevauchements peuvent aussi être possibles si un adulte veut répondre à la première vocalisation d'un jeune, et se retrouve à vocaliser en même temps que la seconde du jeune. De plus, les cris de contacts chez les mones de Campbell semblent plus longs que ceux des mangabeys (ils dépassent les 200ms ; Lemasson and Hausberger, 2004) ce qui faciliterait les chevauchements chez cette espèce. D'autres éléments pourraient expliquer qu'on puisse retrouver des chevauchements de ces cris de contact comme par exemple si des comportements vocaux peuvent être utilisés comme indices d'une production vocale (ouverture de la bouche, inspiration/gonflement d'un sac phonatoire, posture spécifique). Ce n'est en revanche pas le cas chez les mangabeys à collier, où la production d'un cri de contact n'implique pas de tels comportements (par exemple, la bouche n'est pas ouverte lors de leurs productions, obs. pers.) et semble donc difficile à anticiper.

Dans notre étude portant sur les interactions parent-nourrisson, nous avons mesuré les délais de réponses des nourrissons, mais aussi ceux des parents. En effet, les délais de réponses au cours d'interactions avec un nourrisson préverbal peuvent différer de ceux produits lors d'interactions avec un autre adulte capable de parler. Nous n'observons pourtant pas de différences puisque les parents avaient un pic de fréquence entre 0 et 200ms (sans différence entre les pères et les mères), tout comme le pic de fréquence des nourrissons, ce qui peut paraître surprenant. C'est pourtant conforme aux résultats de Gratier et al. (2015) qui ont montré que dès 3 mois, les nourrissons produisaient une grande proportion de vocalisations immédiatement à l'issue du tour du parent, avec des latences de réponses inférieures à 50ms et sans chevauchement (*latched turns*). Ces vocalisations sans latences de réponse représentaient en effet 45% des vocalisations en tour de parole avec les mères, ce qui interroge sur la capacité des nourrissons à anticiper la fin des tours sans avoir recours à des indices lexico-syntaxiques puisqu'étant trop jeunes. Une piste pourrait être qu'ils perçoivent des variations acoustiques liées à la prosodie, mais Casillas et Frank (2013) ont montré chez des enfants de 2 ans que la prosodie seule ne permet pas d'anticiper les changements de tours, ils ont besoin de la combinaison d'indices prosodiques et lexico-syntaxiques. L'autre hypothèse évoquée par Gratier et al (2015) serait que la mère interromprait volontairement son tour en voyant son nourrisson se préparer à parler, mais cela ne semblait pas être le cas dans leur étude puisque les productions vocales des mères précédant

ces transitions ne sont pas plus courtes que celles qui précèdent des transitions avec pauses. Il reste donc des éléments à élucider dans la manière dont les nourrissons préverbaux anticipent la fin des tours, d'où l'intérêt d'étudier leur perception des transitions de tours.

3. De la normalité à la normativité : mettre en évidence des normes

3.1 Qu'est-ce qu'une norme ?

Nous avons vu en introduction que le tour de rôle, plus qu'un pattern, pouvait être considéré comme une règle lorsqu'il était perçu par les individus qui le pratiquent. Ce terme de règle que nous avons employé sans discuter, fait référence à la notion de norme qui comporte une ambivalence. Si une norme peut être simplement définie comme des règles socialement partagées de comportements appropriés, de comportements auxquels on s'attend, elle peut avoir soit un sens descriptif en désignant ce que les individus font, soit un sens prescriptif en désignant ce que les individus doivent faire, une injonction. Il est donc important de toujours préciser lorsqu'on parle de norme, s'il s'agit d'une simple régularité d'un comportement et donc d'une norme comme normalité, ou bien si en plus d'une régularité, il peut être attaché un jugement moral vis-à-vis des comportements qui dévient et donc s'il s'agit d'une norme comme normativité. Pour savoir si un comportement est normatif chez des adultes humains, il suffit de leur poser des questions sur une situation présentée expérimentalement. C'est ainsi que des études ont montré que le tour de parole était clairement normatif et pas simplement normal chez les adultes humains : les individus qui coupent souvent la parole sont explicitement jugés négativement (Robinson and Reis, 1989b; ter Maat et al., 2010). En revanche, mettre en évidence une telle norme chez des individus qui ne peuvent pas parler comme des nourrissons ou des primates non-humains est plus difficile.

3.2 Des normes conversationnelles chez les nourrissons humains ?

Mettre en évidence des normes chez des nourrissons préverbaux nécessite des indicateurs comportementaux. Depuis un article fondateur montrant une préférence visuelle pour la nouveauté chez les nourrissons de 6 mois (Fantz, 1964), les regards ont été largement utilisés dans les études de psychologie du développement pour étudier la perception des nourrissons. Pourtant, des interprétations contradictoires de ces regards rendent difficile toute conclusion sur les processus cognitifs sous-jacents au-delà de la simple détection ou discrimination (Aslin, 2007). Une solution pour pallier à ces difficultés est d'avoir recours à des tests de choix

impliquant la préhension d'un jouet par exemple, mais ce genre d'alternatives ne peut se faire qu'avec des nourrissons assez mobiles et donc proches de la fin de leur première année.

Kinzler et al (2007) ont ainsi montré que des nourrissons de 5-6 mois exposés à des vidéos de locutrices parlant soit leur langue native soit une langue étrangère manifestaient une préférence visuelle pour les premières. Cette préférence a été confirmée chez des nourrissons de 10 mois qui choisissaient le jouet offert par la locutrice de leur langue native plutôt que celui offert par la locutrice d'une langue étrangère. Hamlin et al. (2007) ont montré que des nourrissons de 6 et 10 mois étaient capables d'évaluation sociale, c'est-à-dire de manifester une préférence pour un individu sur la base de ses comportements. Les chercheurs ont ainsi présenté aux nourrissons plusieurs saynètes présentant plusieurs personnages avec des attitudes différentes vis-à-vis d'un autre qui cherchait à accomplir une action : un qui l'aidait, un qui le gênait, et un qui restait neutre. Après cette exposition, lorsqu'ils devaient choisir un personnage (en l'attrapant), ils choisissaient le personnage aidant par rapport au gênant ou au neutre, et le neutre par rapport au gênant. Dans une étude postérieure, Hamlin et al. (2010) ont trouvé des résultats similaires chez des nourrissons encore plus jeunes (3 mois), mais ces derniers étant trop jeunes pour pouvoir saisir les personnages dans le test de choix, ils ont utilisé des temps de regards pour évaluer la préférences des nourrissons.

Ces études suggèrent que des normes avec attribution de valeurs (normatives) pouvaient être présentes très précocement chez les nourrissons et mises en évidence par des indicateurs comportementaux. En ce qui concerne nos expériences, si nous avons observé des variations de temps de regards en fonction de situations conversationnelles, notre dispositif ne nous permet pas de déterminer la raison de ces variations. Par exemple, nous ne pouvons savoir si, lorsque les nourrissons regardent plus longtemps une locutrice c'est parce qu'elle a produit un comportement nouveau pour eux, familier, ou bien parce qu'une valeur (positive ou négative) lui est attribuée. En revanche, nos expériences montrent que certains nourrissons sont capables de distinguer plusieurs situations conversationnelles sur la base des transitions de tours, ce qui est une étape préliminaire *sine qua non* avant d'attribuer une valence positive ou négative à ces dernières.

3.3 Des normes conversationnelles chez les primates non-humains ?

Le terme de norme est rarement employé lorsque l'on parle d'animaux non-humains. Pourtant, de nombreuses espèces vivent dans des groupes sociaux, c'est-à-dire des groupes permanents

où les individus ont des interactions répétées dans le temps (Hinde, 1976). La vie sociale nécessite pour les individus de prédire les interactions futures et leurs issues, et doivent donc se baser sur des régularités. Les primates non-humains sont remarquables à ce titre, car presque toutes les espèces de ce taxon vivent de cette manière, avec des individus qui doivent gérer plusieurs relations sociales et ont recours abondamment à la communication vocale pour ce faire (Dunbar, 2003). Des règles de communication respectées par tous les individus sont donc nécessaires pour gérer ces relations sociales (Pougnault et al., 2020a).

Il y a donc un réel intérêt de mettre en évidence les normes régissant la communication vocale des primates non-humains. Cela peut se faire en observant les régularités et la réaction des individus lorsque d'autres présentent des comportements qui dévient de cette régularité, une approche similaire de celle appliquée chez les humains dans l'analyse des conversations (Logue and Stivers, 2012). Chow et al. (2015) ont par exemple montré que les adultes marmosets réagissaient différemment lorsque les petits répondaient vocalement de manière inappropriée. C'est le même principe qui conduit les chercheurs à repasser des stimuli vocaux manipulés pour violer les attentes des individus sauvages ou captifs testés et ainsi étudier leurs réactions (Zuberbühler, 2014). De même que pour les paradigmes expérimentaux utilisés avec les nourrissons humains, le temps de regards est une variable très utilisée pour mesurer la réaction des primates non-humains. Comme chez les nourrissons humains, il est parfois difficile de savoir si un temps de regard est causé par une réaction à la nouveauté, à la familiarité, ou à l'attribution d'une valeur positive ou négative au stimulus repassé. Le pattern des temps de regards au cours de ces expériences varie grandement en fonction de l'espèce. Ainsi, les expériences de repasses portant sur les mones de Campbell ou les gorilles montrent que le stimulus le plus approprié suscite plus de regards (Lemasson et al., 2011; Pougnault et al., 2020c), tandis que chez les macaques ou les marmosets ce sont les stimuli qui violent les attentes (Wilson et al., 2013). De nouveaux outils comme la thermographie sont prometteurs dans le cadre d'expériences de repasses, car ils donnent des informations sur les changements d'états physiologiques qui ne se traduisent pas par des regards. Ainsi, une étude portant sur l'évaluation sociale d'échanges vocaux chez les marmosets a par exemple mis en évidence que ces derniers percevaient ces échanges de manière « holistique », comme des conversations plutôt que comme la somme de cris individuels (Brügger et al., 2021).

Dans notre expérience de repasses chez les mangabeys à collier, nous n'avons pas pu mettre en évidence de différences de temps de regards en fonction des conditions testées. Les individus ont en revanche manifesté un fort intérêt pour toutes les repasses, ce qui peut s'expliquer par

l'identité des émettrices des cris que nous avons utilisées pour constituer nos stimuli (il s'agissait de femelles inconnues des individus mâles testés). Si le tour de rôle chez les mangabeys est bel et bien une norme, elle serait en tout cas d'une importance relative chez cette espèce, notamment en comparaison de l'identité des émetteurs codée dans les cris de contacts.

