NNT : 0000AIXM0000

THÈSE DE DOCTORAT

Soutenue à Aix-Marseille Université
le 28 novembre 2024 par

Floor MEEWIS

In search of linguistic features in event representations of non-human primates and preverbal infants

Discipline
Sciences cognitives

École doctorale
356 Cognition, Langage, Éducation

Laboratoire/Partenaires de recherche
Centre de Recherche en Psychologie et Neurosciences
Centre Nationale de la Recherche Scientifique

Composition du jury

Gergely CSIBRA — Rapporteur
Professor; Central European University, Vienna, Austria

Klaus ZUBERBÜHLER — Rapporteur
Professor; University of Neuchâtel, Neuchâtel, Switzerland; University of Zurich, Zürich, Switzerland; University of St Andrews, Scotland

Jennifer CULBERTSON — Examinatrice
Professor; The university of Edinburgh, Edinburgh, Scotland

Liuba PAPEO — Présidente du jury
Directrice de recherche; CNRS, Claude Bernard University Lyon 1, Lyon, France

Isabelle DAUTRICHE — Co-directrice de thèse
Chargée de recherche; CNRS, Aix Marseille University, Marseille, France

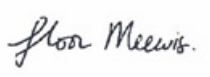
Nicolas CLAIDIÈRE — Directeur de thèse
Directeur de recherche; CNRS, Aix Marseille University, Marseille, France

Affidavit

I, undersigned, Floor MEEWIS, hereby declare that the work presented in this manuscript is my own work, carried out under the scientific supervision of Isabelle DAUTRICHE and Nicolas CLAIDIÈRE, in accordance with the principles of honesty, integrity and responsibility inherent to the research mission. The research work and the writing of this manuscript have been carried out in compliance with both the French national charter for Research Integrity and the Aix-Marseille University charter on the fight against plagiarism.

This work has not been submitted previously either in this country or in another country in the same or in a similar version to any other examination body.

Place: Marseille, date: 3 octobre 2024



Cette œuvre est mise à disposition selon les termes de la Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International.

List of publications and participation in conferences

1) List of publications produced as part of the thesis project:

    1. Meewis, F., Barezzi, I., Fagot, J., Claidière, N. & Dautriche, I. (accepted). A comparative study of causal perception in Guinea baboons (*Papio papio*) and human adults.

    2. Meewis, F., Barezzi, I., Hababou-Bernson, M., Fagot, J., Claidière, N. & Dautriche, I. (submitted). No evidence for agent-patient role attribution in human infants, human adults and Guinea baboons (*Papio papio*).

    3. Meewis, F., Fagot, J., Claidière, N. & Dautriche, I. (under review). Agent preference in chasing interactions in Guinea baboons (*Papio papio*): uncovering the roots of subject-object order in language.

2) Participation in conferences and summer schools during the thesis period:
    1. Summer school of the Institute of Language, Communication and the Brain 2021
    2. Summer school of the Institute of Language, Communication and the Brain 2022
    3. Colloque de la Société Francophone de Primatologie 2022 (poster)
    4. Budapest CEU Conference on Cognitive Development 2023 (oral presentation)
    5. Retreat of the Institute of Language, Communication and the Brain 2023 (oral presentation)
    6. Workshop Hand, Voice & Brain: Ontogeny and Evolution 2023 (oral presentation)
    7. Annual Meeting of the Cognitive Science Society 2023 (poster (online))
    8. Journée des Babylabs Français 2023 (oral presentation)
    9. Budapest CEU Conference on Cognitive Development 2024 (oral presentation)
    10. European Federation of Primatology meeting 2024 (oral presentation)
    11. Annual Meeting of the Cognitive Science Society 2024 (oral presentation)

Résumé

L'évolution du langage peut être étudiée en comparant les primates humains et non-humains afin d'identifier les éléments du langage pouvant être hérités d'un ancêtre commun. Chez les humains, il est établi que la description des événements montre des parallèles frappants avec les représentations d'événements non-linguistiques, soulevant ainsi la question de savoir si les primates non-humains traitent et représentent les événements de manière similaire. Cette thèse explore donc si les représentations d'événements chez les primates non-humains contiennent des composants pertinents pour le langage.

L'accent est mis sur les représentations d'événements dynamiques à deux acteurs, car celles-ci sont exprimées de manière étonnamment cohérente à travers les langues : en mentionnant d'abord l'agent (celui qui fait) avant le patient (celui qui subit), en mettant l'accent sur la relation causale. Trois études ont été menées pour explorer la sensibilité à la causalité et aux rôles agent-patient chez les babouins de Guinée (*Papio papio*), les nourrissons humains et les adultes. La première étude a montré que, tandis que les adultes humains catégorisent les événements de Michotte en fonction de la causalité, les babouins ne le font pas. La deuxième étude n'a révélé aucune sensibilité aux rôles agent-patient chez les nourrissons, les adultes ou les babouins, remettant en cause la fiabilité de la méthode d'inversion de rôle pour l'attribution des rôles. La troisième étude a révélé que les babouins montrent un biais attentionnel envers l'agent dans les interactions de poursuite, en accord avec les tendances linguistiques.

Nous n'en sommes qu'au début de la découverte de l'étendue des représentations d'événements chez les primates non-humains. Ces résultats préliminaires suggèrent que la capacité humaine à transformer le monde extérieur en représentations mentales pourrait être héritée d'un ancêtre commun, mais la capacité de l'exprimer par le langage pourrait être propre aux humains.

Mots-clés : cognition des événements, représentation des événements, évolution du langage, approche comparative, causalité, agent-patient

Abstract

One way to study the evolution of language is by comparing human and non-human primates to identify which parts of language may be inherited from a common ancestor. In humans, it has been established that event descriptions show striking parallels to non-linguistic event representations, leading to the question whether non-human primates may process and represent events in a similar way. This thesis thus explores whether event representations in non-human primates contain linguistically relevant components.

A specific focus lies on the representations of two actor dynamic events, because these are expressed in surprisingly consistent ways across languages: by mentioning the doer, agent, before the undergoer, patient, of the event and focussing on the causal relation between them. Three studies were conducted to explore the sensitivity to causality and agent-patient roles in Guinea baboons (*Papio papio*), human infants and human adults. The first study demonstrated that while human adults categorise Michottean events according to causality, Guinea baboons do not. The second study found no sensitivity to agent-patient roles in preverbal infants, adults, or baboons, using a well-established role reversal method with Michottean events, calling into question the reliability of this method for event role attribution. The third study revealed that Guinea baboons possess an attentional bias towards the agent in chasing interactions which aligns with crosslinguistic tendencies.

Currently, we stand only at the beginning of uncovering to what extent event representations in non-human primates contain linguistically relevant features. These initial findings suggest that human's ability to transform the outside world into mental representations may be inherited from a shared ancestor, but the ability to express this using language may be uniquely human.

Keywords: event cognition, event representation, evolution of language, comparative approach, causality, agent-patient

Acknowledgements

First, of course, I would like to thank my thesis supervisors Isabelle Dautriche and Nicolas Claidière for all of their guidance over the past three years. Isabelle, you have been an incredible mentor. I am very appreciative of the attentive support you have given me throughout this project. It has been a pleasure to learn from your vast knowledge, originality and grounded approach. Thanks for encouraging me to see and describe (!) exciting results for what they are: exciting! Nico, you have been a constant, reliable pillar of support, thanks for stepping up at the right moments. And most of all, thank you for your amazingly positive outlook.

Next, I would like to thank Gergely Csibra, Klaus Zuberbühler, Jennifer Culbertson and Liuba Papeo for accepting to be in my thesis jury. Spanning a range of different disciplines, I am excited to learn about your viewpoints on the questions that I tried to tackle in this thesis. I am looking forward to receiving your feedback and discuss with you.

I am very happy to have had Arnaud Rey and Brent Strickland participating in my Comité Suivi Individuel de Thèse. It was great to receive some insightful feedback and encouraging words along the way.

Thanks to the members of the old Cognition Comparée team and the new DePhy team for your feedback on experimental designs and ongoing projects. I owe a special thanks to Joël Fagot, who shared his expertise of designing baboon experiments and writing papers, and stimulated me to come up with experiments that were not too complicated.

Another thanks for all the help I received with the experiments with the baboons from Julie Gullstrand and Sebastien Barniaud and with the babies from Marielle Hababou-Bernson and Caroline Coindre. It could not have been possible without you!

Merci à Colette Pourpe, Fred Lombardo and Nadia Melili de m'avoir aidé à naviguer dans la bureaucratie du système universitaire français. Ce n'était pas du tout facile !

Also a big thanks to all the parents who participated with their kids.

And a special shout out to the amazing coworkers that were there all along the last three years. And to those who helped me through the writing process, Diya (Shreejata) Gupta, Siham Bouziane, Marielle Hababou-Bernson again, Leo(nardo) Pinto Arata, Jacob Maaz and Florence Bouhali, I know I have told you this many times over the last weeks, but I can do it again a hundred times: thank you!

Table of contents

Affidavit 1
List of publications and participation in conferences 2
Résumé 3
Abstract 4
Acknowledgements 5
Table of contents 6
1. Introduction 9
Aims & research question 11
1.1. Cognition and language interface 12
1.1.1. Revealing the putative cognitive underpinnings of language universals 13
1.2. Event representations 18
1.2.1. Event cognition in human adults 18
1.2.2. Event cognition in preverbal infants 23
1.2.3. Event cognition in non-human primates 27
1.3. Outline of this thesis 30
1.3.1. Participants 30
1.3.2. A word regarding methodological considerations 32
2. Article 1 35
A comparative study of causal perception in Guinea baboons (*Papio papio*) and human adults 35
Abstract 35
Background 36
Methods 38
Ethics statement 38
Participants 39
Procedure 39
Experimental phases and conditions 41
Statistical analyses 43
Results 45
Discussion 48
Acknowledgments 52
References 52
Supporting information 55
3. Article 2 58
No evidence for agent-patient role attribution in human infants, human adults and Guinea baboons (*Papio papio*) 58

Abstract 58
Introduction 59
Methods 62
Ethics statement 63
Participants 63
Stimuli 64
Procedure and design 66
Analyses 68
Results 70
Human infants 70
Baboons 71
Human adults 71
Discussion 72
Acknowledgements 75
References 75
Supplementary information 78
4. Article 3 **80**
Agent preference in chasing interactions in Guinea baboons (*Papio papio*): uncovering the roots of subject-object order in language 80
Abstract 80
Introduction 81
Method 84
Ethics statement 84
Participants 84
General procedure 84
Stimuli and conditions 85
Analyses 87
Results 88
Training phase 88
Test phase 88
Discussion 91
Acknowledgements 92
References 93
Supporting Information 98
5. Discussion **100**
5.1. Summarising the main findings 100
5.1.1. Contributions and limitations 100
5.2. Implications 104
5.2.1. Linking the current findings to language evolution 105
5.3. Outstanding questions and reflections 107

5.3.1. What constitutes an agent in the eyes and mind of a Guinea baboon or a human infant? 107
5.3.2. Beyond agent and patient event roles 108
5.3.3. A brief examination of static topological events 109
5.3.4. Why are many mental representations not expressed in non-human communication? 111
5.4. Conclusion 114
References **115**
ANNEXES **131**
A. Cogsci 2023 proceedings paper 132

1. Introduction

Language is used on a daily basis, often effortlessly, and mastered early in life. Yet, at the same time, it is highly complex and, as such, distinct from communication systems in other animals. Humans are the only species who have language as part of their communicative repertoire. This makes language one of the defining features of humankind and has led many to wonder how this capacity came about. The answer to this question remains elusive, but one angle to take, is to compare in what respects language is similar to and different from communication systems of other, non-human, species. As such, theories about the evolution of language draw heavily on comparative evidence from non-human primates to identify homologues of language that may have evolved gradually from a recent common ancestor.

This comparative approach asks for a pluralistic view by considering language to be a composite of many relevant features and subsequently trying to determine which of those are shared with non-human species and which are not (e.g. Fitch, 2010). Capturing this theoretically, some have argued that there are certain unique features in humans that are critically lacking in other animals. Famously, Hockett drew up a list of what he deemed 'design features of language', and pointed out that some were shared by other animals, and a few that were unique to humans (Hockett, 1960). For example, he claimed that many species share the use of the vocal-auditory channel for communication, but a feature like displacement, the ability to talk about things beyond the present time and place, might be an exclusively human characteristic. This line of thinking perseveres; for example, Hauser *et al.* (2002) introduced the faculty of language 'in the broad sense' which encompasses all of the mechanisms involved in language (acquisition and use), of which many may be shared with other species, and contrasted this with the faculty of language 'in the narrow sense' which was meant to highlight those mechanisms which are uniquely human. They proposed that only the feature of recursion (hierarchical self-embedded structures) may make up the faculty of language in the narrow sense.

In search for commonalities in the communication systems of humans and non-humans, vocal and gestural communication in non-human primates has gained the most research attention (e.g. Arbib et al., 2008). Vocalisations are sometimes even approached from an explicitly linguistic viewpoint to ameliorate comparisons (Schlenker, Chemla, Schel, et al., 2016; Schlenker, Chemla, & Zuberbühler, 2016). Most research endeavours concerning vocalisations, have led to the consensus that non-human primates are rather restricted when it comes to production of calls, which are found to be largely genetically determined and part of a small repertoire (Cheney & Seyfarth, 2005; Fedurek & Slocombe, 2011). This does not mean that vocalisations are completely rigid and automatic; some species have been shown to withhold calls (e.g. le Roux et al., 2013), produce them intentionally (Schel et al., 2013) or combine them into sequences (e.g. Arnold & Zuberbühler, 2006), although the syntactical functioning of these is debated (Schlenker et

al., 2023). On the perception side, it has become clear that listeners can extract 'meaning' about an external event, as evidenced by a specific response, based on the acoustics and emitting circumstances, termed as 'functionally referential calls' (Macedonia & Evans, 1993; Townsend & Manser, 2013). The famous example of functional reference concerns the three alarm calls for different predators that all result in distinct, adaptive responses in vervet monkeys (*Chlorocebus pygerythrus*) (Seyfarth et al., 1980). More flexible use in production can be found in the gestural domain, where intentionality has been evidenced for many gestures of great apes as a sensitivity to the attention of the receiver and persistence when the goal has not been reached (Byrne et al., 2017).

Because external expressions are necessarily required for communication, much research attention has naturally been directed towards these visible and audible signals. However, this may present a somewhat narrow view, since language draws on many cognitive capacities, such as social cognition, perception, memory, categorisation etc., that are not immediately evident from this external form. It is therefore a fruitful complementary approach to look for commonalities between humans and other primates regarding the cognitive functions that support language (e.g. Fagot et al., 2019; Premack, 1976). Shared cognitive functioning may culminate in a (partly) shared mental representational system, whereas the ability to express this through communication may very well be unique to humans (e.g. Fitch, 2019; Hurford, 2007; Sperber & Wilson, 1986 p. 173).

Attempts to probe mental representations which are expressed linguistically in humans, in non-human primates, have been proven fruitful in many different domains. For example, non-human primates represent singular vs. plural (Barner et al., 2008), objects as kinds (Phillips & Santos, 2007), same vs. different (Fagot & Thompson, 2011; Thompson et al., 1997), numbers (Matsuzawa, 1985), the logical concept of connectedness (Chemla et al., 2019), and can manipulate their representations, for example to make transitive inferences (Jensen et al., 2019) and combine them compositionally (Dautriche et al., 2022).

These studies thus illustrate that, although it is not reflected in their communication, non-human primates appear to possess a rich mental life. However, a more systematic theory explaining which areas of cognitive representations relating to language functioning in humans are likely to be shared, would increase explanatory power and generate testable predictions. Recently, it has been argued that representations of events could provide such an area that is conserved across species and highly relevant for humans' linguistic productions (e.g. Wilson et al., 2022).

Given that the correspondence between linguistic expressions and the external world is mediated by the mind, representations of events are thought to be highly influential for language structure in humans by determining which event components are delineated, which are given priority and which will be subsequently expressed (Rissman & Majid, 2019; Ünal, Ji, et al., 2021). In this way, cognitive biases, such as those derived from event cognition, may have played in important role in the emergence of language universals through language evolution (Strickland, 2017).

To understand language evolution, we should thus aim to understand whether cognitive biases may have shaped the structure of human languages, and whether these cognitive roots may be present in other species as well.

Aims & research question

The present thesis is concerned with the question **whether event representations in non-human primates contain linguistically relevant components**. Given the link between linguistic descriptions and event cognition (Strickland, 2017; Ünal, Ji, et al., 2021; Wilson et al., 2022), event representations present a promising avenue to explore the cognitive roots of language that may exist even in the absence of language. By adopting a comparative approach between Guinea baboons (*Papio papio*) and preverbal infants, this thesis aims to shed light on whether the way of representing events is shared across species and could have given rise to language structures in humans.

1.1. Cognition and language interface

Language provides a way to describe the world; yet it does not present an exact one-to-one mapping to the outside world. Some parts of the event will be captured in the linguistic expression, whereas others may not. For example, when describing a scene as "The PhD student writes her dissertation.", this excludes a lot of information in the description that is present in the actual event, such as whether she uses pen and paper or a computer to do so, whether she does this at her desk, or what her mental state is like during the writing process. Different languages across the world showcase high variability in their event descriptions, but they also share some salient similarities, for example with regards to the above-chance frequency of specific word orders over others (e.g. in the noun phrase: Martin et al., 2024) or of the conversational roles that are expressed (Maldonado & Culbertson, 2022). Languages do not vary randomly, it rather appears that this variation is to some extent limited making languages look complexly ordered. The structural similarities across languages are known as language universals. Famously, Greenberg noted a list of such universals, based on a language sample of thirty languages (Greenberg, 1963). This propelled the idea that certain linguistic features have cooccur across different, even un- or distantly related, languages. Although the notion of language universals is debated (e.g. Evans & Levinson, 2009), many agree that statistical universals, where some patterns are probabilistically more likely to occur than others, rather than absolute ones, exist (e.g. Christiansen et al., 2009).

What are the possible origins of language universals? There are several non-mutually exclusive factors that can contribute to explaining these observed patterns. They may be relics that are the result of language contact and history (Dunn et al., 2011). Another possibility is that they have arisen through cultural evolution amongst speakers (Kirby et al., 2008). There are also functionally motivated explanations; according to a usage-based account, similar grammars will arise from similar, conventionalised usage across cultures (Bybee, 2006, 2009). Also communicative efficiency may be a driving factor (Gibson et al., 2019). Another intriguing possibility is that language universals are the result of how the human mind operates (Chomsky, 1957; Christiansen & Chater, 2008). Such cognitive capacities, which could have evolved for non-linguistic purposes, may culminate in biases driving linguistic patterns (Culbertson & Kirby, 2016; Martin et al., 2024; Strickland, 2017).

Here, I will focus on the specific language universals regarding the expression of simple dynamic two actor events, in which one of those actors acts on the other. Many languages express causality using causative syntax (Dixon, 2000; Song, 2013ab) and in active transitive clauses this includes three elements: subject (S), object (O) and verb (V). Given that there are six possible orders of these three building blocks, it is remarkable that frequencies of occurrence of these orders are far from uniformly distributed cross-linguistically. Using a sample of 1187 diverse languages with dominant word order from across the world, Dryer (2013) mapped out that SOV and SVO are disproportionately prevalent (48% and 41% respectively), VSO is less common (8%), and in particular VOS (2%), OVS (1%) and OSV (<1%) are found to be rare orderings, see Figure 1. Similarly,

generalising from a sample of 42 sign languages, it was concluded that SOV is always grammatical and used most frequently, followed by SVO (Napoli & Sutton-Spence, 2014). This reveals a clear pattern: in 97% of languages, the subject precedes the object, reflecting Greenberg's Universal 1 (1963). This word ordering is thus a convincing (statistical) language universal that calls for explanations of why it is so pervasive.

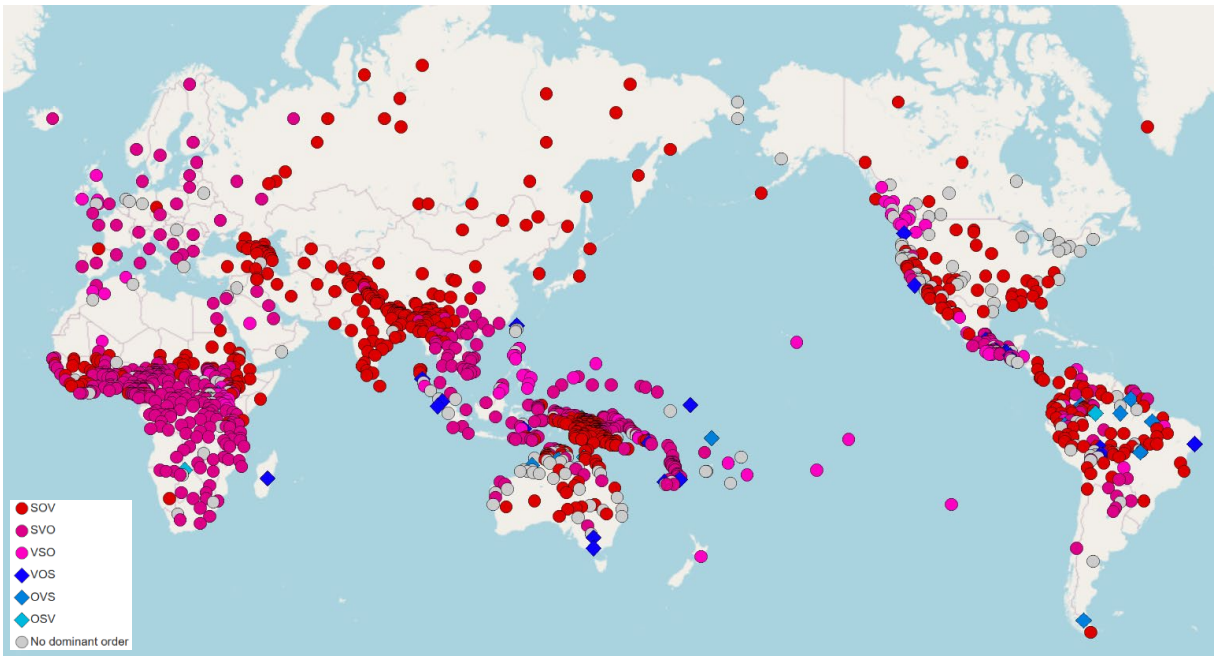


Figure 1: This map, taken from the World Atlas of Language Structures, shows the ordering of subject (S), object (O), and verb (V) in transitive clauses in languages across the world (Dryer, 2013). SOV and SVO are the most frequent orderings.

1.1.1. Revealing the putative cognitive underpinnings of language universals

After having identified cross-linguistic generalisations, such as the prevalence of SOV and SVO ordering, by means of typological inquiries, the question arises of whether such a pattern is driven by cognitive biases. There are a few methods at disposal to study potential cognitive biases underlying these linguistic regularities in human adults. First, one can look at whether the same universal patterns arise when learning an artificial language. During such an experiment, participants are taught new words in a training phase and are asked to describe pictures with these new words in a test phase. How these words are spontaneously combined is subsequently measured (e.g. Martin et al., 2024). Another way is asking people to communicate in a different modality, such as by improvised gesturing. In a typical experiment, participants describe a picture or video of an event using improvised gestures (e.g. Goldin-Meadow et al., 2008). The artificial language learning and improvised gesturing approaches have in common that they try to surpass the conventional way of using language, to see what sort of structures are easiest

to produce and understand. Since there is no specific established convention on how to communicate about the depicted event, participants will have to order their utterances on the spot, which provides the opportunity for potential cognitive biases to skew the ordering. To argue for a bias convincingly, people with knowledge of differently ordered languages must all (or mostly) converge towards one ordering – regardless of their native language. Without including a variety of languages, one cannot be sure that a cognitive bias really underlies the observed language universal, since patterns may occur through language knowledge spilling over, either because the same rules as in the spoken language are applied or because that specific language has effectively restructured cognition (e.g. Gleitman & Papafragou, 2005; Majid et al., 2004). The strongest evidence thus comes from speakers of languages that do *not* show the specific tendency, who yet reveal a bias towards such a pattern, thereby overriding their linguistic rules. Also, populations that have invented their own gestural communication systems with little influence from other languages, such as homesigners or signers of recently emerged sign languages, can be studied to see what structures emerge under those circumstances and whether those resemble the cross-linguistic patterns. Homesign is the system of self-invented gestures which are adopted by deaf individuals to communicate with their (hearing) family. Because of the lack of linguistic input, the communication systems that are invented by these children, provide a good model of what linguistic properties (either resembling conventional languages or not) may result from cognitive roots. Moreover, relatively new sign languages, which developed in secluded deaf communities, pose similar opportunities to study which structures get implemented in a successful communication system without systematic input from (spoken) languages. When the signing system of homesigners and recently emerged sign languages show patterns that resemble statistical universals in other languages, this suggests that these features have a special status in the human cognitive system. Lastly, on the comprehension side, neurophysiological measures can be implemented to see how language structures are processed and which may be perceived as surprising. Again, it is best if such measurements are implemented in speakers of a variety of languages.

The methods described above have culminated in a large body of evidence showing that cognitive factors are (at least partly) behind some linguistic universals in a variety of domains, such as consonant clustering (Berent et al., 2008), telicity of verbs (Strickland et al., 2015) and ordering in noun phrases (Martin et al., 2024) to name a few. I will discuss studies dealing with the ordering of subject, object and verb below.

As mentioned previously, expressing simple causal events often occurs using active transitive sentences. Theoretically, six possible orders of subject, object and verb are possible to do so, but the frequent occurrence of SOV and SVO in spoken and sign languages may be driven by underlying biases.

A study on a homesigner showed that he was able to learn to establish causality in his expressions over the years (Rissman & Goldin-Meadow, 2017), showing that this is a dimension that is accessible without formal access to language. Considering the word ordering universal of subject, object and verb, results from a pantomime experiment led Goldin-Meadow and colleagues (2008) to conclude that SOV ordering is prevalent because

it is the product of the natural sequencing in which events are represented mentally: as agent-patient-action. They had demonstrated that speakers of four distinct languages with either SOV or SVO ordering, converge to use SOV when gesturally describing events. In a second part of their experiment, participants also used this SOV ordering when they had to stack cards depicting the relevant event elements in a non-communicate task (Goldin-Meadow et al., 2008), reinforcing the idea that this ordering reflects how the events are represented mentally. This finding to gesture SOV regardless of the native language's order has been replicated for similar simple events (Gibson et al., 2013; Langus & Nespor, 2010) and this order is additionally easiest to comprehend for listeners (Langus & Nespor, 2010). Similarly, the Al-Sayyid Bedouin Sign Language, a young sign language that has arisen about 90 years ago without significant external influence, follows a SOV ordering, unlike surrounding languages (Sandler et al., 2005). So, some evidence from different sources point to SOV as the ordering reflecting a cognitive representation.