4 Homologies et convergences, un puzzle évolutif à reconstruire

4.1 La pro-socialité qui motive notre moteur interactionnel

Depuis les écrits de Grice (1975, 1957), nous savons qu'un principe de coopération est au cœur de nos conversations, permettant le partage d'intentions de manière ostensive. Bien qu'une coopération puisse être le fruit d'intentions égoïstes pour atteindre un objectif, cette coopération est bien l'œuvre de motivations prosociales (Rossano, 2018), pro-socialité qui peut être définie comme la production spontanée de comportements altruistes. Certaines scènes de la vie quotidienne suffisent à nous faire nous rendre compte du haut degré de coopération dont nous sommes capables : nous pouvons aider un parfait inconnu à trouver son chemin dans la rue en lui donnant des informations vraies sans contrepartie, et cet inconnu les reçoit sans s'imaginer que nous puissions le tromper. Selon Wacewicz et Żywicznyński (2018), la communication humaine repose en fait sur une niche socio-écologique nommée « plateforme de confiance », qui permet une coopération à grande échelle en nous faisant considérer nos congénères comme des interlocuteurs fiables par défaut.

Les études de Hamlin et al. (2010, 2007) montrent en effet que les nourrissons manifestent dès les premiers mois une préférence pour les individus aux comportements prosociaux par rapport aux individus aux comportements anti-sociaux. La précocité de ces évaluations sociales de comportements suggère donc des bases instinctives aux jugements moraux (Hamlin, 2013). La pro-socialité serait ainsi une norme universelle ancrée biologiquement, par opposition aux normes arbitraires susceptibles de varier culturellement et nécessitant un apprentissage (Burkart et al., 2018). Cette pro-socialité ne semble pas présente chez des espèces pourtant proches : Krupenye et Hare (2018) ont transposé l'expérience de Hamlin et al. (2007) chez des bonobos (espèce réputée pour sa tolérance sociale), et ont montré que contrairement aux nourrissons humains, les individus testés ont manifesté une préférence pour les personnages au comportement anti-social. A l'inverse, des espèces plus éloignées comme les marmosets sont capables d'évaluer socialement des congénères sur la base d'interactions vocales, et manifestent une préférence pour les individus inconnus qui auraient eu des interactions vocales positives,

par rapport à des individus qui auraient eu des interactions vocales négatives (Brügger et al., 2021). Ceci suggère des mécanismes de convergence évolutive plus que d'homologies concernant l'apparition de la pro-socialité humaine.

Cette pro-socialité a probablement une influence sur l'apparition du moteur interactionnel humain. Ce dernier comporte plusieurs composantes, issues de voies évolutives différentes et dont on peut retrouver des traces de chacune d'elles chez diverses espèces de primates non-humains, mais dont l'articulation est spécifiquement humaine (Levinson, 2022, 2006). Plusieurs de ces composantes peuvent être liées à la pro-socialité, comme la réversibilité des rôles par exemple (Wacewicz et al., 2017). La réversibilité des rôles du locuteur et de l'auditeur dans une conversation (Levinson, 2006; Sacks et al., 1974; Schegloff, 2000) peut être qualifiée d'égalitaire car chaque individu est par principe libre d'intervenir ou non, et le rôle n'est pas prédéterminé par la nature du lien entre les individus. Dans notre étude observationnelle sur les mangabeys, nous avons vu qu'ils ne respectaient que partiellement cet aspect, car les individus les plus dominants se retrouvaient plus en position de répondre à une vocalisation d'un autre individu que l'inverse. Il ne semble pas que cela soit un rôle absolument prédéterminé à l'avance comme ça peut être le cas dans certaines fonctions de la communication (alarme anti-prédateurs, vocalisations de cour) mais bien une propension liée à des aspects motivationnels, puisqu'aucun des individus étudiés n'a été observé à vocaliser uniquement en réponse à un autre individu, de même qu'aucun individu n'a été observé à ne jamais vocaliser en réponse à un autre individu. Une autre composante du moteur interactionnel humain qui pourrait être facilitée par la pro-socialité est la multimodalité. En effet, en même temps que nous parlons, nous coordonnons des indices de plusieurs modalités tels que des regards, des mimiques faciales ainsi que des gestes (Holler, 2022). Cette multimodalité est permise par le fait que nos conversations se font en interactions face-à-face à très courte distance (moins d'un mètre), ce qui permet aux interlocuteurs d'interpréter cet ensemble de signaux simultanément. Or, interagir en face-à-face à courte distance nécessite une certaine confiance entre les partenaires puisqu'ils sont alors vulnérables par leur proximité physique à des agressions. De plus, chez les espèces primates non-humains où la dominance a une part importante, les regards peuvent être des signes de menaces, et ainsi les différents individus d'un groupe cherchent à les éviter (Harrod et al., 2020). La dominance a en effet un rôle important dans l'organisation sociale des mangabeys, pouvant empêcher la proximité entre certains individus (Dolado and Beltran, 2011). Chez les mangabeys que nous avons observé par exemple, les échanges vocaux avaient lieu entre des individus parfois distants de quelques mètres qui ne se regardaient pas. Les

mangabeys sont pourtant capables de produire des gestes et expressions faciales de manière intentionnelle (Aychet et al., 2021; Schel et al., 2022), ce qui pourraient enrichir les échanges vocaux. Ils sont en outre capables de s'assurer de l'attention visuelle d'un expérimentateur pour obtenir une récompense en se déplaçant dans son champ de vision (Aychet et al., 2020). La raison pour laquelle ils n'ont pas recours à une telle flexibilité dans les échanges vocaux ne semble donc pas être un problème de capacité mais bien de motivation.

4.2 Les relations mère-nourrisson comme « îlots de confiance » temporaires

Un contexte particulier est susceptible de fournir une motivation à la pro-socialité : la relation mère-nourrisson. Cette relation est en effet caractérisée par des comportements de soins de la part de la mère qui doit porter attention à l'état physique et mental de son nourrisson, et aussi par les besoins que le nourrisson a et qu'il doit combler en partageant ses intentions. Levinson (2022) qualifie ces relations de niche pour une « théorie de l'esprit spécialisée », du fait que les mères doivent prendre en compte les états physiques et mentaux de leur nourrisson pour ajuster leurs comportements de soins. Waciewicz et al. (2017) qualifient quant à eux ces relations « (d)'îlots de confiance » en référence à notre « plateforme de confiance » humaine, et seraient donc une niche idéale pour des comportements coopératifs prosociaux à échelle réduite. Les études portant sur les interactions mère-nourrisson chez les primates non-humains, que nous n'avons pas pu étudier au cours de cette thèse, devraient donc être plus complexes que les interactions entre primates adultes du fait de ces aspects motivationnels.

En observant des dyades mère-nourrisson macaques rhésus pendant les 23 premiers jours, Ferrari et al. (2009a) ont rapporté des regards mutuels fréquents entre les mères et leurs nourrissons, ainsi que de nombreux comportements de claquements de lèvres de la part des mères envers leurs nourrissons. La fréquence de ces comportements (regards mutuels et claquements de lèvres) augmente au cours des premières semaines avant de diminuer radicalement après le premier mois, conjointement avec l'augmentation du nombre de séparations des enfants et de leurs mères. Surtout, les auteurs ont mis en évidence que les mères pouvaient chercher activement le regard de leur nourrisson en manipulant la tête de ces derniers vers leur face, et que lorsqu'elles étaient un peu éloignées (entre 20 et 40cm), elles orientaient leur face vers celle de leur nourrisson en agitant leur tête de haut en bas. Les nourrissons sont d'ailleurs actifs durant ces interactions, puisqu'ils sont capables de répondre à ces claquements de lèvres par des claquements similaires, démontrant des capacités d'imitations dont l'existence

a été confirmée expérimentalement dès la naissance (Ferrari et al., 2009b). Cela rappelle les proto-conversations parent-nourrisson humaines, car les interactions face-à-face sont cruciales pour le développement des comportements sociaux des macaques. Dettmer et al. (2016) ont en effet manipulé l'expérience sociale de plusieurs nourrissons macaques rhésus en les exposant plusieurs fois par jour à des soigneurs humains qui avaient différentes consignes : certains d'entre eux devaient tenir le nourrisson en cherchant activement un regard mutuel et effectuer des claquements de lèvres, d'autres devaient simplement le tenir dans les bras avec leur face recouverte par un masque, et enfin un troisième groupe servant de contrôle devait délivrer les soins standards. Après quatre semaines, seuls les nourrissons exposés à des regards mutuels et des claquements de lèvres de soigneurs humains ont passé plus de temps à interagir avec leurs pairs (d'autres nourrissons) (Dettmer et al., 2016). De plus, quand on leur présentait deux types de stimuli vidéos, ils ont été les seuls à manifester une préférence pour le stimulus social (présentant un toilettage social entre plusieurs macaques rhésus) par rapport au stimulus abstrait (des formes abstraites mouvantes) (Dettmer et al., 2016).

D'autres études portant sur les grands singes ont montré que les relations mère-nourrisson étaient propices à des interactions complexes. Fröhlich et al. (2016) ont mis en évidence l'existence de tour de rôles gestuels coopératifs. Les chercheurs ont en effet analysé les comportements qui initient les déplacements conjoints (le nourrisson monte sur le dos de sa mère pour quitter un endroit) dans deux communautés de bonobos et deux communautés de chimpanzés. Ils ont montré que les mères comme les nourrissons pouvaient intervertir les rôles et être alternativement celui qui initie la séquence pour partir ou celui qui doit répondre à la demande. Les mères et les enfants s'assurent d'avoir l'attention de l'autre lorsque les premiers gestes sont produits grâce à des regards ou l'orientation de leur corps. Surtout, une réponse est attendue à la suite de chaque geste et les gestes sont répétés ou d'autres sont produits jusqu'à ce que l'action conjointe soit réalisée (Fröhlich et al., 2016). Cette flexibilité se retrouve même chez les orangs-outans, dont le mode de vie solitaire semble avoir réduit la capacité à interagir en suivant un tour de rôle vocal (Pougnault et al., 2022). Pourtant, Knox et al. (2019) ont montré qu'au sein des dyades mère-enfant, les individus ajustaient la modalité des signaux en utilisant des gestes visuels quand le receveur montrait de l'attention visuelle et des gestes tactiles quand il n'en montrait pas. Bien que rares, des vocalisations pouvaient être combinées à l'utilisation de gestes au cours de ces interactions, montrant que plusieurs modalités pouvaient être impliquées simultanément. Enfin, les individus étudiés étaient en général très réactifs aux gestes

et répondaient dans la seconde à ces derniers, montrant par là qu'il s'agissait vraiment de tour de rôle.

Ces études semblent donc confirmer que les relations mère-nourrisson chez les primates non-humains peuvent être considérées comme des « îlots de confiance » à l'intérieur desquelles des partages d'intentions au travers d'informations multimodales sont possibles. Ces « îlots de confiance » sont toutefois temporaires puisque ce genre d'interactions disparaît avec la séparation croissante des jeunes vis-à-vis de leur mère.

4.3 L'hypothèse de l'élevage coopératif

Un trait en lien avec la relation mère-enfant distingue notre espèce des grands singes dont nous sommes le plus proche phylogénétiquement : c'est l'élevage coopératif. Des auteurs pensent que la pro-socialité généralisée de notre espèce résulte de ce phénomène (Burkart et al., 2009). Les mères humaines peuvent en effet compter sur de nombreux congénères (apparentés ou non) pour les assister dans les soins nécessaires aux jeunes. L'existence de l'élevage coopératif semble être lié à notre taux de reproduction particulièrement rapide (au regard des standards des grands singes) couplé à un temps de développement très long : les mères humaines ont ainsi un second enfant avant que le premier ne puisse se débrouiller seul (Burkart et al., 2009; Hrdy and Burkart, 2020). Ceci rend l'élevage coopératif absolument nécessaire, bien que nous ne sachions pas si c'est l'élevage coopératif qui est la cause de ces autres traits ou si c'est l'inverse.

Parmi les primates non-humains, c'est dans la famille des callitrichidés (qui regroupe les marmosets et les tamarins) que l'élevage coopératif est le plus développé. Les callitrichidés vivent dans des petits groupes multimâles/multifemelles (moins d'une dizaine d'individus), généralement composés d'un unique couple reproducteur et de sa progéniture adulte et/ou juvénile, avec parfois des individus non-apparentés (Garber et al., 2016). Chez les callitrichidés aussi, le taux de reproduction est rapide, si bien que la nouvelle progéniture naît avant que la précédente ne soit indépendante. Les petits ont donc besoin des soins des autres membres du groupe, et les mères tolèrent qu'ils soient transférés auprès de ces aidants. L'élevage coopératif des callitrichidés se distingue de l'aloparentalité qu'on retrouve chez d'autres espèces de primates dans le haut degré de tolérance avec lequel les mères laissent ces aidants s'occuper de son petit sans se rebeller. Tous les membres du groupe donnent ainsi régulièrement de la nourriture aux immatures sans contrepartie et sans qu'ils aient besoin de quémander (Brügger et al., 2018; Guerreiro Martins et al., 2019). Ces dons de nourriture ont même lieu entre adultes

chez les tamarins (Guerreiro Martins et al., 2019). Il y aurait donc une convergence évolutive entre les callitrichidés et les humains qui expliqueraient leur propension à la pro-socialité en tant que conséquences de l'élevage coopératif. C'est ce qu'une expérience standardisée pour évaluer un score de pro-socialité a pu confirmer. Menée sur 15 espèces de primates, les callitrichidés et l'humain ont obtenu des meilleurs scores que les autres espèces, montrant que l'élevage coopératif prédisait mieux le score de pro-socialité que d'autres hypothèses retenues dans la littérature comme l'auto-domestication, la force des liens, la coordination lors d'activités de fourrages ou encore la taille du cerveau (Burkart et al., 2014).

Cette pro-socialité s'exprime aussi à travers le comportement vocal. Les tamarins lions dorés produisent en effet une vocalisation spécifique pour attirer les juvéniles auxquels ils veulent donner de la nourriture déjà extraite, mais lorsque ces juvéniles sont assez vieux pour avoir déjà pu trouver de la nourriture en fourrageant par eux-mêmes, ils produisent cette vocalisation dans un contexte différent, en présence d'une source de nourriture qui n'est pas encore extraite (Rapaport, 2011). Ceci suggère un comportement complexe de transmission d'informations selon des motivations prosociales proches d'un comportement d'enseignement. Au-delà d'apprendre à trouver de la nourriture, les callitrichidés sont aussi capable de guider l'apprentissage des juvéniles à interagir vocalement : les adultes sont moins susceptibles de continuer un échange vocal avec les juvéniles s'ils ont répondu trop tôt après une vocalisation, ou avec la mauvaise vocalisation (Chow et al., 2015b). Le renforcement positif que procure une réponse vocale peut d'ailleurs être suscitée par les juvéniles eux-mêmes. En effet, les juvéniles produisent de longues suites de vocalisations différentes qui ressemblent aux vocalisations des adultes sans en être véritablement : elles ressemblent au babillage qu'on observe chez les nourrissons humains, d'autant qu'elles stimulent les réponses de la part des adultes (Elowson et al., 1998).

Ainsi, l'élevage coopératif est un catalyseur de comportements communicatifs en ce qu'il étend la relation de confiance classique entre une mère et son nourrisson à l'ensemble des membres du groupe. La survie des juvéniles reposant sur les comportements de soins des adultes, ils ont une motivation forte à attirer leur attention pour interagir avec eux, et ceux-là sont motivés en retour à subvenir à leurs besoins (Hrdy and Burkart, 2020). Ces aspects motivationnels qui semblent manquer à la plupart des primates non-humains sont présents chez les callitrichidés, qui auraient ainsi un moteur interactionnel convergent à celui de l'humain, en faisant un bon modèle d'étude du langage (Burkart et al., 2022).

4.4 Retour aux exigences d'une vie sociale complexe

Si l'humain partage ce facteur motivationnel qu'est la pro-socialité induite par l'élevage coopératif avec les callitrichidés, chez ce premier seulement, cette adaptation vient se combiner à un héritage cognitif spécialisé dans la reconnaissance d'intentions en partage avec les grands singes (Burkart et al., 2009; Hrdy and Burkart, 2020). Les humains ont en effet hérité d'une société avec une dynamique de fission-fusion assez éloigné des groupes familiaux (ou pseudo-familiaux) restreints des callitrichidés. Les groupes sociaux caractérisé par ce genre de dynamique, où des sous-groupes de tailles et de compositions variables se forment de manière temporaire, sont considérés comme plus complexes que les groupes qui n'ont pas cette dynamique (Aureli et al., 2008). La complexité sociale, définie par le fait que des individus interagissent fréquemment dans des contextes très différents avec de nombreux individus, suscite de la complexité vocale pour que les individus puisse gérer leurs relations sociales (Freeberg et al., 2012). Dans ces sociétés de fission-fusion où les relations sociales doivent être constamment négociés, les échanges vocaux ont une place primordiale, ce qui justifie l'étude d'espèce de primates non-humains avec des organisation sociales de ce type. Briseño-Jaramillo et al. (2022) ont en effet montré comment que les échanges vocaux jouaient un rôle dans la fusion de sous-groupes chez les singes-araignées. Il pourrait ainsi être intéressant d'étudier des mangabeys à collier sauvages, puisqu'ils vivent eux aussi dans des groupes avec des dynamiques de fission-fusion mais légèrement différentes car saisonnières, donc où les fissions et les fusions de sous-groupes se font beaucoup moins régulièrement. C'est en tout cas dans la comparaison d'espèces de primates non-humains avec des systèmes sociaux très différents que nous pourrons en savoir plus sur les pressions évolutives ayant façonné notre propre moteur interactionnel.

5 Perspectives

Afin de poursuivre et d'étayer ce travail préliminaire, de nombreuses études avec d'autres méthodes que celles que nous avons utilisées peuvent voir le jour.

En ce qui concerne les primates non-humains, le fait que nous n'avons pu mettre en évidence clairement que les mangabeys à collier percevaient la règle du tour de rôle grâce à des temps de regards peut nous inviter à reconsidérer cet indicateur comportemental. D'autres méthodes existent, dont une qui a pu faire ses preuves dans le domaine de la perception de repasses de type conversationnelles : la thermographie (Brügger et al., 2021). Cette étude a en effet pu

démontrer que l'écoute d'une repasse de conversations pouvaient créer un changement d'état physiologique associé à un état émotionnel. L'écoute d'une repasse avec un pattern temporel conversationnel pourrait générer un état de surprise perceptible grâce à la chaleur générée au niveau du visage des mangabeys. Au-delà d'utiliser un autre indicateur comportemental pour mesurer la réaction des mangabeys à une repasse, on pourrait repasser des cris de manière interactives, en fonction des cris émis par l'individu (King, 2015). Ainsi, on pourrait manipuler précisément selon quel pattern temporel une vocalisation repassée interagirait avec une vocalisation émise par le sujet : il réagirait peut-être différemment à des échanges vocaux dans lequel il est impliqué.

Toujours en ce qui concerne les primates non-humains, nous avons évoqué les interactions complexes qui pouvaient exister dans le cadre de ces îlots de confiance que sont les relations mère-nourrisson. Il pourrait être plus intéressant de comparer ce type de proto-conversations aux proto-conversations des nourrissons humains, du fait des mécanismes similaires d'attachements et de confiance impliqués. Surtout, une étude longitudinale pourrait permettre de voir si ces interactions permettent d'acquérir des compétences conversationnelles, en faisant le lien entre la propension d'un nourrisson à interagir selon un tour de rôle avec sa mère et plus tard, sa propension à le faire avec d'autres adultes en plus de sa perception de patterns conversationnelles via des repasses. Enfin, dans ce cadre des interactions mère-nourrisson, des études comparatives entre plusieurs espèces aux systèmes sociaux différents seraient très pertinents, pour voir s'il y a des différences dans la complexité de ces interactions et dans leur persistance à l'âge adulte en fonction de ces systèmes sociaux.

En ce qui concerne les nourrissons humains, nous voulions au départ faire le lien entre la performance des nourrissons lors d'interactions observées avec leurs parents (en situation standardisée ou au foyer), et leur perception aux différentes situations conversationnelles lors de l'expérimentation. Le recrutement étant plus difficile que prévu (notamment à cause du COVID-19), nous n'avons pas réussi à observer suffisamment de nourrissons en interactions avec leurs parents pour faire le lien avec la performance de ces mêmes nourrissons lors de l'expérimentation. De même, les effectifs étant trop faibles, nous n'avons pas pu lier la structure des interactions parent-nourrisson avec des facteurs socio-démographiques dont on a montré qu'ils pouvaient avoir une influence sur la perception.

Il pourrait être intéressant d'étudier les proto-conversations entre enfants en bas-âge. Nous ne savons en effet pas si la structure temporelle de ces dernières serait différente que celles observées dans les interactions parent-nourrisson. Surtout, relativement peu d'étude ont pu faire

le lien entre les conversations parent-enfant au foyer et les conversations que l'enfant va avoir avec ses pairs à la crèche ou à l'école. On peut citer l'étude de Black and Logan (1995), qui mériterait donc d'être approfondie pour voir sa reproductibilité et voir si un lien peut être fait au niveau des structures temporelles de ces conversations.