On the other hand, the tendency to use SOV may be mitigated under specific circumstances. For example, when participants of a gesturing experiment were presented with more complex scenes including one event embedded in another, they tended to gesture main clause before the subordinate clauses, thus following SVO (Langus & Nespor, 2010), indicating that complexity and embeddedness also influences ordering. An important point to consider is that, for most spontaneous gesturing studies, events with humans acting on objects were portrayed. Four studies showed that this tendency towards SOV may be limited to such specific configurations in events and may switch to SVO depending on semantic and functional considerations. For example, Gibson *et al.* (2013) proposed similarly a tendency to follow SOV, which may switch to SVO in events that are considered reversible, to avoid potential confusion about 'who is doing what'. In their experiment, SVO was produced when both the agent and patient were human (reversible), whereas SOV was the order participants chose to gesture when a human agent acted on a non-human patient. Hall *et al.* (2013) showed similar results, but posited that this is to avoid role conflicts from the producer's side. Another experiment also found SOV for events with an animate agent acting on an inanimate patient, but this was only the case for intensional events, which have a concrete object, and turned to SVO for extensional events (e.g. "thinking of"), indicating that fine-grained semantic factors can influence constituent ordering (Schouwstra & de Swart, 2014). However, a recent study proposes a different explanation and demonstrates that the order preferences are related to the iconicity of the event, such that the place of the object is dependent on whether the object is visible before (leading to OV) or after the action (leading to VO) (Schlenker et al., 2024). This visibility generalisation theory focuses thus on a more direct relation between what is seen and how it is described.

While SOV and SVO are both frequent, one universally present feature remains throughout: the fact the subject is placed before object. Focussing on this feature could therefore be a more fruitful avenue in trying to uncover the precise ways in which cognitive functioning may underlie universals. However, from the pantomime studies described above, *none* featured speakers of a language that placed the object before the subject. As explained previously, this poses a serious concern when arguing for a cognitive

bias. Future studies should thus try to probe what would be spontaneously produced by speakers of a language that has this infrequent OS ordering (for instance, Malagasy may be an accessible option for such a study). The free gesturing studies mentioned here are therefore only indicative of a cognitive bias (in the sense that O is rarely placed before S), but influences of the gesturers' spoken language cannot be excluded. In this respect, it is worth mentioning again that the Al-Sayyid Bedouin Sign Language generally places subject before object (Sandler et al., 2005). The same is true for two other young sing languages: Israeli Sign Language and Kafr Qasem Sign Language (Meir et al., 2017). Likewise, a study on three homesigners (without any formal education and knowledge of written language) demonstrated that they mark subject, and place this first in their gestured sentences (Coppola & Newport, 2005). The animate-inanimate divide may be an important determiner for word ordering; in gestured sentences of signers of young sign languages and of gesturing speakers of three different languages, the likelihood of producing object before subject increased when the object was a human or animate (Meir et al., 2017), but, overall, SO remained most frequent. Even though the visibility generalisation theory proposed by Schlenker *et al.* (2024) for the ordering of OV and VO does not explicitly address the preference for the positioning of the subject, the same line of reasoning can be applied: the order of the arguments may reflect the temporal order in which the event plays out and is perceived, and since the subject is often visible before the object (temporally), this iconicity would lead to a tendency to place S before O.

On the comprehension side, it has been theorised that people tend to interpret the first element of a linearly ordered sentence as the agent as a disambiguation heuristic, and that this is one of the features of a proto-language that is still visible in present-day languages (Jackendoff, 1999). Support for this idea comes from how people seem to understand gestured events and construe that the firstly gestured constituent must be the agent, following an agent-first principle (Hall et al., 2015). It should be noted that those participants were all speakers of languages that placed the subject first to begin with (SOV or SVO). However, it is still telling that young children tend to overgeneralise the agent-first rule, even when it is not appropriate. For example, children interpret the first noun phrase as an agent even if passive voice is used, such as in "the dog is chased by the cat" (Abbot-Smith et al., 2017), or when the sentence describes a two person intransitive event, such as "the dog and the cat are running" (Gertner & Fisher, 2012). Even more convincing evidence that this processing heuristic is used, comes from another study that used physiological measurements and showed that speakers of Hindi, who do not necessarily place the agent as subject first (because of subject-drop and ergative marking), are nonetheless surprised when the noun phrase placed first turns out *not* to be the agent (Bickel et al., 2015).

From these studies, it becomes apparent that, in cases in which conventional languages cannot be used (i.e. for homesigners and in elicited pantomime experiments) or when a conventionalised system has emerged only recently (i.e. for young sign languages), the subject is placed before the object, but there is a non-negligible effect of animacy that plays a role. Tentative evidence coming from listeners seems to confirm that the first element in a sentence is often construed as the agent. These findings thus attest that

certain grammatical categories and structures, like placing subject before object, may exist outside of typical language and align patterns earmarked as universal. These experiments appear to reveal probabilistic biases driven by people's common, universal mental representations. In other words, the cognitive underpinnings of SO order that come to light, show that people probably perceive and represent simple events in the same way. For example, it has been argued that word order results from the temporal order in which the event plays out visually (Schlenker et al., 2024) and that the linguistic subject-first tendency that is found in languages is instantiated because it matches the sequential ordering of how events are processed in the brain (Kemmerer, 2012). What do we know about how events are mentally represented?

1.2. Event representations

Granting that there could be some universal representations culminating in cognitive biases driving observed linguistic patterns, it is a valid question to ask where those representations may come from. A prominent and relevant proposal is that these representations are simply derived from how humans naturally perceive and process the world around them.

It is well-established that there is a link between the way events are represented and described. For example, theories about language production are often built around the idea that a conceptualisation phase precedes speech (Levelt, 1993). According to this hypothesis, during conceptualisation, the message is a package of information which represents the speaker's communicative intent, which will get transformed into linguistic form. This does not necessarily mean that the linguistic representation of the message before uttering a sentence must have the same form as the representations for understanding or remembering an event (Clark, 2003). Language may influence cognition as remarkable cross-linguistic differences can appear during event processing, in particular when the task involves using linguistic knowledge, such as when describing a scene (e.g. Gleitman & Papafragou, 2005; Ünal & Papafragou, 2016). So, the format of the representation may depend on whether the task at hand draws upon linguistic knowledge, surfacing as crosslinguistic differences, or not (Malt et al., 2003; Slobin, 1987). Another way to look at it, is that communication with language is an externalised and visible form of the underlying grammar-governed representational system (Sperber & Wilson, 1986). These theories converge on the idea that there are mental representations that could be shared across people, which in turn are transformed into language.

In the domain of events, there are striking similarities between conceptual event structure and linguistic structure (e.g. Papafragou & Grigoroglou, 2019). There are indications that the surface realisation in language reflects the conceptual structure of events, in particular with regards to delineating event boundaries and the identification of event participants (Ünal, Ji, et al., 2021).

First, in 1.2.1 it will be discussed how events are represented in human adults. However, since knowledge of a specific language sometimes influences the way events are perceived, leading to cross-linguistic variation in conceptualisation, this can form a chicken-and-egg situation. To truly identify the structure of event cognition in the absence of language, one can turn to preverbal infants. 1.2.2 will therefore explore what is known about how events are represented in human infants, before they master language. Lastly, 1.2.3 will go through the evidence about event representations in non-humans, mostly in non-human primates. This will simultaneously give insights in the evolutionary roots of the cognitive functions studied.

1.2.1. Event cognition in human adults

Instead of a tumultuous incoming flux of sensory information, people experience, remember and talk about their surroundings as structured. *Events* therefore play a central role in how humans perceive their environment. *Event cognition* encompasses all

cognitive mechanisms that are responsible for transforming the incoming information into events, and is composed of event perception in particular, which deals with how events are perceived and represented, and its interplay with other capacities such as memory and language (Zacks, 2020).

Defining an event is not straightforward (see: Yates et al., 2023). The definition for two actor events that I use here, is thus composed of several sources. An event can be defined as follows: an event A) has a spatiotemporal dimension, such that it has a location and a beginning and an end, B) has an internal structure, such that it is composed of several subelements, and C) expresses a causal relation, such that a change takes place. Here, I will discuss these fundamental properties of events in more detail.

1.2.1.1. Spatiotemporal dimension of events

An event is considered a unit with a spatiotemporal dimension: happening at a location and with a duration. Such a unit is the result from a dissection mechanism in which the event is singled out from a continuous stream of incoming information (Baldwin & Kosie, 2021; Kuperberg, 2021; Zacks & Swallow, 2007; Zacks & Tversky, 2001). People are fairly consistent in judging where the boundaries between events lie (Magliano et al., 2014; Newtson & Engquist, 1976; Zacks et al., 2007). This can either be tested by explicitly instructing participants to find the event boundaries (e.g. Newtson & Engquist, 1976), or by the means of more implicit measures such as using neuroimaging (e.g. Zacks et al., 2001) or the dwell time paradigm (Hard et al., 2011). In the dwell-time paradigm people skim through slides of events and are found to dwell longer at event boundaries. Across these different measurements there is a concurrence in people's judgements of event boundaries, suggesting that this concerns a general and automatic cognitive mechanism (Zacks & Swallow, 2007).

Segmentation of the stream occurs according to a perceived beginning and end of the event. This can happen to a finer or coarser scale, leading to a hierarchical organisation of events where coarser events can be made up from several smaller-grained events (Zacks et al., 2001; Zacks & Tversky, 2001). For example, the action sequence of baking bread would be considered to be a coarse-grained event, whereas a part of that, such as folding the dough, would be considered as a fine-grained event.

Judging where an event boundary lies, has been proposed to result from prediction error in the Event Segmentation Theory (EST; Radvansky & Zacks, 2011; Zacks et al., 2007). According to this theory, people make predictions about what will happen next, and if this prediction does not match the subsequent situation well enough, it is experienced as a break (Reynolds et al., 2007; Zacks et al., 2007). Such a break then constitutes the boundary between one event and the next. A prediction error especially occurs when a high level of visual change (Reynolds et al., 2007; Zacks et al., 2007) or content change (Swallow et al., 2018) is registered, which results in the perceived event boundary. However, according to the information-optimisation account, surprisal is not sufficient to explain event segmentation (Baldwin & Kosie, 2021). Looking times during the dwell-paradigm were found to actually increase at event boundaries in people that had prior knowledge about the statistical regularities of the event sequences (Hard et al.,

2019), rather than decrease because they were less surprised. Thus, people may actually shift attention towards areas of expected low predictability leading to a more efficient processing of the event structures (Baldwin & Kosie, 2021).

All in all, segmentation of incoming information into individuated events is an important aspect of how humans perceive their surroundings and represent them mentally.

1.2.1.2. Internal structure of events

Once an event is segmented from the information stream, it is not considered to be represented in just one piece, but rather an assemblage of the different objects and actors (Altmann & Ekves, 2019; Radvansky & Zacks, 2011).

Possibly this ensemble is made up of separate objects which change state over the time course of the event (Altmann & Ekves, 2019) or, more commonly believed, of entities that are defined based on the role they play in the event (Rissman & Majid, 2019). There have been a variety of event roles identified, of which the agent and patient are most studied (Rissman & Majid, 2019). A simple event can consist of just these two event roles, and more extended events also include other roles such as source, goal and instrument (Rissman & Majid, 2019; Ünal et al., 2024). The focus will lie on the agent and patient.

How are agent and patient defined? The *agent* of an event is the doer of the action, which can often be recognised by its volitional behaviour, awareness and motility, and can exist on its own (Dowty, 1991). The *patient* is the undergoer of the action, whose behaviour is dependent on the agent's behaviour, it goes through a change and is less dynamic compared to the agent (Dowty, 1991). It follows that a patient only exists relative to the agent, making them part of a relation (see below).

To avoid any terminological misunderstandings, both event roles can show properties of agency, although the agent probably does so to a greater extent than the patient. Agency means having the capacity to become the causal source of events (Leslie, 1995) and thus has to do with the ability to exert power, which is often related to perceived goal-directedness of the actions. Then, the notion of animacy is related to agency, but not the same. Animacy is delineated by whether an entity is alive or not, which is mostly equated to whether the entity can move by itself. In most typical cases, the agent is both animate and agentive.

There is now ample of evidence that people extract agent and patient roles from events. For example, when presented with very brief (37 or 73 ms) pictures of a two-participant event, human adults were successful to answer yes/no questions about who was the agent and who the patient (Hafri et al., 2013). This even occurred unconsciously; when the participants are engaged in an irrelevant task of deciding on which side of the screen the target actor wearing a specific coloured t-shirt is present, they were faster to respond when the agent and patient roles stayed the same as in the previous trial (Hafri et al., 2018). Similarly, participants' responses to name an action were facilitated after being primed with a briefly presented picture of the same action involving different entities followed by a mask (Zwitserlood et al., 2018). Since the specific identity of the entities

involved in the action does not matter for such facilitating effects to emerge, these findings suggest that agent patient roles are represented abstractly.

The agent and patient roles do not appear to stand on equal ground though; agents are recognised to attract people's attention more than patients. While watching events, human adults first direct their gaze toward the agent of an action (Webb et al., 2010). Similarly, in cartoons, agents are looked at longer than patients (Cohn & Paczynski, 2013). Furthermore, adults are faster to orient to agents when asked to detect agents than to orient to patients when asked to detect patients (Wilson et al., 2011). From short displays of drawings, agents are systematically better identified than patients (Dobel et al., 2007). However, a recent study that used a change detection task, found the opposite: that patients were detected faster than agents, but the authors attribute this finding to the fact that they applied the change to an attribute of the agent rather than the whole entity (as they did for the other event elements) (Ünal et al., 2024).

What cues are used to identify these roles? Attributes such as agency and animacy may function as heuristics to identify roles in events involving an animate acting on an inanimate. Since animacy attracts attention (e.g. New et al., 2007), it cannot easily be disentangled whether the attention was attracted because of the agent role or simply because the entity was animate. Webb *et al.* tried to circumvent this issue by using stimuli in which a human acted on a box, while having a second human and a second box present in the same video, and nonetheless participants looked first towards the agent and not towards the second human, either because the agent performed the agent role or simply because they were moving. To avoid such simple explanations based on animacy only, it is thus relevant to look at events including a similar looking agent and patient, e.g. both animate or both inanimate. For example, the studies by Hafri and colleagues (2013, 2018) which included pictures portraying two humans, point to posture as bearing cues of who is agent and who patient. In their stimuli the agent had a more active posture, with outstanding limbs, and the patient had a more passive posture. This is corroborated by cartoons in which humanlike figures are identified as agent and patient (Cohn & Paczynski, 2013), again based on posture. By using artificial stimuli, it has also become clear that motion by itself can identify the agent role. For example, in 2D chasing stimuli where the two objects look exactly the same, the chaser is identified faster than the chasee (Meyerhoff et al., 2014b); indicating that the motion pattern of these objects alone is enough to derive who is performing which role.

For human adults, there is thus ample evidence that they do not process events holistically, but rather pick up on the important event elements, such as the agent and patient. Based on a wide variety of stimuli, it can be concluded that human adults represent agent and patient in an abstract sense with a particular prominence for the agent.

1.2.1.3. Events as causal relation

An event is not just an assemblage of separate elements, rather they are bound together causally (Altmann & Ekves, 2019; Radvansky et al., 2014; Radvansky & Zacks, 2011). Simple two participants events can be schematically described as an agent generating a

force vector which acts on the patient and changes the patient as a result vector (Gärdenfors, 2024). Indeed, one of the defining characteristics of the agent role is that it causes an event or state change in the other participant and of the patient role that it is causally affected by the other participant (Dowty, 1991).

There are many indications that humans extract causality from ongoing events. For example, people readily interpret scenes as causal and falsely remember causal contact (Kominsky et al., 2021; Strickland & Keil, 2011). Even to simple shapes that are moving contingently, people apply causal interpretations of their behaviour (Heider & Simmel, 1944). Similarly, in schematic 2D stimuli in which an object A moves towards a stationary object B, and B starts moving as soon as A stops and touches B, people see that A *caused* B to move (Michotte, 1963; for review: Scholl & Tremoulet, 2000), see Figure 2. However, an impression of causality is not evoked for highly similar events, where a temporal gap and/or spatial gap is inserted between the movements of object A and B, see Figure 2. For these specific stimuli, it appears that the causal impression upon seeing such a launching event is a perceptual phenomenon (Kominsky & Scholl, 2020; Moors et al., 2017; Rolfs et al., 2013). Importantly, recent findings point to the idea that many relations are perceived rather than cognitively deduced (Hafri & Firestone, 2021).

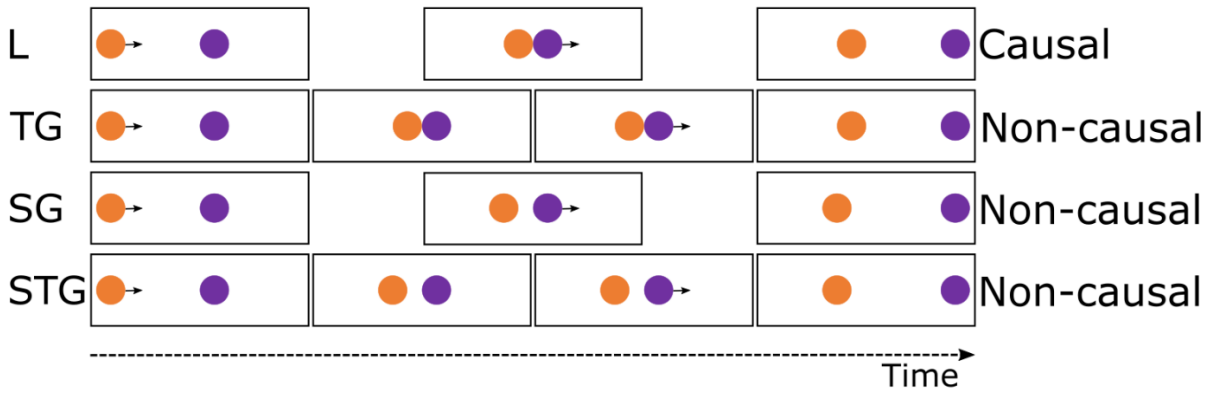


Figure 2: Four Michottean events. The launching event (L) is perceived as causal by human adults. The other similar events where a temporal gap (TG), spatial gap (SG) or both (STG) is added between the movements of the two objects, do not result in a causal impression.

In line with this, human adults appear to be specifically attuned to recognise social human-human interactions, for example, by using cues of body postures. Papeo and colleagues showed in a series of experiments that humans are particularly sensitive to configurations of two bodies oriented towards one another (Papeo, 2020). They demonstrated this by making use of the inversion effect, which holds that upside-down images of, typically, faces are processed slower (i.e. a high cost of inversion) than upside-down images of other familiar objects, because of visual sensitivity finetuned to the perception of faces. Importantly, the inversion effect is not restricted to faces and also emerges for dyads that are facing one another compared to those who have the bodies

turned away from one another (Papeo et al., 2017), regardless of visibility of the face (Papeo & Abassi, 2019).

Human adults thus quickly pick up on relations and are particularly apt to attribute causality to them.

1.2.2. Event cognition in preverbal infants

There is evidence that infants are sensitive to some of the key components of events – well before having the capacity to express them with full-fledged language. Many theories of language development are built around the idea that non-linguistic concepts are in place before word learning begins (e.g. Clark, 2004; Gleitman, 1990). It has also been posited that such representations may be universal in young infants, and can later be altered through knowledge of a specific language (Hespos & Spelke, 2004). It is thus relevant to study preverbal infants to establish whether event representations are the result of language or can exist before full command of language.

1.2.2.1. Spatiotemporal dimension of events

There are now indications that young infants can dissect events from an incoming stream of information (Levine et al., 2019). For example, 10 month-old infants look longer at events that have pauses implemented in the middle of the event than at the endpoints (Baldwin et al., 2001), suggesting that they are sensitive to natural breakpoints between events. A subsequent study confirmed this by using a different design. When presented with two side-by-side distinct videos displaying an actor performing an action sequence, infants of 9 to 11 months old looked towards the video which corresponded to a series of tones which marked the event endings that was presented simultaneously (Saylor et al., 2007).

In another study, 6-to-8-month-old infants who were habituated to sequences of events, successfully detected which separate events they had previously seen in the sequence and which were completely novel, indicating that they were capable of extracting events from the sequence (Hespos et al., 2009). However, when the authors applied the same method on the transitions between the events in the sequence, the infants did not differentiate between novel and familiar transitions. In a later experiment, the authors showed that babies of this same age group looked significantly longer to sequences that contained a novel event with a familiar transition than a familiar event with a novel transition, which confirmed that event content is more salient than event transitions (Hespos et al., 2010). Possibly more experience is necessary to focus on the event breakpoints. Infants seem to employ statistical learning to determine where event boundaries lie. For example, infants of 7 to 9 months old who were familiarised with a series of actions of which some always followed each other, looked longer at test at units which crossed those boundaries than intact units (Stahl et al., 2014). That experience may be necessary to finetune the parsing of events, is confirmed by a study that revealed neural differences in how adults and infants perceive event streams (Yates et al., 2022). Using fMRI data, in which events could be measured as stable activity patterns that shifts

at event boundaries, infants appear to detect events from a continuous stream like adults, but they do so at different breakpoints and at a coarser scale.

The current picture thus points to infants having more interest in event content than event boundaries, and to a developmental trajectory where the detection of event transitions changes over time, presumably dependent on experience.

1.2.2.2. Internal structure of events

There is a large body of research studying how infants perceive event roles in the broad sense (e.g. Wagner & Lakusta, 2009). Considering the agent and patient specifically, few studies have looked at the discrimination of the two roles. A seminal paper by Leslie & Keeble (1987) suggested that infants of 6-month-old may represent agent-patient roles. The authors presented infants with Michottean launching events: schematic 2D animations depicting an object A seemingly launching an object B, see Figure 3 top. To test whether the infants perceived the launching events as having an internal structure, they first habituated infants to launching or a similar event with a temporal gap (perceived as non-causal by human adults) and presented the infants with a directionally reversed version at test where B moved first, followed by A. They demonstrated that infants were more surprised to see a reversal of the launching event than a reversal of the temporal gap event, and proposed that this effect was due to a role reversal that had taken place in which the launcher had become the launchee, see Figure 3 bottom. Since an agent-patient structure is present in the launching event only (launcher and launchee) and not in the temporal gap event (two separately moving objects), this finding is often taken to indicate that 6-month-old infants represent agent and patient. However, subsequent studies failed to replicate this effect with different materials (Bélanger & Desrochers, 2001; Kominsky & Carey, 2024). Another problem concerns the lack of a control condition which takes into account that infants may have differential responses to launching and temporal gap events in general (as they have at birth (Mascalzoni et al., 2013)), potentially generating differences at test that do not necessary have to do with the role reversal. These points made the results and the following impact of the study arguable, which will be discussed in detail in **Article 2.**

Another study looking at preverbal infants' ability to delineate agent and patient, showed that 5-month-olds have an attentional bias towards the agent in a 2D chasing interaction (Galazka & Nyström, 2016), similar to human adults (Meyerhoff et al., 2014b). Somewhat later, at 9 months old, this looking behaviour at the chaser evokes neural responses related to social perception (Galazka et al., 2016), lending support to the idea that the attentional bias is related to the agentive characteristics of the chaser and in general that such moving shapes are really perceived as a relation. This bias towards the agent may facilitate word learning too as 14-month-olds learn more easily a label for a chaser than for a chasee (Yin & Csibra, 2015).

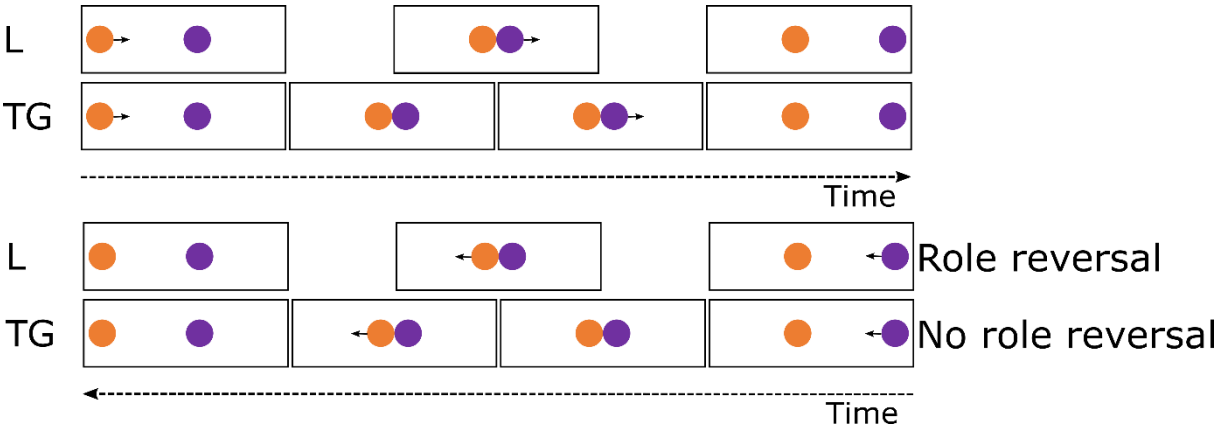


Figure 3: Leslie and Keeble (1987)'s role reversal paradigm on Michottean events. Top: Habituation events are a launching event (L) and a temporal gap (TG) event displayed from left to right where object A (here depicted as orange) moves first and then object B (here: purple). Bottom: At test, the direction is reversed from right to left and now object B (purple) moves first, followed by object A (orange). Since only the L has an agent-patient structure, the reversal of L leads to a role reversal.

However, Leslie and Keeble's (and further replication attempts) and Galazka *et al.*'s studies did not examine infants' ability to generalise the agent-patient discrimination across different entities or relations. Until recently, it had not been tested whether they constitute abstract representations rather than being restricted to representations of a launcher and a launchee or a chaser and a chasee. Importantly, a recent study showed that infants as young as 7 months old, may have abstract representations of agent and patient roles based on their posture and positioning (Papeo et al., 2024). After seeing a variety of still images in which a human agent acts on a human patient while they are facing each other, infants are surprised, as evidenced by dishabituation and pupil dilation, upon seeing a display in which the two roles have been switched around such that the previous patient has become the agent and the previous agent the patient. Importantly, this effect was not observed in a crucial control condition in which the same postures were used but back-to-back, indicating that the postural cues were only indicative of event roles if they were interacting. Additionally, after a new stream of different agent-patient scenes where the two faced each other, pupil dilation was observed when the agent flipped its position facing away from the patient, but not the patient, showing that the infants generalise from posture to positioning to assign the agent and patient roles.

So, it is recognised that infants may be sensitive to the agents and patients in chasing events, although for launching events the grounds are less stable. It has long been hypothesised that preverbal infants represent event roles mentally, and recent evidence suggests they do so abstractly.

1.2.2.3. Events as causal relation

Many view that perception of causality from simple Michottean events (see Figure 2) may form the basis of causal understanding (Michotte, 1963; Muentener & Bonawitz, 2017; but see: Saxe & Carey, 2006). For this reason, there is a large number of studies on

how infants process launching events and whether they also see them as portraying a causal relation. Just after being born, babies preferentially look towards the continuous movements of direct launches compared to events with a temporal delay (Mascalzoni et al., 2013). These discrimination abilities may have to do with causality, but a more parsimonious explanation is that they are based on the specific spatiotemporal features that differentiate between causal and non-causal events (i.e. the presence or absence of the gaps, see Figure 2). Evidence that infants may rely on causality for categorial differentiation, rather than spatiotemporal features, appears later in life. Habituation-dishabituation studies have shown that six-month-old infants, but not younger ones, group different non-causal events (with a temporal gap and with a spatial gap) together (Cohen & Amsel, 1998; Oakes, 1994). This suggests a developmental progression where younger infants rely on a feature-based strategy, using spatiotemporal cues (such as the presence or absence of a spatial or temporal gap), to differentiate Michottean events and from around six months of age, this shifts into a category-based strategy. These categories can be explained as representing continuous vs. non-continuous movement, but they can also be interpreted as hinting at causal vs. non-causal categories. Further confirmation that such events are truly perceived in terms of causality depends on demonstrating that the causal roles of agent and patient are distinguished, or that subsequent inferences are made about the entities based on their causal role (Kominsky & Carey, 2024). For example, 8-month-old infants expect to see the two objects make contact during a launching event occurring behind an occluder. When the occluder is removed, they are surprised to see a gap between the objects (Adibpour & Hochmann, 2023). When it comes to forming expectations that could point to the extraction of causality, only somewhat older infants seem to be able to do so. For example, 8-month-olds do not readily apply event roles to boxes covered in fur (a surface feature indicative of the entity being an animate agent), engaged in a launching event; they did not look longer when the launcher box was revealed to be hollow, which would be unexpected if they had attributed the agent role to it. On the other hand, 14-month-olds are surprised to see the patient of a launching event engaged in self-propelled movement, but not the agent (Cicchino et al., 2011), indicating that they have inferred likely behaviour based on the roles in the events. In a similar vein, when 10- and 12-month-old infants see objects being thrown but the source is obscured, they infer that a human hand, and not an inanimate object must be the causal agent of this action (Saxe et al., 2005). These studies go thus beyond the discrimination of launching and gap events, and are therefore more convincing at establishing a grasp of causality, but show that the ability to make inferences may arrive later in life, possibly after 10 months old.