Enfin, il pourrait être intéressant d'étudier plus précisément que nous l'avons fait les interactions parent-nourrisson dans une approche multimodale. Les conversations humaines étant pleinement multimodales (Holler, 2022; Holler and Levinson, 2019), il serait intéressant de voir au cours d'études longitudinales comment l'articulation de plusieurs modalités se met en place au cours du développement. On pourrait avoir recours un répertoire comportemental plus complet et précis que ce qui est généralement fait dans les études classiques portant sur les proto-conversations en prenant exemple de la précision des répertoires comportementales des animaux non-humains. Cette meilleure précision pourrait permettre de mettre en évidence des variations plus fines qui pourrait être intéressantes.

Références

- Amphaeris, J., Blumstein, D.T., Shannon, G., Tenbrink, T., Kershenbaum, A., 2023. A multifaceted framework to establish the presence of meaning in non-human communication. *Biol. Rev.* <https://doi.org/10.1111/brv.12989>
- Arlet, M., Jubin, R., Masataka, N., Lemasson, A., 2015. Grooming-at-a-distance by exchanging calls in non-human primates. *Biol. Lett.* **11**, 20150711–20150714. <https://doi.org/10.1098/rsbl.2015.0711>
- Arnold, K., Zuberbühler, K., 2013. Female Putty-Nosed Monkeys Use Experimentally Altered Contextual Information to Disambiguate the Cause of Male Alarm Calls. *PLoS ONE* **8**, e65660. <https://doi.org/10.1371/journal.pone.0065660>
- Aslin, R.N., 2007. What's in a look? *Dev. Sci.* **10**, 48–53. <https://doi.org/10.1111/j.1467-7687.2007.00563.x>
- Augusti, E., Melinder, A., Gredebäck, G., 2010. Look Who's Talking: Pre-Verbal Infants' Perception of Face-to-Face and Back-to-Back Social Interactions. *Front Psychol.* **1**, 161. <https://doi.org/10.3389/fpsyg.2010.00161>
- Aureli, F., Schaffner, C.M., Boesch, C., Bearder, S.K., Call, J., Chapman, C.A., Connor, R., Fiore, A.D., Dunbar, R.I.M., Henzi, S.P., Holekamp, K., Korstjens, A.H., Layton, R., Lee, P., Lehmann, J., Manson, J.H., Ramos-Fernandez, G., Strier, K.B., Schaik, C.P. van, 2008. Fission-Fusion Dynamics: New Research Frameworks. *Curr. Anthropol.* **49**, 627–654. <https://doi.org/10.1086/586708>
- Aychet, J., Blois-Heulin, C., Palagi, E., Lemasson, A., 2021. Facial displays in red-capped mangabeys (*Cercocebus torquatus*): Repertoire, social context, and potential intentionality. *J. Comp. Psychol.* **135**, 98–113. <https://doi.org/10.1037/com0000252>
- Aychet, J., Pezzino, P., Rossard, A., Bec, P., Blois-Heulin, C., Lemasson, A., 2020. Red-capped mangabeys (*Cercocebus torquatus*) adapt their interspecific gestural communication to the recipient's behaviour. *Sci. Rep.* **10**, 1–12. <https://doi.org/10.1038/s41598-020-69847-6>
- Barbu, S., Nardy, A., Chevrot, J.-P., Guellaï, B., Glas, L., Juhel, J., Lemasson, A., 2015. Sex Differences in Language Across Early Childhood: Family Socioeconomic Status does not Impact Boys and Girls Equally. *Front. Psychol.* **6**. <https://doi.org/10.3389/fpsyg.2015.01874>
- Barthel, M., Meyer, A.S., Levinson, S.C., 2017. Next Speakers Plan Their Turn Early and Speak after Turn-Final “Go-Signals.” *Front. Psychol.* **8**. <https://doi.org/10.3389/fpsyg.2017.00393>
- Berglund, E., Eriksson, M., Westerlund, M., 2005. Communicative skills in relation to gender, birth order, childcare and socioeconomic status in 18-month-old children. *Scand. J. Psychol.* **46**, 485–491. <https://doi.org/10.1111/j.1467-9450.2005.00480.x>
- Biben, M., Symmes, D., Masataka, N., 1986. Temporal and Structural Analysis of Affiliative Vocal Exchanges in Squirrel Monkeys (*Saimiri sciureus*). *Behaviour* **98**, 259–273.

- Bigelow, A.E., 1998. Infants' sensitivity to familiar imperfect contingencies in social interaction. *Infant Behav. Dev.* 21, 149–162. [https://doi.org/10.1016/S0163-6383\(98\)90060-1](https://doi.org/10.1016/S0163-6383(98)90060-1)
- Black, B., Logan, A., 1995a. Links between Communication Patterns in IVlothei Child, Father-Child, and Child-Peer Interactions and Children's Social Status. *Child Dev.* 66, 255–274.
- Black, B., Logan, A., 1995b. Links between Communication Patterns in Mother-Child, Father-Child, and Child-Peer Interactions and Children's Social Status. *Child Dev.* 66, 255. <https://doi.org/10.2307/1131204>
- Bloom, K., 1975. Social elicitation of infant vocal behavior. *J. Exp. Child Psychol.* 20, 51–58. [https://doi.org/10.1016/0022-0965\(75\)90025-9](https://doi.org/10.1016/0022-0965(75)90025-9)
- Bloom, K., Russell, A., Wassenberg, K., 1987. Turn taking affects the quality of infant vocalizations. *J. Child Lang.* 14, 211. <https://doi.org/10.1017/S0305000900012897>
- Bonicco, C., 2007. Goffman et l'ordre de l'interaction : un exemple de sociologie compréhensive. *Philonsorbonne* 31–48. <https://doi.org/10.4000/philonsorbonne.102>
- Bornstein, M.H., Arterberry, M.E., 2022. *Infancy: The Basics*. Taylor & Francis.
- Bornstein, M.H., Haynes, O.M., Pascual, L., Painter, K.M., Galperín, C., 1999. Play in Two Societies: Pervasiveness of Process, Specificity of Structure. *Child Dev.* 70, 317–331. <https://doi.org/10.1111/1467-8624.00024>
- Bouchet, H., Blois-Heulin, C., Pellier, A.S., Zuberbühler, K., Lemasson, A., 2012. Acoustic variability and individual distinctiveness in the vocal repertoire of red-capped mangabeys (*Cercocebus torquatus*). *J. Comp. Psychol.* 126, 45–56. <https://doi.org/10.1037/a0025018>
- Bouchet, H., Koda, H., Lemasson, A., 2017. Age-dependent change in attention paid to vocal exchange rules in Japanese macaques. *Anim. Behav.* 129, 81–92. <https://doi.org/10.1016/j.anbehav.2017.05.012>
- Bouchet, H., Pellier, A.-S., Blois-Heulin, C., Lemasson, A., 2010. Sex differences in the vocal repertoire of adult red-capped mangabeys (*Cercocebus torquatus*): a multi-level acoustic analysis. *Am. J. Primatol.* 72, 360–375. <https://doi.org/10.1002/ajp.20791>
- Briseño-Jaramillo, M., Berthet, M., Estrada, A., Biquand, V., Lemasson, A., 2021. Socially mediated overlap in vocal interactions between free-ranging black howler monkeys. *Am. J. Primatol.* 83, e23297. <https://doi.org/10.1002/ajp.23297>
- Briseno-Jaramillo, M., Ramos-Fernández, G., Palacios-Romo, T.M., Sosa-López, & J.R., Lemasson, & A., Sosa-López, J.R., Lemasson, A., 2018. Age and social affinity effects on contact call interactions in free-ranging spider monkeys. *Behav. Ecol. Sociobiol.* 72, 192. <https://doi.org/10.1007/s00265-018-2615-2>
- Briseño-Jaramillo, M., Sosa-López José, R., Ramos-Fernández, G., Lemasson, A., 2022. Flexible use of contact calls in a species with high fission-fusion dynamics. <https://doi.org/10.1101/2022.05.25.492876>

- Brügger, R.K., Kappeler-Schmalzriedt, T., Burkart, J.M., 2018. Reverse audience effects on helping in cooperatively breeding marmoset monkeys. *Biol. Lett.* 14, 20180030. <https://doi.org/10.1098/rsbl.2018.0030>
- Brügger, R.K., Willems, E.P., Burkart, J.M., 2021. Do marmosets understand others' conversations? A thermography approach. *Sci. Adv.* 7, eabc8790. <https://doi.org/10.1126/sciadv.abc8790>
- Burkart, J.M., Adriaense, J.E.C., Brügger, R.K., Miss, F.M., Wierucka, K., Van Schaik, C.P., 2022. A convergent interaction engine: vocal communication among marmoset monkeys. *Philos. Trans. R. Soc. B Biol. Sci.* 377, 20210098. <https://doi.org/10.1098/rstb.2021.0098>
- Burkart, J.M., Allon, O., Amici, F., Fichtel, C., Finkenwirth, C., Heschl, A., Huber, J., Isler, K., Kosonen, Z.K., Martins, E., Meulman, E.J., Richiger, R., Rueth, K., Spillmann, B., Wiesendanger, S., van Schaik, C.P., 2014. The evolutionary origin of human hyper-cooperation. *Nat. Commun.* 5, 4747. <https://doi.org/10.1038/ncomms5747>
- Burkart, J.M., Brügger, R.K., van Schaik, C.P., 2018. Evolutionary Origins of Morality: Insights From Non-human Primates. *Front. Sociol.* 3.
- Burkart, J.M., Hrdy, S.B., Van Schaik, C.P., 2009. Cooperative breeding and human cognitive evolution. *Evol. Anthropol. Issues News Rev.* 18, 175–186. <https://doi.org/10.1002/evan.20222>
- Byrne, R.W., Cartmill, E.A., Genty, E., Graham, K.E., Hobaiter, C., Tanner, J.E., 2017. Great ape gestures: intentional communication with a rich set of innate signals. *Anim. Cogn.* 20, 755–769. <https://doi.org/10.1007/s10071-017-1127-1>
- Cäsar, C., Byrne, R.W., Hoppitt, W., Young, R.J., Zuberbühler, K., 2012. Evidence for semantic communication in titi monkey alarm calls. *Anim. Behav.* 84, 405–411. <https://doi.org/10.1016/j.anbehav.2012.05.010>
- Casillas, M., Frank, M.C., 2017. The development of children's ability to track and predict turn structure in conversation. *J. Mem. Lang.* 92, 234–253. <https://doi.org/10.1016/j.jml.2016.06.013>
- Casillas, M., Frank, M.C., 2013. The development of predictive processes in children's discourse understanding, in: Knauff, M., Pauen, M., Sebanz, N., Wachsmuth, I. (Eds.), *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*. Cognitive Society, Berlin, pp. 299–304.
- Caskey, M., Stephens, B., Tucker, R., Vohr, B., 2014. Adult Talk in the NICU With Preterm Infants and Developmental Outcomes. *PEDIATRICS* 133, e578–e584. <https://doi.org/10.1542/peds.2013-0104>
- Chambliss, C.A., Feeny, N., 1992. Effects of Sex of Subject, Sex of Interrupter, and Topic of Conversation on the Perceptions of Interruptions. *Percept. Mot. Skills* 75, 1235–1241. <https://doi.org/10.2466/pms.1992.75.3f.1235>
- Chow, C.P., Mitchell, J.F., Miller, C.T., 2015a. Vocal turn-taking in a non-human primate is learned during ontogeny. *Proc. R. Soc. B Biol. Sci.* 282, 20150069. <https://doi.org/10.1098/rspb.2015.0069>