Sensitivity to information cueing an interaction between humans may aid infants to extract a causal relation. In that sense it is relevant that 6-month-old infants discriminate facing from non-facing dyads, as they look longer to the latter (Goupil et al., 2022). Since the infants also differentiated between a non-facing dyad and a non-reciprocated dyad, in which one person was facing the back of the other, but not between the non-reciprocated and facing dyad, this suggests that the infants perceive a relationship when one person is facing the other. The preference for looking at non-facing dyads contrasts with human

adults, and a later study revealed that this preference slowly turns towards facing dyads over the first years of life (Goupil et al., 2024).

So, preverbal infants are attuned to circumstances that cue a causal interaction, such as the positioning of two actors and the spatiotemporal properties tying them together, but it takes time to develop the capacity to extract causality.

1.2.3. Event cognition in non-human primates

Human's linguistic abilities are built from the building blocks inherited from our primate ancestors, and one of those may be in the form of shared event representations. This has been argued for mechanical relations (Hespos & Spelke, 2004), but recently the broader view of event cognition has also gained attention. Zuberbühler (2019, 2020, 2022) proposed that there may be a shared way of perceiving events in humans and non-human primates, which became externalised in the human communication system (i.e. language) as one of the theories explaining the emergence of syntax and grammar rules in language. Elaborating on this, it has been argued that non-human primates show signs of understanding event relations similarly to humans, but do not express them in their communication, revealing a dichotomy between their mental representations and production of communicative signals (Zuberbühler & Bickel, 2022). Next, Wilson *et al.* (2022) have reviewed a large body of literature and have concluded that event decomposition with a specific focus on agency is probably not unique to humans. They focus on four cognitive components of event decomposition, physical causality, social representation, psychological causality and agency attributions, and conclude that there is some evidence that non-human primates function similarly to humans in the first two categories, but less information is available for the latter two categories. Below, I will use the same framework for event representations as I have done for human adults and human infants, and review what is known for non-human primates.

1.2.3.1. Spatiotemporal dimension of events

To my knowledge, there are no studies with non-human animals about if and how events are singled out from a continuous stream of incoming information. It remains thus an open question whether boundaries are perceived at similar timings in non-human primates and in humans. Adaptations to the dwell time paradigm or implementations of looking time paradigms that have been used on preverbal infants, could be potentially relevant ways to tackle this topic in non-linguistic species. I would like to pursue such questions further in future projects.

1.2.3.2. Internal structure of events

Do non-humans extract relevant event elements, such as agent and patient from events? One line of work has revealed that baboons (*Papio hamadryas ursinus*) and chimpanzees (*Pan troglodytes*) detect incongruent interactions from vocalisation playbacks in which a lower-ranked individual threatens a higher-ranking one, indicating that they discriminate the aggressor and the victim, and thus extract who does what to whom in these events (resp. Bergman et al., 2003; Slocombe et al., 2010). A similar effect

was found in the visual modality as male long-tailed macaques (*Macaca fascicularis*) looked longer at pictures displaying an anomalous interaction including threatening facial expressions from a lower-ranking towards a higher-ranking individual (Overduin-de Vries et al., 2016). However, since these inferences are based on knowledge about the social relationships in the community and the responses specific to the recognised individuals in the playback or picture, they cannot reveal whether the representations also exist in the abstract form of agent and patient.

A recent comparative study between great apes (chimpanzees (*Pan troglodytes)*, gorillas (*Gorilla gorilla*) and orangutans (*Pongo abelii*)) and humans (adults and children) tried to address agent and patient representations directly by determining whether an agent bias exists across species, but yielded highly variable results which could largely be explained by effects of animacy (Brocard et al., 2024). The participants were presented with naturalistic videos of agent-patient interactions of conspecifics or of the other tested species, and they had to touch the screen when the video froze, after which it was measured whether they had touched the agent or the patient. Only in the animate-acting-on-inanimate condition, the agent was reliably chosen over the patient across species, which can be explained by a preference for animates (or by the training in which the participants were taught to press an animate actor). The idea that an agent bias contributed to this effect is further undermined by the lack of a consistent agent bias even among human adults. This raises doubts about whether the task is suitable for probing an agent bias. A subsequent eye-tracking study using similar naturalistic videos found that human adults, infants, chimpanzees, gorillas, and orangutans did not show a higher proportion of looking time toward agents throughout the entire video (Wilson et al., 2023). However, that study did not specifically test for an initial attentional preference or agent bias at the beginning of the video. To dive into this question from a different angle, we tried to test an attentional bias towards agents based on highly controlled stimuli of chasing interactions, which have been effective in probing an agent bias in human adults and preverbal infants, which is discussed in Article 3.

So far, there has not been much research into the question of whether non-human primates represent agent and patient in an abstract sense. Recent attempts to probe these capacities have not painted a clear picture yet. Demonstrating the existence of such representations remains thus essential to corroborate the theory that event representations are similar across species and can underly humans' linguistic abilities.

1.2.3.3. Events as causal relation

Many studies have looked at the abilities of non-human animals to reason causally (Penn & Povinelli, 2007; Schloegl & Fischer, 2017; Völter & Call, 2017). For example, in the Primate Cognition Test Battery, there are tests included tapping into understanding of causality in the physical domain considering a produced sound, an appearance change, and using tools (Herrmann et al., 2007) and many primate species have been shown to successfully solve these sort of tasks (e.g. Fichtel et al., 2020; Herrmann et al., 2007, 2010; Joly et al., 2017; Schmitt et al., 2012). However, some specific types of causal relations may be treated even at the perceptual level (e.g. Hafri & Firestone, 2021), and whether non-

human primates also *perceive* causal relations from ongoing events has received less attention.

One area of interest has been to monitor animals' responses to violations of physical causality. From such studies it has become clear that young Eurasian jays (*Garrulus glandarius*) and chimpanzees look longer at videos where an object moves without contact (resp. Davidson et al., 2017; O'Connell & Dunbar, 2005), although it is hard to determine whether this looking behaviour is because of the violation of causal structure or just because of the unusuality of the events.

In the realm of physical causality, many experiments also made use of 2D Michottean events, see Figure 2. Indeed, several species discriminate launching from other events, which are perceived as non-causal by humans, such as events with a spatial or temporal gap. For example, newly-hatched chicks (*Gallus gallus*) have an innate predisposition to orient towards the launcher in causal direct launch, but such a preference does not exist for a non-causal temporal gap event (Mascalzoni et al., 2010). Also, pet dogs (*Canis familiaris*) look longer and show more pupil dilation to a non-causal event with a spatial gap compared to a causal event (Völter & Huber, 2021). On the other hand, pigeons (*Columba livia*) had difficulties to learn a causal distinction and instead focussed on low-level features (Young et al., 2006). While these studies show that certain species can differentiate between various Michottean events, they do not clarify the basis of this distinction – whether it is truly driven by an understanding of causality or merely by low-level perceptual features. We designed a task that can discriminate between these two strategies and tested whether Guinea baboons (*Papio papio*) perceive causality in Michottean launches in Article 1. Chimpanzees, however, may be able to distil causality from Michottean events, as they have been shown to be sensitive to the causal capture effect (Matsuno & Tomonaga, 2017).

Whether non-human primates are sensitive to some cues indicative of a potential social relation has been studied as well. Using more realistic depictions of relations, rhesus macaques discriminate between facing and non-facing dyads, by looking longer at the facing dyads (Goupil et al., 2024). This shows that rhesus macaques are sensitive to a configuration that is a prerequisite for most social interactions, but does not necessarily entail that they extract a causal relation from this.

From these studies it thus seems that, even though many non-human primates perform well on tasks tapping into causal reasoning in the physical domain, evidence regarding whether they also perceptually treat simple causal relations is currently not sufficient.

1.3. Outline of this thesis

This thesis describes three studies regarding two principles related to the event cognition and language interface: causality and agent and patient roles. As pointed out in the previous sections, this thesis tackles three gaps in the existing literature. First, in Article 1, it will be asked whether Guinea baboons perceive causality in Michottean events, since these events are frequently used for human adults and human infants, and even some non-human species, but a common pitfall regarding these studies is that discrimination of launching from similar events with a temporal or spatial gap is equated to sensitivity to causality. A novel task which is designed to decide whether causality or only spatiotemporal properties are used for classification of Michottean events is implemented for Guinea baboons and human adults. Second, in Article 2, we ask whether human infants, Guinea baboons and human adults extract agent and patient roles from Michottean launches by using the role reversal method as developed by Leslie and Keeble (1987). Because the original study with preverbal infants has had a major effect on the field, but lacked a critical control condition, we aimed to replicate it. Lastly, to further probe the ability of Guinea baboons to assign agent and patient roles, in Article 3, a novel paradigm was implemented that used controlled stimuli of 2D chasing interactions to test whether Guinea baboons have an attentional preference towards agents.

1.3.1. Participants

This thesis probes the linguistically relevant cognitive capacities that may exist in the absence of language, following the comparative approach in three populations: non-human primates (specifically: Guinea baboons (*Papio papio*)), human infants (specifically: 6-month-old French infants), and human adults. These participant groups are non-, pre- and linguistic respectively and are chosen because they will provide the opportunity to make a comparison regarding the cognitive capacities of these groups while simultaneously controlling for the influence of language.

1.3.1.1. Guinea baboons

Guinea baboons are part of the genus *Papio* in the family of Cercopithecidae (Old World monkeys) within the primate order. They are a diurnal and terrestrial animal, endemic to the most western part of West Africa, whose diet is opportunistic but consists mostly of plants (Fischer et al., 2017, 2019). The social organisation can be described as nested multi-level with female-biased dispersal and is considered to be tolerant and largely affiliative (Fischer et al., 2017). To communicate, Guinea baboons use a variety of vocal signals (Kemp et al., 2017; Maciej et al., 2013) and gestures (Dal Pesco & Fischer, 2018). Guinea baboons share a common ancestor with humans who lived around 25 Mya (Springer et al., 2012).

Because of similarities to (early) humans, baboons have long been recognised as an informative model to study human evolution (Fagot et al., 2019; Fischer et al., 2019; Fischer & Zinner, 2020; Jolly, 1970; King, 2022; Swedell & Plummer, 2012). These similarities include that baboons and humans are both primates with a relatively large

body size, that they are diurnal and terrestrial, that they have an omnivorous diet and, most importantly, that they live in complex social systems. The resemblances may be due to the fact that, besides having a recent common ancestor, these two species faced similar ecological pressures as they have evolved in very comparable environments of the African savanna (Fischer et al., 2019). The more recent diversification of *Papio* species around 1.5 Mya in sub-Saharan Africa (Rogers et al., 2019; Zinner et al., 2009), mirrors that of the *Homo* genus, of which today only *Homo sapiens* exist (Rogers et al., 2019).

Guinea baboons have also been identified as a good model species for the topic of language evolution (Fagot et al., 2019). This is due to their phylogenetic closeness to humans and the fact that a lot of data about baboons is available. A focus to study domain general cognitive functions in baboons that are implicated in language in humans, may elucidate our understanding of language evolution (Fagot et al., 2017). From a cladistic perspective, comparing humans to baboons can provide insights about the common ancestor of humans and baboons who lived 25-30 My ago, which could have implications for elucidating evolutionary trajectories within the Old World monkeys part of the primate lineage (e.g. Byrne, 2000).



Figure 4: Left, the baboon enclosure at the Station de Primatologie. The ALDM's are located in the bungalows and are freely accessible for the baboons to participate in the experiments. Right: A Guinea baboon participating in a touch screen task.

The 23 Guinea baboons (16 females) that have participated in the studies that make up the main part of this thesis, are captive individuals living in a facility called the Station de Primatologie in Rousset, in the south of France. They are housed in two groups of eighteen and five individuals. The enclosure has a large outdoor area (700 m^2) which is enriched with big stones as climbing structures, and a small indoor area which is accessible for shelter, see Figure 4. The baboons are not water- or food-deprived: they always have access to water and are fed daily (monkey chow, fruits, vegetables). The baboons have *ad libitum* access to testing booths with touch screens, which is called the automatic learning device for monkeys (ALDM) developed by Fagot & Paleressompoulle (2009), see Figure 4. The baboons participated voluntarily in computerised tasks employing an operant conditioning method. This testing method, in which the baboons can come and go as they

please, reduces stress levels compared to periods during which the ALDMs are not accessible (Fagot et al., 2014).

1.3.1.2. Human infants

6-month-old infants are considered preverbal, since the production of the first word generally occurs around one year old (e.g. Tardif et al., 2008). However, comprehension of language starts well before that. For example, infants of 6-month have already some knowledge of common nouns (Bergelson & Swingley, 2012; Tincoff & Jusczyk, 1999). Besides some early linguistic knowledge, infants as young as 6 months old already possess a rich understanding of the world around them, for example in the domains that have been described as belonging to "core knowledge": representations of objects, actions, numbers, places, and social partners (Spelke et al., 1992; Spelke & Kinzler, 2007).