- Chow, C.P., Mitchell, J.F., Miller, C.T., 2015b. Vocal turn-taking in a non-human primate is learned during ontogeny. *Proc. R. Soc. B Biol. Sci.* 282, 20150069. <https://doi.org/10.1098/rspb.2015.0069>
- Clearfield, M.W., Nelson, N.M., 2006. Sex Differences in Mothers' Speech and Play Behavior with 6-, 9-, and 14-Month-Old Infants. *Sex Roles* 54, 127–137. <https://doi.org/10.1007/s11199-005-8874-1>
- Cooke, C.A., 2012. The feeding, ranging and positional behaviors of *Cercocebus torquatus*, the red-capped mangabey, in Sette Cama Gabon: a phylogenetic perspective (Doctoral dissertation). Ohio State University, USA.
- Crockford, C., Wittig, R.M., Mundry, R., Zuberbühler, K., 2012. Wild chimpanzees inform ignorant group members of danger. *Curr. Biol.* 22, 142–146. <https://doi.org/10.1016/j.cub.2011.11.053>
- Dettmer, A.M., Kaburu, S.S.K., Simpson, E.A., Paukner, A., Sclafani, V., Byers, K.L., Murphy, A.M., Miller, M., Marquez, N., Miller, G.M., Suomi, S.J., Ferrari, P.F., 2016. Neonatal face-to-face interactions promote later social behaviour in infant rhesus monkeys. *Nat. Commun.* 7, 11940. <https://doi.org/10.1038/ncomms11940>
- Digweed, S.M., Fedigan, L.M., Rendall, D., 2007. Who cares who calls? Selective responses to the lost calls of socially dominant group members in the white-faced capuchin (*Cebus capucinus*). *Am. J. Primatol.* 69, 829–835. <https://doi.org/10.1002/ajp.20398>
- Dolado, R., Beltran, F.S., 2012. Emergent patterns of social organization in captive *Cercocebus torquatus*: Testing the GrooFiWorld agent-based model. *J. Biosci.* 37, 777–784. <https://doi.org/10.1007/s12038-012-9231-5>
- Dolado, R., Beltran, F.S., 2011. Dominance hierarchy and spatial distribution in captive red-capped mangabeys (*Cercocebus torquatus torquatus*): Testing Hemelrijk's agent-based model. *Interact. Stud.* 12, 461–473. <https://doi.org/10.1075/is.12.3.05dol>
- Dolado, R., Cooke, C., Beltran, F.S., 2016. How Many for Lunch Today? Seasonal Fission-Fusion Dynamics as a Feeding Strategy in Wild Red-Capped Mangabeys (*Cercocebus torquatus*). *Folia Primatol. Int. J. Primatol.* 87, 197–212. <https://doi.org/10.1159/000449220>
- Dominguez, S., Devouche, E., Apter, G., Gratier, M., 2016. The Roots of Turn-Taking in the Neonatal Period. *Infant Child Dev.* 25, 240–255. <https://doi.org/10.1002/icd.1976>
- Dunbar, R.I.M., 2003. The Social Brain: Mind, Language, and Society in Evolutionary Perspective. *Annu. Rev. Anthropol.* 32, 163–181. <https://doi.org/10.1146/annurev.anthro.32.061002.093158>
- Dunbar, R.I.M., 1995. Coevolution of neocortical size, group size and language in humans. *Behav. Brain Sci.* 16, 681–735. <https://doi.org/10.1017/S0140525X00039042>
- Dunbar, R.I.M., 1991. Functional significance of social grooming in primates. *Folia Primatol. (Basel)* 57, 121–131.
- Dunbar, R.I.M., Marriott, A., Duncan, N.D.C., 1997. Human conversational behavior. *Hum. Nat.* 8, 231–246. <https://doi.org/10.1007/bf02912493>

- Dunn, J., Shatz, M., 1989. Becoming a Conversationalist despite (Or Because of) Having an Older Sibling. *Child Dev.* 60, 399–410. <https://doi.org/10.2307/1130985>
- Elowson, A.M., Snowdon, C., Lazaro-Perea, C., 1998. Infant “Babbling” in a Nonhuman Primate: Complex Vocal Sequences with Repeated Call Types. *Behaviour* 135, 643–664. <https://doi.org/10.1163/156853998792897905>
- Eriksson, M., Marschik, P., Tulviste, T., Almgren, M., Pérez-Pereira, M., Wehberg, S., Umek, L., Gayraud, F., Kovačević, M., Gallego, C., 2012. Differences between girls and boys in emerging language skills: Evidence from 10 language communities. *Br. J. Dev. Psychol.* 30, 326–43. <https://doi.org/10.1111/j.2044-835X.2011.02042.x>
- Farley, S.D., 2008. Attaining Status at the Expense of Likeability: Pilfering Power Through Conversational Interruption. *J. Nonverbal Behav.* 32, 241–260. <https://doi.org/10.1007/s10919-008-0054-x>
- Fernald, A., Marchman, V.A., Weisleder, A., 2013. SES differences in language processing skill and vocabulary are evident at 18 months. *Dev. Sci.* 16, 234–248. <https://doi.org/10.1111/desc.12019>
- Ferrari, P.F., Paukner, A., Ionica, C., Suomi, S.J., 2009a. Reciprocal Face-to-Face Communication between Rhesus Macaque Mothers and Their Newborn Infants. *Curr. Biol.* 19, 1768–1772. <https://doi.org/10.1016/j.cub.2009.08.055>
- Ferrari, P.F., Paukner, A., Ruggiero, A., Darcey, L., Unbehagen, S., Suomi, S.J., 2009b. Interindividual Differences in Neonatal Imitation and the Development of Action Chains in Rhesus Macaques. *Child Dev.* 80, 1057–1068. <https://doi.org/10.1111/j.1467-8624.2009.01316.x>
- Fischer, J., Price, T., 2016. Meaning, intention, and inference in primate vocal communication. *Neurosci. Biobehav. Rev.* <https://doi.org/10.1016/j.neubiorev.2016.10.014>
- Freeberg, T.M., Dunbar, R.I.M., Ord, T.J., 2012. Social complexity as a proximate and ultimate factor in communicative complexity. *Philos. Trans. R. Soc. B Biol. Sci.* 367, 1785–1801. <https://doi.org/10.1098/rstb.2011.0213>
- Fröhlich, M., Kuchenbuch, P., Müller, G., Fruth, B., Furuichi, T., Wittig, R.M., Pika, S., 2016. Unpeeling the layers of language: Bonobos and chimpanzees engage in cooperative turn-taking sequences. *Sci. Rep.* 6. <https://doi.org/10.1038/srep25887>
- Galsworthy, M.J., Dionne, G., Dale, P.S., Plomin, R., 2000. Sex differences in early verbal and non-verbal cognitive development. *Dev. Sci.* 3, 206–215. <https://doi.org/10.1111/1467-7687.00114>
- Garber, P.A., Porter, L.M., Spross, J., Di Fiore, A., 2016. Tamarins: Insights into monogamous and non-monogamous single female social and breeding systems. *Am. J. Primatol.* 78, 298–314. <https://doi.org/10.1002/ajp.22370>
- Gardner, R.A., Gardner, B.T., 1969. Teaching Sign Language to a Chimpanzee. *Science* 165, 664–672.
- Gautier-Hion, A., Colyn, M., Gautier, J.-P., 1999. Histoire naturelle des primates d’Afrique centrale. ECOFAC, Gabon.

- Geissmann, T., Orgeldinger, M., 2000. The relationship between duet songs and pair bonds in siamangs, *Hylobates syndactylus*. *Anim. Behav.* 60, 805–809. <https://doi.org/10.1006/anbe.2000.1540>
- Ginsburg, G.P., Kilbourne, B.K., 1988. Emergence of vocal alternation in mother-infant interchanges. *J. Child Lang.* 15, 221–235. <https://doi.org/10.1017/S0305000900012344>
- Goldstein, M.H., King, A.P., West, M.J., 2003. Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proc. Natl. Acad. Sci.* 100, 8030–8035. <https://doi.org/10.1073/pnas.1332441100>
- Goodsitt, J.V., Morgan, J.L., Kuhl, P.K., 1993. Perceptual strategies in prelingual speech segmentation. *J. Child Lang.* 20, 229–252. <https://doi.org/10.1017/S0305000900008266>
- Goodwin, C., Heritage, J., 1990. *Conversation Analysis* 26.
- Gouzoules, S., Gouzoules, H., Marler, P., 1984. Rhesus Monkey (*Macaca mulata*) Screams: Representational Signalling in the Recruitment of Agonistic Aid. *Anim. Behav.* 32, 182–193.
- Gratier, M., Devouche, E., Guellai, B., Infanti, R., Yilmaz, E., Parlato, E., 2015. Early development of turn-taking in vocal interaction between mothers and infants. *Front. Psychol.* 6. <https://doi.org/10.3389/fpsyg.2015.01167>
- Grice, H.P., 1975. Logic and conversation, in: Cole, P., Morgan, J.L. (Eds.), *Syntax and Semantics 3: Speech Acts*. Academic Press, New York, pp. 41–58.
- Grice, H.P., 1957. Meaning. *Philos. Rev.* 66, 377–388. <https://doi.org/10.2307/2182440>
- Gros-Louis, J., West, M.J., Goldstein, M.H., King, A.P., 2006. Mothers provide differential feedback to infants' prelinguistic sounds. *Int. J. Behav. Dev.* 30, 509–516. <https://doi.org/10.1177/0165025406071914>
- Guerreiro Martins, E.M., Moura, A.C.D.A., Finkenwirth, C., Griesser, M., Burkart, J.M., 2019. Food sharing patterns in three species of callitrichid monkeys (*Callithrix jacchus*, *Leontopithecus chrysomelas*, *Saguinus midas*): Individual and species differences. *J. Comp. Psychol.* 133, 474–487. <https://doi.org/10.1037/com0000169>
- Hamlin, J.K., 2013. Moral Judgment and Action in Preverbal Infants and Toddlers: Evidence for an Innate Moral Core. *Curr. Dir. Psychol. Sci.* 22, 186–193. <https://doi.org/10.1177/0963721412470687>
- Hamlin, J.K., Wynn, K., Bloom, P., 2010. Three-month-olds show a negativity bias in their social evaluations. *Dev. Sci.* 13, 923–929. <https://doi.org/10.1111/j.1467-7687.2010.00951.x>
- Hamlin, J.K., Wynn, K., Bloom, P., 2007. Social evaluation by preverbal infants. *Nature* 450, 557–559. <https://doi.org/10.1038/nature06288>
- Hammerschmidt, K., Fischer, J., 2013. Constraints in Primate Vocal Production.