The importance of studying event representations in preverbal infants is that they have less linguistic experience than adults and are not yet capable of producing event descriptions. When a certain capacity is thus found in preverbal infants, it can be inferred that it does not depend on the ability to express that linguistically, but is rather linked to non-linguistic (or, pre-linguistic) representations. In this way, preverbal infants make for a fairer comparison with non-human species than human adults.

The participants who partook in the present thesis were French infants of approximately 6 months old. They were tested in the babylab called Labo des Minots in Marseille, France. They have been recruited through posts on social media, invitations sent after consultation of the birth certificate registers of the city hall and word of mouth. Infants were at all times accompanied by their parent(s).

1.3.1.3. Human adults

Most human adults are proficient users of their native language. Illustrative of this, it has been estimated that native American English speakers know about 42000 lemmas by the age of twenty years old (Brysbaert et al., 2016).

The people who participated in the present studies, were recruited through Prolific, an online platform where experiments can be hosted. We set our criteria to accept only participants who reported to speak English fluently, so that they could understand the instructions, but the participants had a variety of native languages.

1.3.2. A word regarding methodological considerations

This thesis adopts the comparative approach. This approach aims to use data on extant species to make cladistic inferences about the traits of an extinct common ancestor to identify evolutionary continuity (or discontinuity) in terms of behaviour. Comparative studies with non-human primates, human's closest living relatives, are suitable to identify potential homologue traits given the relatively recent evolutionary divergence. To trace back the cognitive capacities of the common, extinct ancestor, ideally, one would need evidence for the characteristic under question in many living descendants, because only then it is safe to draw firm conclusions based on the parsimony argument that it would be

unlikely that a certain trait evolved independently in many lineages. Such an approach has been applied widely in order to understand the evolution of language (see: Weiss & Newport, 2006). Here, this tradition is continued by comparing two primate species: humans and Guinea baboons.

To study event representations in these two species, we aim to apply paradigms that are accessible without language knowledge. For this reason, inspiration is taken from the long history of doing experiments with non-human primates and preverbal infants. For these populations, the use of simple and controlled stimuli has an important advantage. As outlined by Kominsky and colleagues (2022) for infant research, the use of artificial stimuli increases measurement validity, making it more certain to tap into the construct of interest without unintended confounding variables that could explain the results or distract the participant. On the other hand, this simultaneously reduces the ecological validity, making the results less generalisable over different contexts. However, given that currently little is known about the format of event representations in non-human animals and whether it is similar or different from those in humans (e.g. Wilson et al., 2022), it appears to be the best strategy to start off with more controlled stimuli as a proof of concept, before trying to extrapolate to more real-life situations. For human adults, it has long been recognised that contingent motions of simplified stimuli, such as those developed by Heider and Simmel, can give rise to higher-order interpretations in terms of intentionality (Heider & Simmel, 1944; Scholl & Tremoulet, 2000). The use of simplified stimuli is not without its critics (e.g. Packer & Moreno-Dulcey, 2022). Recently, a study showed different looking durations for human adults and rhesus macaques when they were observing Heider-and-Simmel animations, which prompted the authors to conclude that the monkeys did not attribute mental states to the simple shapes (Schafroth et al., 2021). However, these conclusions are quite a leap to take from free viewing. Possibly, when imbedded in an actual task that engages the non-human participants, results indicative of such perceptions may not be impossible. In all studies (Article 1, Article 2, Article 3) presented in this thesis, I take an approach with simplified events (embedded in a task for the non-human primates).

For the infants we made use of a habituation-dishabituation design. The idea is that looking times can be informative to infer mental capacities, such that the unexpected draws attention, giving insights in the representations that may underlie these expectations (Aslin, 2007). During the habituation-dishabituation paradigm (which is employed in Article 2), significant recovery of looking time for a novel stimulus after being habituated, is interpreted as showing that the test trial is discriminated from the repeated habituation trials. Next, the level of dishabituation is taken as a measure of surprisal, such that conditions in which looking time at test is highest, appear to be least expected. If such methods are applied with simple stimuli, the precise factor triggering such a response from babies can be individuated (without overinterpreting such looking differences (Tafreshi et al., 2014)). Habituation-dishabituation can thus be a powerful tool to assess whether infants discriminate stimuli and by using which features (Oakes, 2010).

For the Guinea baboons we made use of a match-to-sample paradigm in Article 1 and Article 2. In match-to-sample tasks, a stimulus is presented which is followed by two

comparison stimuli of which the participant has to choose one, which is followed by feedback. When there is an exact match of the sample stimulus and the delay between them is close to zero, all animal species have good performance (Lind et al., 2015). We adapted this paradigm with moving stimuli and comparison stimuli as arbitrary representatives of a category rather than an exact match.

In Article 3, we adapted a change detection paradigm. These types of tasks have shown that attention plays an important role in detecting changes (Rensink et al., 1997). Changes are noticed faster in areas in which attention has been attracted, this explains why changes to animate entities are noticed faster than to inanimate entities (Altman et al., 2016; New et al., 2007, 2010). A similar rationale is used for dot-probe tasks which show faster response times for touching the dot behind an attention-attracting picture and this type of paradigm has been successfully applied to non-human primates (van Rooijen et al., 2017).

2. Article 1

A comparative study of causal perception in Guinea baboons (*Papio papio*) and human adults

Floor Meewis[1,2*], Iris Barezzi[1], Joël Fagot[1,2], Nicolas Claidière[1,2], Isabelle Dautriche[1]

1. Centre de Recherche en Psychologie et Neurosciences, UMR7077, Aix-Marseille University, CNRS, Marseille, France
2. Station de Primatologie-Celphedia, CNRS UAR846, Rousset, France

Abstract

In humans, simple 2D visual displays of launching events ("Michottean launches") can evoke the impression of causality. Direct launching events are regarded as causal, but similar events with a temporal and/or spatial gap between the movements of the two objects, as non-causal. This ability to distinguish between causal and non-causal events is perceptual in nature and develops early and preverbally in infancy. In the present study we investigated the evolutionary origins of this phenomenon and tested whether Guinea baboons (*Papio papio*) perceive causality in launching events. We used a novel paradigm which was designed to distinguish between the use of causality and the use of spatiotemporal properties. Our results indicate that Guinea baboons successfully discriminate between different Michottean events, but we did not find a learning advantage for a categorisation based on causality as was the case for human adults. Our results imply that, contrary to humans, baboons focused on the spatial and temporal gaps to achieve accurate categorisation, but not on causality *per se*. Understanding how animals perceive causality is important to figure out whether non-human animals comprehend events similarly to humans. Our study hints at a different manner of processing physical causality for Guinea baboons and human adults.

Keywords: causality; nonhuman primate; Michotte; launching event; event categorisation task; event perception

Background

In humans, the impression of physical causality can be induced upon seeing simple 2D visual displays of moving geometric shapes (for review, see: Scholl & Tremoulet, 2000). When an object A moves towards a stationary object B, and if B starts moving as soon as A stops and touches B (a launching event, see Fig 1), people infer that A *caused* B to move. Introducing a temporal gap or spatial gap between A and B eliminates this impression of causality. This phenomenon has first been studied by Michotte in the 1940's (Michotte, 1963) and it has inspired a wealth of research ever since.

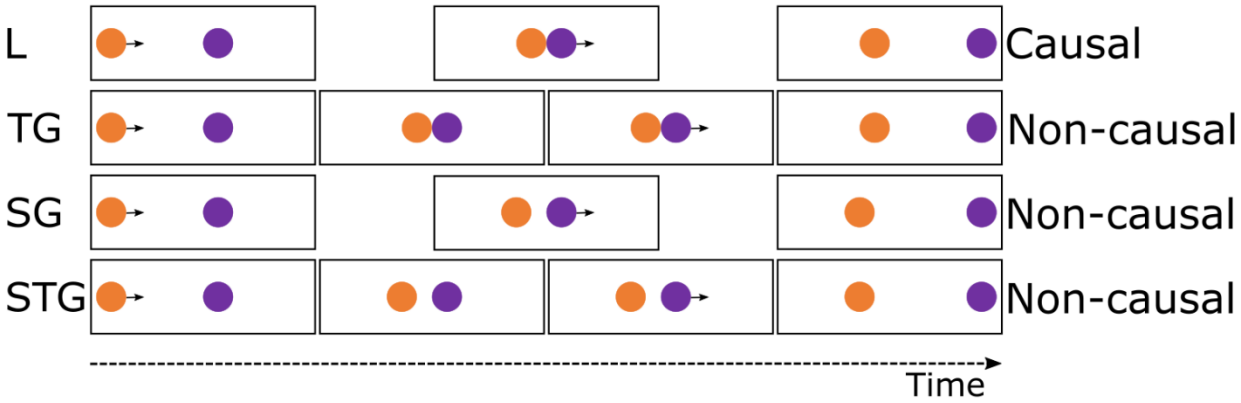


Fig 1: Four different Michottean events.
Michottean events with the relevant event frames in temporal order from left to right. From top to bottom: Direct launching event (L), temporal gap event (TG), spatial gap event (SG), spatial+temporal gap event (STG). Only the direct launching event is regarded as causal, the other three as non-causal.

Much debate has revolved around the mechanisms of causality inference and whether it involves deliberate and conscious cognitive mechanisms or can be explained as a perceptual process (e.g. Rips, 2011; Schlottmann & Shanks, 1992). Michotte himself adhered to this latter interpretation (1963) and recent studies demonstrate that – at least for simple launching events – causal perception is indeed a perceptual phenomenon (Kominsky & Scholl, 2020; Moors et al., 2017; Rolfs et al., 2013).

Already at birth, babies show an attentional bias towards the continuous movements of direct launches compared to events with a temporal delay (Mascalzoni et al., 2013). Newborns appear thus to be sensitive to the specific spatiotemporal features that differentiate between causal and non-causal events, but this does not necessarily mean they use causality to distinguish between them. This ability is suggested to develop later in life: habituation-dishabituation studies revealed that six-months-old infants, but not younger infants (but see: Rakison & Krogh, 2012), group together different non-causal events, thereby categorically distinguishing causal from non-causal events (Cohen & Amsel, 1998; Oakes, 1994). Consequently, there seems to be a developmental trajectory where younger infants employ a feature-based discrimination strategy, based on spatiotemporal cues (i.e., the presence or the absence of a spatial or a temporal gap), to

distinguish Michottean events, which transforms into a category-based strategy, based on causality, from six months onwards. It is important to note that when actual toys are used as stimuli, this categorical distinction is only evidenced from ten months old (Oakes & Cohen, 1990). This may be because the more realistic events distract infants from focussing on causality, suggesting that the use of simple and abstract figures is a strong method for studying causal perception (Kominsky et al., 2022).

Whilst the developmental origins of causal perception have received some attention, much less is known about its phylogenetic roots. Causal perception has been studied in only a limited number of non-human animal species using protocols that test the animals' sensitivity to violations of physical causality. For example, young Eurasian jays (*Garrulus glandarius*) and chimpanzees (*Pan troglodytes*) look longer at videos of "spatial gap events" where a food object is moved without contact (resp. Davidson et al., 2017; O'Connell & Dunbar, 2005). Similar experiments have been conducted on other species using 2D Michottean events. Newly-hatched chicks (*Gallus gallus*) have an innate predisposition to orient towards the launcher in causal direct launch, but such a preference does not exist for a non-causal temporal gap event (Mascalzoni et al., 2010). Similarly, pet dogs (*Canis familiaris*) look longer and show more pupil dilation to a non-causal event with a spatial gap compared to a causal event (Völter & Huber, 2021). Similar to looking preference designs in human infants, an important limitation of these studies is that they cannot disclose whether this discrimination is based solely on low-level features or (also) on the concept of causality. In a go/no-go task where pigeons (*Columba livia*) were trained to peck to discriminate between Michotte's causal and non-causal stimuli, the causal distinction was difficult to learn and instead the pigeons focussed on low-level features (Young et al., 2006). Therefore, it remains unclear whether non-human animals can experience an impression of causality with such stimuli. One study, however, revealed that chimpanzees are susceptible to the 'causal capture effect' by showing that synchronous presentation of a causal event (bouncing) increases the experience of a second, ambiguous bounce/stream animation as causal (Matsuno & Tomonaga, 2017). However, this study did not use the same Michottean launching events as are usually presented in other studies on human adults and infants, preventing direct comparisons.

Extracting causality from ongoing events is at the core of how humans represent events (Radvansky & Zacks, 2011; Zacks & Tversky, 2001), but it is yet to be established whether non-human animals process the world around them in the same way (Wilson et al., 2022). A recent study showed that, similarly to humans, apes may have a bias towards animate agents in realistic videos, but causality was not specifically addressed (Brocard et al., 2024). Since the chimpanzee may be the only non-human species for whom causal perception for Michottean-like events is evidenced, it is relevant to investigate whether species with whom humans share a less recent common ancestor also perceive causality. For some of those species, it has been reported that they respond differently to causal and non-causal events (e.g. pet dogs), but there is no conclusive evidence whether this is because of the causal dimension. Therefore, to further investigate the evolutionary origins of causal perception with the classic Michottean events, we tested whether Guinea baboons (*Papio papio*) categorise these events based on causality. We subsequently tested

human adults on the same categorisation task to compare the two species. Guinea baboons and humans share a common ancestor that lived about 30 Mya, so this comparison could, in concert with the previous findings in chimpanzees, shed light on the evolution of causal perception in a wider part of the primate lineage. Additionally, Guinea baboons are considered an excellent model species for studying human cognitive evolution (Fagot et al., 2019; Fischer et al., 2017, 2019) and are therefore a promising species to investigate causal perception. Because causal perception has been frequently documented in humans, shown to be early developing and perceptually processed, we hypothesised that both baboons and human adults would perceive causality and therefore use it to classify Michottean events.

In our study, we firstly trained the participants to distinguish a direct launch (Fig 1A) from a similar event which included a spatial and temporal gap (Fig 1D). Next, we introduced events with just a temporal (Fig 1B) or just a spatial gap (Fig 1C) and tested whether it was easier to categorise these events according to a division based on causality. We expected to find a faster learning rate for a categorisation based on causality compared to control, based on the assumption that the saliency of the causal events would set them apart from the non-causal events.