- Harrod, E.G., Coe, C.L., Niedenthal, P.M., 2020. Social Structure Predicts Eye Contact Tolerance in Nonhuman Primates: Evidence from a Crowd-Sourcing Approach. *Sci. Rep.* 10, 6971. <https://doi.org/10.1038/s41598-020-63884-x>
- Hart, B., Risley, T.R., 2003. The Early Catastrophe. The 30 Million Word Gap by Age 3. *Am. Educ.* 27, 4–9.
- Hauser, M.D., Chomsky, N., Fitch, W.T., 2002. The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science* 298, 1569–1579. <https://doi.org/10.1126/science.298.5598.1569>
- Hazen, N.L., Black, B., 1989. Preschool Peer Communication Skills: The Role of Social Status and Interaction Context. *Child Dev.* 60, 867–876. <https://doi.org/10.2307/1131028>
- Heldner, M., 2011. Detection thresholds for gaps, overlaps, and no-gap-no-overlaps. *J. Acoust. Soc. Am.* 130, 508–513. <https://doi.org/10.1121/1.3598457>
- Heldner, M., Edlund, J., 2010. Pauses, gaps and overlaps in conversations. *J. Phon.* 38, 555–568. <https://doi.org/10.1016/j.wocn.2010.08.002>
- Henry, L., Craig, A.J.F.K., Lemasson, A., Hausberger, M., 2015. Social coordination in animal vocal interactions. Is there any evidence of turn-taking? The starling as an animal model. *Front. Psychol.* 6, 1416. <https://doi.org/10.3389/fpsyg.2015.01924>
- Hewes, G.W., 1973. Primate communication and the gestural origin of language. *Curr. Anthropol.* 14, 65–84.
- Hilbrink, E.E., Gattis, M., Levinson, S.C., 2015. Early developmental changes in the timing of turn-taking: a longitudinal study of mother–infant interaction. *Front. Psychol.* 6. <https://doi.org/10.3389/fpsyg.2015.01492>
- Hinde, R.A., 1976. Interactions, Relationships and Social Structure. *Man* 11, 1–17. <https://doi.org/10.2307/2800384>
- Hoff, E., 2006. How social contexts support and shape language development☆. *Dev. Rev.* 26, 55–88. <https://doi.org/10.1016/j.dr.2005.11.002>
- Hoff, E., 2003. The Specificity of Environmental Influence: Socioeconomic Status Affects Early Vocabulary Development Via Maternal Speech. *Child Dev.* 74, 1368–1378. <https://doi.org/10.1111/1467-8624.00612>
- Hoff-Ginsberg, E., 1998. The relation of birth order and socioeconomic status to children’s language experience and language development. *Appl. Psycholinguist.* 19, 603–629. <https://doi.org/10.1017/S0142716400010389>
- Holler, J., 2022. Visual bodily signals as core devices for coordinating minds in interaction. *Philos. Trans. R. Soc. B Biol. Sci.* 377, 20210094. <https://doi.org/10.1098/rstb.2021.0094>
- Holler, J., Levinson, S.C., 2019. Multimodal Language Processing in Human Communication. *Trends Cogn. Sci.* 23, 639–652. <https://doi.org/10.1016/j.tics.2019.05.006>

- Hrdy, S.B., Burkart, J.M., 2020. The emergence of emotionally modern humans: implications for language and learning. *Philos. Trans. R. Soc. B Biol. Sci.* 375, 20190499. <https://doi.org/10.1098/rstb.2019.0499>
- Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., Hedges, L.V., 2010. Sources of Variability in Children's Language Growth. *Cognit. Psychol.* 61, 343–365. <https://doi.org/10.1016/j.cogpsych.2010.08.002>
- Indefrey, P., Levelt, W.J.M., 2004. The spatial and temporal signatures of word production components. *Cognition* 92, 101–144. <https://doi.org/10.1016/j.cognition.2002.06.001>
- Jones, C.P., Adamson, L.B., 1987. Language Use in Mother-Child and Mother-Child-Sibling Interactions 12.
- Katsu, N., Yamada, K., Nakamichi, M., 2016. Function of grunts, girneys and coo calls of Japanese macaques (*Macaca fuscata*) in relation to call usage, age and dominance relationships. *Behaviour* 153, 125–142. <https://doi.org/10.1163/1568539X-00003330>
- Katsu, N., Yamada, K., Okanoya, K., Nakamichi, M., 2018. Temporal adjustment of short calls according to a partner during vocal turn-taking in Japanese macaques. *Curr. Zool.* 65, 99–105. <https://doi.org/10.1093/cz/zoy077>
- King, S.L., 2015. You talkin' to me? Interactive playback is a powerful yet underused tool in animal communication research. *Biol. Lett.* 11, 20150403. <https://doi.org/10.1098/rsbl.2015.0403>
- Kinzler, K.D., Dupoux, E., Spelke, E.S., 2007. The native language of social cognition. *Proc. Natl. Acad. Sci.* 104, 12577–12580. <https://doi.org/10.1073/pnas.0705345104>
- Kirchhof, J., Hammerschmidt, K., 2006. Functionally Referential Alarm Calls in Tamarins (*Saguinus fuscicollis* and *Saguinus mystax*) – Evidence from Playback Experiments. *Ethology* 112, 346–354. <https://doi.org/10.1111/j.1439-0310.2006.01165.x>
- Knox, A., Markx, J., How, E., Azis, A., Hobaiter, C., van Veen, F.J.F., Morrogh-Bernard, H., 2019. Gesture Use in Communication between Mothers and Offspring in Wild Orang-Utans (*Pongo pygmaeus wurmbii*) from the Sabangau Peat-Swamp Forest, Borneo. *Int. J. Primatol.* 40, 393–416. <https://doi.org/10.1007/s10764-019-00095-w>
- Krupenye, C., Hare, B., 2018. Bonobos Prefer Individuals that Hinder Others over Those that Help. *Curr. Biol.* 28, 280-286.e5. <https://doi.org/10.1016/j.cub.2017.11.061>
- Kuhl, P.K., 2004. Early language acquisition: cracking the speech code. *Nat. Rev. Neurosci.* 5, 831–843. <https://doi.org/10.1038/nrn1533>
- Kuhl, P.K., Meltzoff, A.N., 1996. Infant vocalizations in response to speech: Vocal imitation and developmental change. *J. Acoust. Soc. Am.* 100, 2425–2438.
- Kuhl, P.K., Tsao, F.-M., Liu, H.-M., 2003. Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proc. Natl. Acad. Sci.* 100, 9096–9101. <https://doi.org/10.1073/pnas.1532872100>

- Kulahci, I.G., Rubenstein, D.I., Ghazanfar, A.A., 2015. Lemurs groom-at-a-distance through vocal networks. *Anim. Behav.* 110, 179–185. <https://doi.org/10.1016/j.anbehav.2015.09.016>
- LaFrance, M., 1992. Gender and Interruptions: Individual Infraction or Violation of the Social Order? *Psychol. Women Q.* 16, 497–512. <https://doi.org/10.1111/j.1471-6402.1992.tb00271.x>
- Landman, R., Sharma, J., Hyman, J.B., Fanucci-Kiss, A., Meisner, O., Parmar, S., Feng, G., Desimone, R., 2020. Close-range vocal interaction in the common marmoset (*Callithrix jacchus*). *PLOS ONE* 15, e0227392. <https://doi.org/10.1371/journal.pone.0227392>
- Leaper, C., Anderson, K.J., Sanders, P., 1998. Moderators of gender effects on parents' talk to their children: A meta-analysis. *Dev. Psychol.* 34, 3–27. <https://doi.org/10.1037/0012-1649.34.1.3>
- Lemasson, A., Blois-Heulin, C., Jubin, R., Hausberger, M., 2006. Female social relationships in a captive group of Campbell's monkeys (*Cercopithecus campbelli campbelli*). *Am. J. Primatol.* 68, 1161–1170. <https://doi.org/10.1002/ajp.20315>
- Lemasson, A., Gandon, E., Hausberger, M., 2010. Attention to elders' voice in non-human primates. *Biol. Lett.* 6, 325–328. <https://doi.org/10.1098/rsbl.2009.0875>
- Lemasson, A., Gautier, J.P., Hausberger, M., 2003. Vocal similarities and social bonds Campbell's monkey (*Cercopithecus campbelli*). *Comptes Rendus - Biol.* 326, 1185–1193. <https://doi.org/10.1016/j.crv.2003.10.005>
- Lemasson, A., Glas, L., Barbu, S., Lacroix, A., Guilloux, M., Remeuf, K., Koda, H., 2011. Youngsters do not pay attention to conversational rules: Is this so for nonhuman primates? *Sci. Rep.* 1, 12–15. <https://doi.org/10.1038/srep00022>
- Lemasson, A., Guilloux, M., Rizaldi, Barbu, S., Lacroix, A., Koda, H., 2013. Age- and sex-dependent contact call usage in Japanese macaques. *Primates* 54, 283–291. <https://doi.org/10.1007/s10329-013-0347-5>
- Lemasson, A., Hausberger, M., 2004. Patterns of Vocal Sharing and Social Dynamics in a Captive Group of Campbell's Monkeys (*Cercopithecus campbelli campbelli*). *J. Comp. Psychol.* 118, 347–359. <https://doi.org/10.1037/0735-7036.118.3.347>
- Lemasson, A., Pereira, H., Levréro, F., 2018. Social basis of vocal interactions in western lowland gorillas (*Gorilla g. gorilla*). *J. Comp. Psychol.* 132, 141–151. <https://doi.org/10.1037/com0000105>
- Levinson, S.C., 2022. The interaction engine: cuteness selection and the evolution of the interactional base for language. *Philos. Trans. R. Soc. B Biol. Sci.* 377, 20210108. <https://doi.org/10.1098/rstb.2021.0108>
- Levinson, S.C., 2019. Interactional Foundations of Language: The Interaction Engine Hypothesis, in: Hagoort, P. (Ed.), *Human Language*. The MIT Press, pp. 189–200. <https://doi.org/10.7551/mitpress/10841.003.0018>
- Levinson, S.C., 2016. Turn-taking in Human Communication – Origins and Implications for Language Processing. *Trends Cogn. Sci.* 20, 6–14. <https://doi.org/10.1016/j.tics.2015.10.010>

- Levinson, S.C., 2006. On the Human “Interaction Engine,” in: Enfield, N.J., Levinson, S.C. (Eds.), *Roots of Human Sociality*. Routledge, pp. 39–69. <https://doi.org/10.4324/9781003135517-3>
- Levinson, S.C., Holler, J., 2014. The origin of human multi-modal communication. *Philos. Trans. R. Soc. B Biol. Sci.* 369. <https://doi.org/10.1098/rstb.2013.0302>
- Levinson, S.C., Torreira, F., 2015. Timing in turn-taking and its implications for processing models of language. *Front. Psychol.* 7, 731. <https://doi.org/10.3389/978-2-88919-825-2>
- Levréro, F., Touitou, S., Frédet, J., Nairaud, B., Guéry, J.-P., Lemasson, A., 2019. Social bonding drives vocal exchanges in Bonobos. *Sci. Rep.* 9. <https://doi.org/10.1038/s41598-018-36024-9>
- Logue, D.M., Stivers, T., 2012. Squawk in interaction: A primer of conversation analysis for students of animal communication. *Behaviour* 149, 1283–1298. <https://doi.org/10.1163/1568539X-00003031>
- Lovas, G.S., 2011. Gender and patterns of language development in mother-toddler and father-toddler dyads. *First Lang.* 31, 83–108. <https://doi.org/10.1177/0142723709359241>
- Macedonia, J.M., Evans, C.S., 1993. Essay on Contemporary Issues in Ethology: Variation among Mammalian Alarm Call Systems and the Problem of Meaning in Animal Signals. *Ethology* 93, 177–197. <https://doi.org/10.1111/j.1439-0310.1993.tb00988.x>
- Masataka, N., Biben, M., 1987. Temporal Rules Regulating Affiliative Vocal Exchanges of Squirrel Monkeys. *Behaviour* 101, 311–319. <https://doi.org/10.1163/156853987X00035>
- Meguerditchian, A., Molesti, S., Vauclair, J., 2011. Right-handedness predominance in 162 baboons (*Papio anubis*) for gestural communication: Consistency across time and groups. *Behav. Neurosci.* 125, 653–660. <https://doi.org/10.1037/a0023823>
- Meguerditchian, A., Vauclair, J., Hopkins, W.D., 2013. On the origins of human handedness and language: A comparative review of hand preferences for bimanual coordinated actions and gestural communication in nonhuman primates. *Dev. Psychobiol.* 55, 637–650. <https://doi.org/10.1002/dev.21150>
- Mehl, M.R., Vazire, S., Ramirez-Esparza, N., Slatcher, R.B., Pennebaker, J.W., 2007. Are women really more talkative than men? *Science* 317, 82. <https://doi.org/10.1126/science.1139940>
- Meltzoff, A., Moore, M.K., 1977. Imitation of Facial and Manual Gestures by Human Neonates. *Science.* <https://doi.org/10.1126/science.198.4312.75>
- Neumann, C., Zuberbühler, K., 2016. Vocal correlates of individual sooty mangabey travel speed and direction. *PeerJ* 4, e2298. <https://doi.org/10.7717/peerj.2298>
- 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. *Proc. Natl. Acad. Sci.* 110, 6318–6323. <https://doi.org/10.1073/pnas.1300337110>
- Oshima-Takane, Y., Robbins, M., 2003. Linguistic Environment of Secondborn Children. *First Lang.* 23, 21–40. <https://doi.org/10.1177/0142723703023001002>

- Ouattara, K., Lemasson, A., Zuberbühler, K., 2009. Campbell's monkeys use affixation to alter call meaning. *PLOS ONE* 4, e7808.
- Owren, M.J., Rendall, D., 2001. Sound on the Rebound: Bringing Form and Function Back to the Forefront in Understanding Nonhuman Primate Vocal Signaling. *Evol. Anthropol.* 10, 58–71.
- Patterson, F.G., 1978. The gestures of a gorilla: Language acquisition in another pongid. *Brain Lang.* 5, 72–97. [https://doi.org/10.1016/0093-934X\(78\)90008-1](https://doi.org/10.1016/0093-934X(78)90008-1)
- Pereira, M.E., Macedonia, J.M., 1991. Ringtailed lemur anti-predator calls denote predator class, not response urgency. *Anim. Behav.* 41, 543–544. [https://doi.org/10.1016/S0003-3472\(05\)80861-9](https://doi.org/10.1016/S0003-3472(05)80861-9)
- Pika, S., Wilkinson, R., Kendrick, K.H., Vernes, S.C., 2018. Taking turns: bridging the gap between human and animal communication. *Proc. R. Soc. B Biol. Sci.* 285, 20180598. <https://doi.org/10.1098/rspb.2018.0598>
- Pine, J.M., 1995. Variation in Vocabulary Development as a Function of Birth Order. *Child Dev.* 66, 272. <https://doi.org/10.2307/1131205>
- Pougnault, L., Levréro, F., Lemasson, A., 2020a. Conversation Among Primate Species, in: Masataka, N. (Ed.), *The Origins of Language Revisited: Differentiation from Music and the Emergence of Neurodiversity and Autism*. Springer, Singapore, pp. 73–96. https://doi.org/10.1007/978-981-15-4250-3_4
- Pougnault, L., Levréro, F., Leroux, M., Paulet, J., Bombani, P., Dentressangle, F., Deruti, L., Mulot, B., Lemasson, A., 2022. Social pressure drives “conversational rules” in great apes. *Biol. Rev. Camb. Philos. Soc.* 97, 749–765. <https://doi.org/10.1111/brv.12821>
- Pougnault, L., Levréro, F., Mulot, B., Lemasson, A., 2020b. Breaking conversational rules matters to captive gorillas: A playback experiment. *Sci. Rep.* 10, 6947. <https://doi.org/10.1038/s41598-020-63923-7>
- Pougnault, L., Levréro, F., Mulot, B., Lemasson, A., 2020c. Breaking conversational rules matters to captive gorillas: A playback experiment. *Sci. Rep.* 10, 1–10. <https://doi.org/10.1038/s41598-020-63923-7>
- Price, T., Wadewitz, P., Cheney, D., Seyfarth, R., Hammerschmidt, K., Fischer, J., 2015. Vervets revisited: A quantitative analysis of alarm call structure and context specificity. *Sci. Rep.* 5, 1–11. <https://doi.org/10.1038/srep13220>
- Prieur, J., Barbu, S., Blois-Heulin, C., Lemasson, A., 2020. The origins of gestures and language: history, current advances and proposed theories. *Biol. Rev.* 95, 531–554. <https://doi.org/10.1111/brv.12576>
- Rapaport, L.G., 2011. Progressive parenting behavior in wild golden lion tamarins. *Behav. Ecol.* 22, 745–754. <https://doi.org/10.1093/beheco/arr055>
- Robinson, L.F., Reis, H.T., 1989a. The effects of interruption, gender, and status on interpersonal perceptions. *J. Nonverbal Behav.* 13, 141–153. <https://doi.org/10.1007/BF00987046>

- Robinson, L.F., Reis, H.T., 1989b. The effects of interruption, gender, and status on interpersonal perceptions. *J. Nonverbal Behav.* 13, 141–153. <https://doi.org/10.1007/BF00987046>
- Rossano, F., 2018. Social manipulation, turn-taking and cooperation in apes. *Interact. Stud.* 19, 151–166. <https://doi.org/10.1075/is.17043.ros>
- Rowe, M.L., 2008. Child-directed speech: relation to socioeconomic status, knowledge of child development and child vocabulary skill*. *J. Child Lang.* 35, 185–205. <https://doi.org/10.1017/S0305000907008343>
- Sacks, H., Schegloff, E.A., Jefferson, G., 1974. A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language* 50, 696–735. <https://doi.org/10.2307/412243>
- Schegloff, E.A., 2000. Overlapping talk and the organization of turn-taking for conversation. *Lang. Soc.* 29, 1–63. <https://doi.org/10.1017/s0047404500001019>
- Schel, A.M., Bono, A., Aychet, J., Pika, S., Lemasson, A., 2022. Intentional gestural communication amongst red-capped mangabeys (*Cercocebus torquatus*). *Anim. Cogn.* <https://doi.org/10.1007/s10071-022-01615-7>
- Schel, A.M., Townsend, S.W., Machanda, Z., Zuberbühler, K., Slocombe, K.E., 2013. Chimpanzee Alarm Call Production Meets Key Criteria for Intentionality. *PLOS ONE* 8, e76674. <https://doi.org/10.1371/journal.pone.0076674>
- Schlenker, P., Chemla, E., Schel, A.M., Fuller, J., Gautier, J.-P., Kuhn, J., Veselinović, D., Arnold, K., Căsar, C., Keenan, S., Lemasson, A., Ouattara, K., Ryder, R., Zuberbühler, K., 2016. Formal monkey linguistics. *Theor. Linguist.* 42, 1–90. <https://doi.org/10.1515/tl-2016-0001>
- Schwab, J.F., Lew-Williams, C., 2016. Language learning, socioeconomic status, and child-directed speech. *WIREs Cogn. Sci.* 7, 264–275. <https://doi.org/10.1002/wcs.1393>
- Scott-Phillips, T.C., 2015. Nonhuman Primate Communication, Pragmatics, and the Origins of Language. *Curr. Anthropol.* 56, 56–80. <https://doi.org/10.1086/679674>
- Seyfarth, R.M., Cheney, D.L., Marler, P., 1980a. Vervet monkey alarm calls: Semantic communication in a free-ranging primate. *Anim. Behav.* 28, 1070–1094. [https://doi.org/10.1016/S0003-3472\(80\)80097-2](https://doi.org/10.1016/S0003-3472(80)80097-2)
- Seyfarth, R.M., Cheney, D.L., Marler, P., 1980b. Monkey Responses to Three Different Alarm Calls: Evidence of Predator Classification and Semantic Communication. *Science* 210, 801–803. <https://doi.org/10.1126/science.7433999>
- Shannon, C.E., 1948. A mathematical Theory of Communication. *Bell Syst. Tech. J.* 27, 379–423, 623–656.
- Shneidman, L.A., Arroyo, M.E., Levine, S.C., Goldin-Meadow, S., 2013. What counts as effective input for word learning? *J. Child Lang.* 40, 672–686. <https://doi.org/10.1017/S0305000912000141>
- Slocombe, K.E., Zuberbühler, K., 2006. Food-associated calls in chimpanzees: responses to food types or food preferences? *Anim. Behav.* 72, 989–999. <https://doi.org/10.1016/j.anbehav.2006.01.030>