Methods

Methods have been preregistered and materials are available here: https://osf.io/vt5g8/.

Ethics statement

Baboons

The study on baboons was carried out in accordance with French and EU standards and received approval from the French Ministère de l'Education Nationale et de la Recherche (#APAFIS-2717-2015111708173794-V3). Procedures used in the present study were also consistent with the guidelines of the Association for the Study of Animal Behaviour. During this study, we collected behavioural data only. The study is non-invasive; no anaesthesia was administered, nor were the baboons sacrificed at the end of the study. Potential suffering is minimised as the baboons can partake in the experiments or not, as they please.

Humans

The study on human adults has received approval from the CEEI of INSERM (#20-733). We provided information about the task to the human participants and they consented to participate voluntarily in the study after having read the instructions by clicking on a button that said "I agree" before entering the online task.

Participants

Baboons

Participants were 23 Guinea baboons (*Papio papio*, 16 females, mean age = 12.9 ± 1.2(SEM) years, age range = 4.5-23.1 years, see S1 Table) who were housed at the Station de Primatologie in Rousset-sur-Arc (France) in two groups of eighteen and five individuals. The enclosure has a large outdoor area (700 m^2) which is enriched with big stones as climbing structures, and a small indoor area which is accessible for shelter. The baboons are not water- or food-deprived: they always have access to water and are fed daily (monkey chow, fruits, vegetables). The health of the monkeys is monitored by the keepers and veterinarians.

For testing, we used the automatic learning device for monkeys (ALDM) developed by Fagot & Paleressompoulle (2009). With this system, the baboons had *ad libitum* access to testing booths with touch screens where they participated voluntarily in computerised tasks employing an operant conditioning method. This testing method, in which the baboons can come and go as they please, reduces stress levels compared to periods during which the ALDMs are not accessible (Fagot et al., 2014). The baboons have been tested with the ALDM for approximately ten years before the start of the current experiment, but they have never been exposed to Michottean events or any other video stimuli. Data collection took place in May and July 2022. For a pilot study held in January, we showed the small group of baboons (N=5) Michottean events in a different task and with different stimuli. Additionally, a technical error prevented a smooth display of some of our stimuli of the first phase (see below) of the first stimulus set in May, after which we adjusted them and restarted our experiment.

Humans

On the 23rd and 24th of May 2023, we recruited 40 participants through the Prolific website, a platform where online experiments can be hosted. We set our criteria to accept only participants who reported to speak English fluently, so that they could understand the instructions. We paid participants £5 for their time.

We analysed the data from 39 participants (female = 21, male = 18; mean age = 27.5 ± 1.4(SEM) years, age range = 20-59 years). Of the recruited people, we excluded one participant because they did not complete the experiment in one go but paused for several hours in the middle. We had piloted our study in April/early May 2023 on twenty-one participants to gain insight into a possible effect size to estimate our needed sample size and to calculate the time needed for participants to finish the task to decide on appropriate participant payment.

Procedure

We trained the monkeys and humans using a match-to-sample (MTS) paradigm to first discriminate and secondly categorise Michottean causal and non-causal events, see Fig 2. For each trial in the MTS task, a video was used as a sample, followed by two distinctive coloured shapes as comparison stimuli of which the participant had to touch one. The

sample video was displayed in the centre of the screen, and the comparison stimuli were displayed for the monkeys in its bottom part and for the humans in the middle, with their right-left location counterbalanced across trials. The comparison stimuli were used as arbitrary depictions of the categories and which response stimulus was assigned to which event category was counterbalanced across participants. Training took place by giving feedback to the participants. For the monkeys, we rewarded correct responses with some wheat grains, whereas incorrect responses were punished with a 3s green time-out screen. Touching the screen during the video interrupted the trial and also led to a 3s green time-out screen. The humans received a 1s green screen with the word 'Yes!' to correct responses and a 3s red screen with the word 'No!' to incorrect responses. We presented the trials in random order in blocks and determined the classification accuracy by calculating the percentage of correct responses for each block. The baboons had blocks of sixty trials and the humans blocks of eight trials in the training phases.

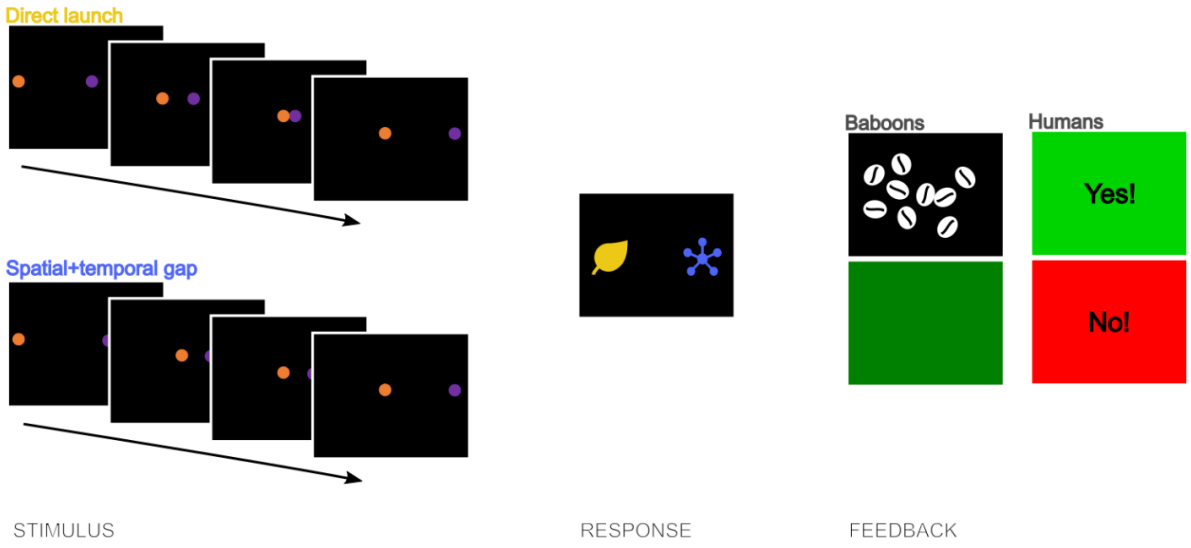


Fig 2. Match-to-sample paradigm.
The participant (baboon or human) views one of the Michottean events play out, which is followed by a screen with two response stimuli. Upon pressing one of the buttons, the participants receive feedback; for the baboons, correct results in a food reward and incorrect in a 3s green time out screen, for the humans, correct leads to a green screen with 'Yes!' and incorrect to a 3s red screen with 'No!'.

The humans received minimal instructions before the start of the experiment. We explained that they would see videos which they had to classify by pressing one of two buttons. We also told them they would receive feedback in some of the trials (as a green or red screen) and in some other trials no feedback in the form of a purple screen. Lastly, we told them that the task's duration would increase with wrong answers, to give them an incentive to try to answer correctly. Causality was not mentioned. After the experiment had finished, the human participants were asked to answer to a question on how they had

categorised the videos. We did this to test whether people would explicitly mention causality (of the 39 participants, 13 mentioned that some videos were causal [8 times in the causal and 5 times in the control condition]). Also, we asked about their age, gender, nationality and first language.

Stimuli

The stimuli were based on Michotte's causal and non-causal events (Michotte, 1963; see Fig 1; Scholl & Tremoulet, 2000). We used four event types: the direct launch (L), the temporal gap event (TG), the spatial gap event (SG) and the spatial and temporal gap event (STG). The temporal gap (present in TG and STG) lasted always for one second and the spatial gap (present in SG and STG) was always a quarter of the area between the start of object A and the end of object B. Note that the STG event always had both of these gaps and the L event had none. Multiple variants were created for each event category in which we changed the speed of the objects and the length of the trajectories of objects A and B, making the defining features of each event type the presence or absence of the gap(s), see S2 and S3 Tables. All videos were between 800 and 2600ms. We used two distinct stimulus sets. The first stimulus set consisted of videos depicting an orange circle as object A and a purple circle as object B with a movement direction from left to right. The second stimulus set used a blue square as object A and a yellow square as object B with a movement direction from right to left. Within each stimulus set, the colours of object A and B are used consistently regardless of event type (L, SG, TG and STG). Stimuli were presented on a black background. The videos were made using the animation features of Microsoft® PowerPoint® and subsequently exported in a mp4 format with a resolution of 944x 720 pixels and a frame rate of 30 frames/s. Videos and response stimuli were presented to the baboons using E-prime (V2) and to the humans using Labvanced (Finger et al., 2017). All videos are available on OSF (in the folder named Stimuli).

Experimental phases and conditions

The experiment consisted of four phases. An overview of the experiment is presented in Fig 3. The participants started with phase 1 and continued in the experiment according to their performance. The baboons participated two times in the experiment to do both conditions, each with a different stimulus set. The humans were randomly assigned to a condition and a stimulus set.

Phase 1: L vs. STG

First, we trained the participants to discriminate one L from one STG video, see videos 1 and 2 in S2A Table. These two events were matched for their total duration (i.e., 1600 ms). Participants were presented with blocks of randomly ordered trials, such that, within a block, there were half of L trials and half of STG trials. To continue to the next phase of the experiment, the baboons had to reach 80% correct classification for both events within the block (60 trials) and the humans had to reach 100% correct classification for all the trials in the block (8 trials).

After reaching these pre-set criteria, two extra events were added, one L and one STG, see videos 3 and 4 in S2B Table. With these two events, we controlled for distinction strategies that employed the following 6 'irrelevant' features : (1) speed, (2) durations, (3) distances of movements of the objects, (4) initial and (5) end positions of the objects, and (6) total duration of the events (see S3 Table). That is, if a participant tried to distinguish the events based on one of the above-mentioned features, they would not succeed. We therefore trained the participants to classify using a strategy which could involve causality or paying attention to the presence or absence of the gaps. Each block consisted of 25% of previously learnt L events, 25% of previously learnt STG events, 25% new L events and 25% new STG events. When the baboons reached the criteria of 70% correct on all four events, they continued to the next phase.

Phase 2: Generalisation 1

To verify whether the participants had not reached high accuracy scores by just memorising the four training videos and how they associate with the response stimuli (which they had seen many times throughout training), we tested whether they could transfer their learnt categorisation to new L and STG events. Here, we tested the performance of the participants on four probe events that they had not seen before: two L and two STG events, see videos 5, 6, 7 and 8 in S2C Table. These probes were different in terms of durations, speed, position of the point of impact, initial position of object B and end position of object A, compared to the training events.

For the baboons, we used blocks of 60 trials, including 8 probe trials (4 L; 4 STG) and 52 previously learnt Phase 1 trials (26 L and 26 STG), and repeated these blocks sixteen times. The probes were randomly rewarded in 80% of cases regardless of the response given to avoid any learning for these trials. For the humans, we presented three times a block of six trials composed of the four probe events, and the two previously learnt events from the first training phase. We did not give the humans feedback on these probe trials, but instead showed a purple screen for two seconds after they made their choice, regardless of the response given, which was explained in the instructions as no feedback.

Before going to phase 3, we also tested the participants' reaction to Michottean events of which the direction had been reversed. The same method as for phase 2 generalisation 1 was employed using these reversed events as probes. This test was designed for a separate study and will not be discussed further in the current paper.

Phase 3: L/STG vs. rest

This crucial phase tested whether a categorisation based on causality would lead to a learning advantage compared to a categorisation which did not involve causality. We trained the participants to classify events with just a spatial gap (SG) and just a temporal gap (TG), into their previously learnt categories either forming a causal vs. non-causal division or a control division. Half of the trials of each block were comprised of the four previously learnt L and STG events and half of the new SG and TG events (each 25%), see videos 9 and 10 in S2D Table.

In the causal vs. non-causal condition, the participants had to learn to press the same response stimulus for the SG and TG events as for the STG event (see Fig 3, Phase 3, right), effectively dividing causal from non-causal events. In the control condition, participants had to press the same response stimulus for SG and TG as for the L event (see Fig 3, Phase 3, left), leading to a mix of causal and non-causal events. Critically, both the SG and the TG events are equally different from the STG event as from the L event in terms of their perceptual features (presence or absence of a gap; see Cohen & Amsel, 1998). This means that learning to group together TG and SG with STG in the causal vs. non-causal condition implies a difference of the size of the spatial gap (always ¼ of the screen) and the size of the temporal gap (1s) and exactly these differences are also present for grouping them with L. Since we do not know whether the spatial gap and temporal gap are of equal salience, we decided to divide the classification such that the participants always had to learn a rule that implicated both the temporal and the spatial domain.

The monkeys were alternatingly assigned to one of the conditions (causal vs. non-causal or control) upon reaching phase 3. A few months later, we ran the second part of the experiment with a new stimulus set; here, the monkeys were admitted to the condition they had not participated in. The humans were randomly assigned to one condition and one stimulus set.

For both conditions, we measured the learning speed as the number of blocks that were needed to reach a pre-determined learning criterion. For the baboons we set this at 70% correct per event type (L, STG, SG and TG) in a block within a maximum of 35 blocks, and for the humans we set it at 100% correct for the whole block.

Phase 4: Generalisation 2

Lastly, we tested whether the categorisation learnt in Phase 3 would transfer to new, unseen SG and TG stimuli. We added two new SG and two new TG events which had different speeds and durations compared to the training videos, see videos 11, 12, 13 and 14 in S2E Table. For the baboons, we introduced eight probe videos per block, which were randomly rewarded in 80% of cases, and they were intermingled with 52 trials of previously learnt events (17 L, 17 STG, 9 SG and 9 TG). We repeated the block sixteen times. For the humans, we showed three times a block of six trials composed of the four probe events and one L and one STG event from Phase 1 with the two-second purple screen as uninformative feedback.

Statistical analyses

All statistical analyses were conducted in R (R Core Team, 2021).