- Snow, C.E., 1977. The development of conversation between mothers and babies. *J. Child Lang.* 4, 1–22. <https://doi.org/10.1017/S0305000900000453>
- Snowdon, C.T., Cleveland, J., 1984. “Conversations” among pygmy marmosets. *Am. J. Primatol.* 7, 15–20. <https://doi.org/10.1002/ajp.1350070104>
- Sperber, A.L., Werner, L.M., Kappeler, P.M., Fichtel, C., 2017. Grunt to go—Vocal coordination of group movements in redfronted lemurs. *Ethology* 123, 894–905. <https://doi.org/10.1111/eth.12663>
- Stern, D.N., Jaffe, J., Beebe, B., Bennett, S.L., 1975. Vocalizing in unison and in alternation: Two modes of communication within the mother- infant dyad. *Ann. N. Y. Acad. Sci.* 263, 89–100. <https://doi.org/10.1111/j.1749-6632.1975.tb41574.x>
- Stivers, T., Enfield, N.J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., Ruitter, J.P. de, Yoon, K.-E., Levinson, S.C., 2009. Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci.* 106, 10587–10592. <https://doi.org/10.1073/pnas.0903616106>
- Striano, T., Henning, A., Stahl, D., 2005. Sensitivity to social contingencies between 1 and 3 months of age. *Dev. Sci.* 8, 509–518. <https://doi.org/10.1111/j.1467-7687.2005.00442.x>
- Sugiura, H., 1993. Temporal and Acoustic Correlates in Vocal Exchange of Coo Calls in Japanese Macaques. *Behaviour* 124, 3–4. <https://doi.org/10.1163/156853993X00588>
- Sung, J., Fausto-Sterling, A., Coll, C.G., Seifer, R., 2013. The Dynamics of Age and Sex in the Development of Mother–Infant Vocal Communication Between 3 and 11 Months. *Infancy* 18, 1135–1158. <https://doi.org/10.1111/inf.12019>
- Swedell, L., 2010. African papionins: diversity of social organization and ecological flexibility, in: Campbell, C.J., Fuentes, A., MacKinnon, K.C., Panger, M., Bearder, S.K. (Eds.), *Primates in Perspectives*. New York, pp. 241–277.
- Takahashi, D.Y., Narayanan, D.Z., Ghazanfar, A.A., 2013. Coupled Oscillator Dynamics of Vocal Turn-Taking in Monkeys. *Curr. Biol.* 23, 2162–2168. <https://doi.org/10.1016/j.cub.2013.09.005>
- ter Maat, M., Truong, K.P., Heylen, D., 2010. How Turn-Taking Strategies Influence Users’ Impressions of an Agent, in: Allbeck, J., Badler, N., Bickmore, T., Pelachaud, C., Safonova, A. (Eds.), *Intelligent Virtual Agents, Lecture Notes in Computer Science*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 441–453. https://doi.org/10.1007/978-3-642-15892-6_48
- Terrace, H.S., Petitto, L.A., Sanders, R.J., Bever, T.G., 1979. Can an Ape Create a Sentence? *Science* 206, 891–902. <https://doi.org/10.1126/science.504995>
- Tinbergen, N., 1963. On aims and methods of Ethology. *Z. Für Tierpsychol.* 20, 410–433.
- Tomasello, M., Call, J., 2019. Thirty years of great ape gestures. *Anim. Cogn.* 22, 461–469. <https://doi.org/10.1007/s10071-018-1167-1>
- Waciewicz, S., Żywicznyński, P., 2018. Language origins: Fitness consequences, platform of trust, cooperation, and turn-taking 13.

- Waciewicz, S., Żywiczyński, P., Chiera, A., 2017. An evolutionary approach to low-level conversational cooperation. *Lang. Sci., Language Evolution: Focus on Mechanisms* 63, 91–104. <https://doi.org/10.1016/j.langsci.2017.01.005>
- Werker, J.F., Tees, R.C., 1984. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behav. Dev.* 7, 49–63. [https://doi.org/10.1016/S0163-6383\(84\)80022-3](https://doi.org/10.1016/S0163-6383(84)80022-3)
- Wheeler, B.C., 2010. Production and perception of situationally variable alarm calls in wild tufted capuchin monkeys (*Cebus apella nigrinus*). *Behav. Ecol. Sociobiol.* 64, 989–1000. <https://doi.org/10.1007/s00265-010-0914-3>
- Wheeler, B.C., Fischer, J., 2012. Functionally referential signals: A promising paradigm whose time has passed. *Evol. Anthropol.* 21, 195–205. <https://doi.org/10.1002/evan.21319>
- Wheeler, B.C., Hammerschmidt, K., 2013. Proximate Factors Underpinning Receiver Responses to Deceptive False Alarm Calls in Wild Tufted Capuchin Monkeys: Is It Counterdeception? *Am. J. Primatol.* 75, 715–725. <https://doi.org/10.1002/ajp.22097>
- Wilson, B., Slater, H., Kikuchi, Y., Milne, A.E., Marslen-Wilson, W.D., Smith, K., Petkov, C.I., 2013. Auditory artificial grammar learning in Macaque and Marmoset monkeys. *J. Neurosci.* 33, 18825–18835. <https://doi.org/10.1523/JNEUROSCI.2414-13.2013>
- Wilson, D., Sperber, D., 1979. Remarques sur l'interprétation des énoncés selon Paul Grice. *Communications* 30, 80–94. <https://doi.org/10.3406/comm.1979.1448>
- Yamaguchi, C., Izumi, A., Nakamura, K., 2009. Temporal rules in vocal exchanges of phee and trills in common marmosets (*Callithrix jacchus*). *Am. J. Primatol.* 71, 617–622. <https://doi.org/10.1002/ajp.20697>
- Yoo, H., Bowman, D.A., Oller, D.K., 2018. The Origin of Protoconversation: An Examination of Caregiver Responses to Cry and Speech-Like Vocalizations. *Front. Psychol.* 9. <https://doi.org/10.3389/fpsyg.2018.01510>
- Yoshida, S., Okanoya, K., 2005. Evolution of turntaking: a bio-cognitive perspective. *Cogn. Stud.* 12, 153–165.
- Zambrana, I.M., Ystrom, E., Pons, F., 2012. Impact of Gender, Maternal Education, and Birth Order on the Development of Language Comprehension: A Longitudinal Study from 18 to 36 Months of Age. *J. Dev. Behav. Pediatr.* 33, 146–155. <https://doi.org/10.1097/DBP.0b013e31823d4f83>
- Zanoli, A., Gamba, M., Lemasson, A., Norscia, I., Palagi, E., 2021. Inter-sexual multimodal communication during mating in wild geladas: the leading role of females. *Curr. Zool.* zoab093. <https://doi.org/10.1093/cz/zoab093>
- Zuberbühler, K., 2014. Experimental field studies with non-human primates. *Curr. Opin. Neurobiol.* 28, 150–156. <https://doi.org/10.1016/j.conb.2014.07.012>
- Zuberbühler, K., 2002. A syntactic rule in forest monkey communication. *Anim. Behav.* 63, 293–299. <https://doi.org/10.1006/anbe.2001.1914>

Zuberbuhler, K., 2000. Referential labelling in Diana monkeys. *Anim. Behav.* 59, 917–927.

Titre : Aux prémices de nos conversations : une étude comparative chez les nourrissons humains et les primates non-humains

Mots clés : tour de rôle, proto-conversation, langage, socialité

Résumé : Les conversations forment la niche principale à l'intérieur de laquelle le langage se déploie et prend son sens. Elles se caractérisent par un tour de parole : une alternance de tours entre plusieurs interlocuteurs qui évitent les chevauchements et minimisent les silences. Ce pattern universel semble avoir des bases biologiques profondes : il est présent très précocement chez les nourrissons préverbaux, ainsi que dans les échanges vocaux de primates non-humains. Nous avons donc voulu savoir dans quelle mesure le tour de rôle faisait sens pour des êtres non-parlants en étudiant des nourrissons humains et des mangabeys à collier (*Cercocebus torquatus*). Pour ces deux modèles biologiques, nous avons analysé d'une part la production des tours de rôles vocaux grâce à des observations, et d'autre part leur perception par les individus qui les pratiquent grâce à des expériences. Ainsi, nous avons pu montrer que les nourrissons préverbaux étaient sensibles aux variations liées au tour de rôle, et

que cette sensibilité dépendait de facteurs démographiques tels que statut socioéconomique, la présence d'une fratrie ou le mode de garde. Nous avons aussi mis en évidence la production de vocalisation à tour de rôle lors d'interactions parents-nourrissons différaient en fonction du sexe du parent ainsi que du nourrisson. Chez les mangabeys, la production de vocalisations à tour de rôle dépendait du statut social des individus et des liens avec les congénères, définis par l'affinité et la dominance. Toutefois, nous n'avons pas réussi à mettre en évidence une sensibilité de la part des individus testés aux variations liés au tour de rôle. Ces résultats confirment la préséance évolutive et développementale des compétences conversationnelles sur le langage, et montre l'importance de l'approche comparative pour éclairer les bases biologiques de nos comportements communicatifs.

Title: The beginnings of our conversations: a comparative study of human infants and non-human primates

Keywords: turn-taking, protoconversations, language, sociality

Abstract: Conversations constitute the main niche within which language evolves and acquires meaning. They are characterized by turn-taking: an alternation of turns between several interlocutors that avoids overlaps and minimizes silences. This universal pattern seems to have deep biological foundations: it occurs very precociously in preverbal infants, as well as in the vocal exchanges of non-human primates. We therefore investigated the extent to which turn-taking makes sense for non-speaking beings by studying human infants and red-capped mangabeys (*Cercocebus torquatus*). For these two biological models, we analyzed both the production of vocal turn-taking through observations, and its perception by the individuals performing it through experiments. We showed that preverbal infants were sensitive to variations

in turn-taking, and that this sensitivity depended on demographic factors such as socioeconomic status, the presence of siblings and daycare. We also found that the production of vocalizations in turn-taking during parent-infant interactions differed according to the sex of both parent and infant. In mangabeys, the production of vocalizations in turn-taking depended on the social status of individuals and their links with conspecifics, defined by affinity and dominance. However, we were unable to demonstrate any sensitivity of the individuals tested to turn-taking variations. These results confirm the evolutionary and developmental precedence of conversational skills over language, and underline the importance of the comparative approach in elucidating the biological basis of our communicative behaviours.