Phase 1 (learning L vs. STG): we did not conduct any statistical analyses, but we calculated how many blocks were needed to reach our pre-set learning criteria for both the baboons and humans. Additionally for the baboons, we tracked which individuals passed the criteria.

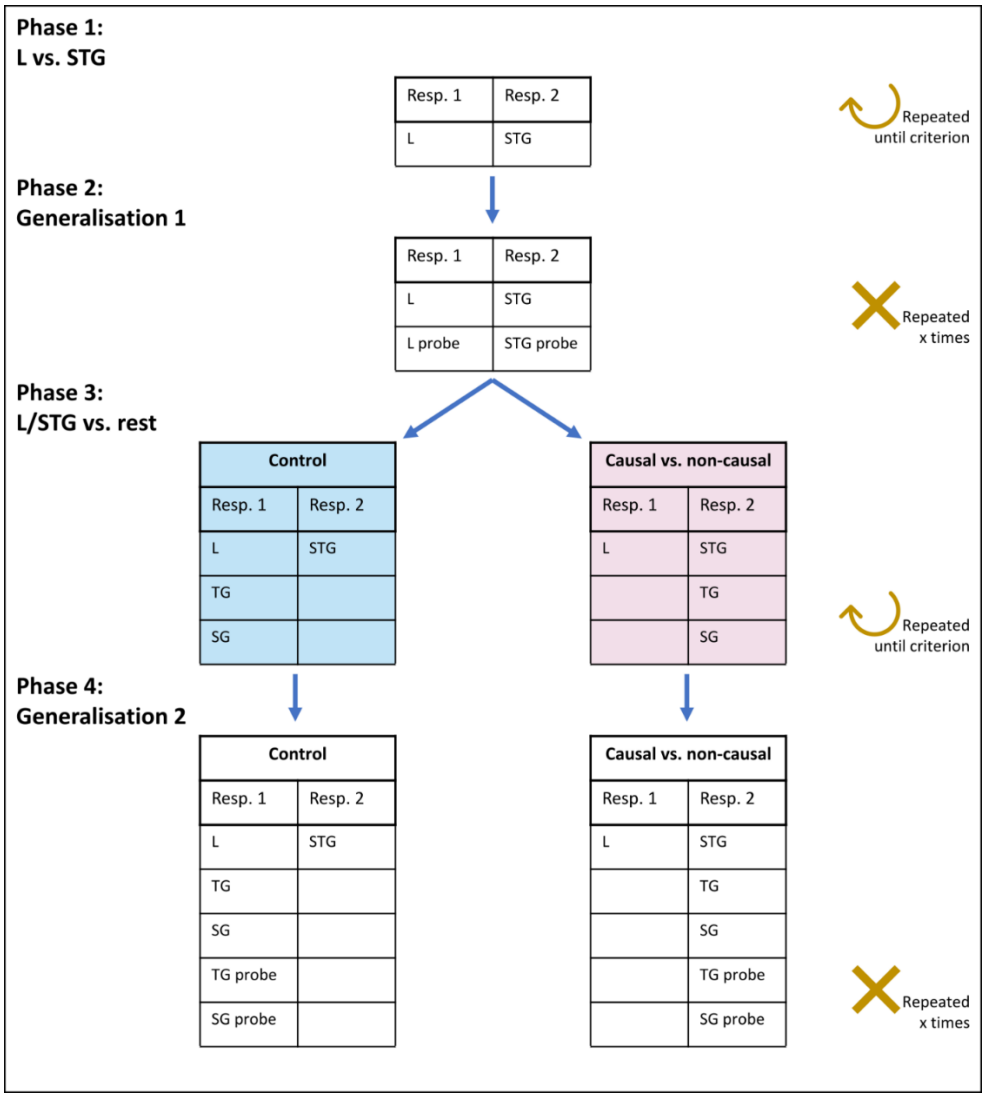


Fig 3. Overview of the experimental phases. The Michottean events that are grouped together under the same response stimulus (resp. 1 or resp. 2) are shown. The participants were trained to discriminate L from STG in Phase 1 and tested on their transfer ability to new probe events in Phase 2. In Phase 3, TG and SG events were added and the participants were trained according to a control (in blue) or causal vs. non-causal (in pink) condition. Transfer to new probe SG and TG events was tested in Phase 4.

Phase 2 and Phase 4 (generalisation phases): For each generalisation phase, we implemented a generalised linear mixed model with binomial error structure using the lme4-package (Bates et al., 2015) to model the effect of trial type (baseline or probe) and the effect of event type (L, STG, SG, TG) on trial accuracy (0/1). We also checked whether the stimulus set (first or second) would affect the accuracy. Adding stimulus set was not preregistered, but we wanted to verify whether the different stimuli would not affect the performance. For the second generalisation phase (Phase 4), we also implemented post-hoc an interaction between the condition (causal or control) and event type, because the distribution of responses for the event types depended on the condition (see Fig 2, Phase

3). We allowed the intercepts to vary per participant. We tested the main effects using likelihood ratio tests.

For Phase 2, the first generalisation phase, we used the following model: trial accuracy ~ trial type + event type + stimulus set + (1|participant).

For Phase 4, the second generalisation phase, we used for the baboons the following model: trial accuracy ~ trial type + event type * condition + (1|participant). We have thus removed the stimulus set, because it did not improve the model fit. Post-hoc, we tested whether it is easier to tell apart L from the other events in the causal vs. non-causal condition compared to the STG in the control condition by using contrasts of the estimated marginal means using the 'emmeans' package (Lenth, 2022). For the humans, the model with interaction between condition and event type did not converge, so we continued with a model without this interaction. Additionally, we removed trial type, since the probes and baseline had non-overlapping event types.

We tested for Phase 2 and Phase 4 whether the percentage of correct responses per event type was above 50% (chance level) by using the intercept values of the model summary and changing the baseline categories accordingly.

Phase 3 (control vs. causal/non-causal conditions): We compared the learning speed for the causal vs. non-causal compared to the control condition, i.e., how many blocks of trials are needed to reach the learning criteria in each condition. For the baboons, we used a within-subjects design, and we tested whether there was a difference in learning speed between the two conditions using a non-parametric sign-test. This method deviates from what was anticipated in the preregistration and was implemented because the intended model with random intercepts per participant did not converge as too few individuals participated in both conditions (N = 6). Additionally, the upper limit of 35 blocks to reach the learning criteria prevented us from accurately determining the shape of the distribution and apply standard parametric tests. To compare learning speeds of the control and causal division in the human data, we implemented a between-subjects design, and we used a Mann Whitney U-test because the data did not show equal variances and normal distributions. Finally, we checked the responses of the human participants to see whether there was mention of causality in their reported strategies.

Results

Phase 1: L vs. STG Twelve baboons learnt to discriminate one L from one STG event for at least one of the stimulus sets, which took on average 53.1 ± 12.3(SEM) blocks of sixty trials. Subsequently, these baboons continued to the next step in the experiment where we added two more events (one L and one STG). From these four training videos, eleven baboons learnt to discriminate L from STG in on average 36.0 ± 10.2(SEM) blocks, see S1 Table for the individual performances. All human participants learnt to discriminate L from STG in two videos after on average 4.5 ± 1.1(SEM) blocks of eight trials and continued in the experiment and again learnt the division for four videos after 2.4 ± 0.5(SEM) blocks.

Phase 2: Generalisation 1 The eleven baboons performed above chance on categorising the newly introduced probe events, for both L and STG, as well as for the previously learnt training events of L and STG (all < .001, all mean accuracy above 62.0±3.4(SEM)%). Additionally, there was an effect of stimulus set on accuracy (χ^2 = 9.44, p = .002); generalisation was more accurate for the first (72.0±1.8(SEM)%) than the second stimulus set (68.9±1.8(SEM)%). Also, our model showed an effect of trial type (χ^2 = 60.66, p < .001) showing that baboons performed more accurately on the training events (73.0±1.4(SEM)%) than on the probe events (67.3±2.2(SEM)%). Event type did not influence accuracy (L vs. STG, χ^2 = 1.35, p = .245).

For the humans, the data of generalisation 1 showed that the participants performed very accurately on all event types and trial types, such that our final sample included 47 incorrect to 655 correct responses. To check whether this excess of 1's would pose a problem, we tested for possible "1-inflation" using the DHARMa package (Hartig, 2022) which indicated no issues (p = .612) and led us to continue with our original binomial model. Our model showed no effect of trial type (χ^2 = 0.63, p = .429) and no effect of event type (χ^2 = 0.20, p = .656). They performed significantly above chance on all shown events (all p < .001, all mean accuracy above 91.45±2.9(SEM)%).

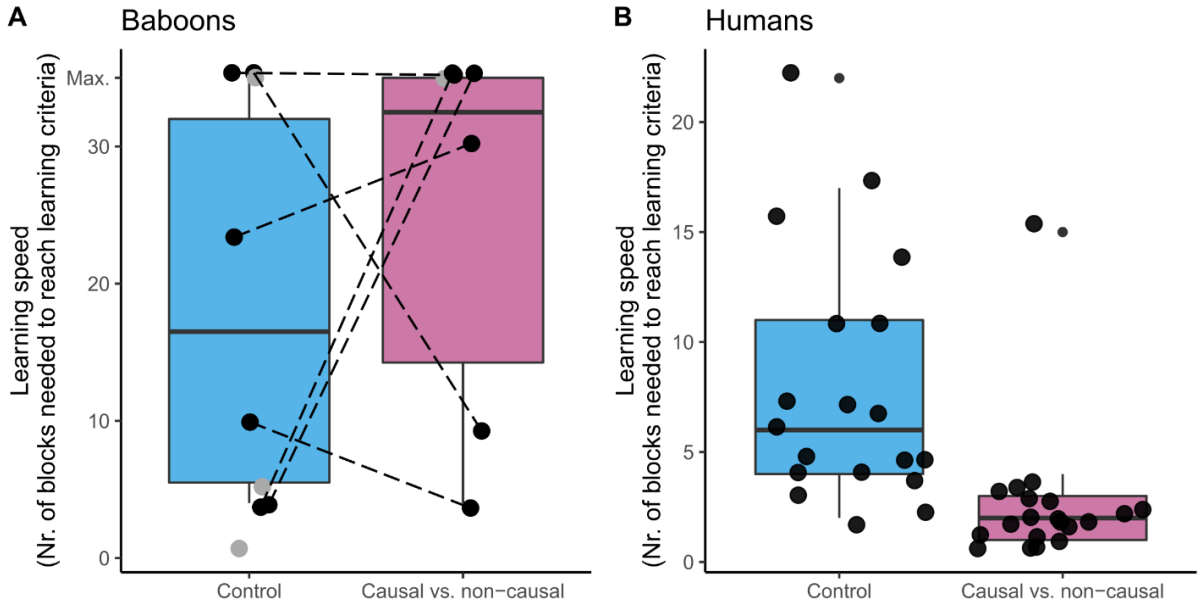


Fig 4. Baboons' and humans' learning speed to learn the control and causal vs. non-causal condition. Learning speed is expressed as number of blocks of trials needed to reach our learning criteria. (A) The baboons were equally fast to learn the control as the causal vs. non-causal condition. Dotted lines connect individuals who have participated in both conditions shown in black, grey dots are individuals who participated in one condition and do not make part of the statistical test. (B) The humans were faster to learn the causal condition than the control condition.

Phase 3: L/STG vs. rest Of the 11 baboons who participated in this phase in at least one of the conditions, 5 participated in both of the conditions and managed to reach our learning criterion in at least one condition. There was no significant difference between the speed at which baboons learnt the causal condition (mean = 24.7 ± 5.8(SEM) blocks) and the speed at which they learnt the control condition (mean = 18.5 ± 5.9(SEM) blocks; N = 5, successes = 2, p = 1), see Fig 4.

All human participants learnt the condition they were assigned to. Humans were significantly faster to learn the causal condition (2.7±0.7(SEM)) compared to the control condition (8.0±1.3(SEM); N = 39, W = 38.5, p < .001).

Phase 4: Generalisation 2 For the baboons, our model showed a significant effect of trial type (χ^2 = 25.00, p < .001) such that baseline trials resulted in higher accuracy than probe trials (80.0±2.6(SEM)% for baseline and 77.0±3.1(SEM)% for probe). Also, we found an interaction effect between condition and event type (χ^2 = 843.56, p < .001). Further exploration showed that the participants performed significantly above chance on the newly introduced probe events (all < .001, all mean accuracy above 66.1±6.7(SEM)%), but, on the previously learnt baseline trials the performance on the launching event (L) in the causal vs. non-causal condition did not differ significantly from the 50% chance level (p = .05, mean accuracy = 54.8 ±10.0(SEM)%), see Fig 5. Accuracy on all other baseline event categories for both conditions remained significantly above chance level (all p < .001, all mean accuracy above 68.0±5.5(SEM)%). To check whether the L event stands out from the others, our post-hoc analysis showed that for the event category that had to be classified separately from the other three (i.e., L in the causal vs. non-causal condition and STG in the control condition), baboons were worse at classifying launching events in the causal vs. non-causal condition than classifying the STG event in the control condition (54.8±10.0(SEM)% vs. 68.0±5.5(SEM)%; estimate = 0.498, SE = 0.227, p < .001).

On the human data, there was a significant effect of condition (χ^2 = 14.12, p < .001), indicating that the participants in the causal vs. non-causal condition were more accurate (93.3±1.7(SEM)%) than participants in the control condition (84.4±2.9(SEM)%). Additionally, there was an effect of event type (χ^2 = 20.95, p < .001), reflecting that the performance on SG events (81.2±4.4(SEM)%) was lower than the three other event types (L = 91.5±2.7(SEM)%, STG = 89.7±3.7(SEM)% and TG 93.6±2.2(SEM)%). Importantly, all event types had an accuracy that was significantly above chance level (all p < .001, all mean accuracy above 64.04±7.13(SEM)%), demonstrating that the humans successfully categorised previously unseen SG and TG events.

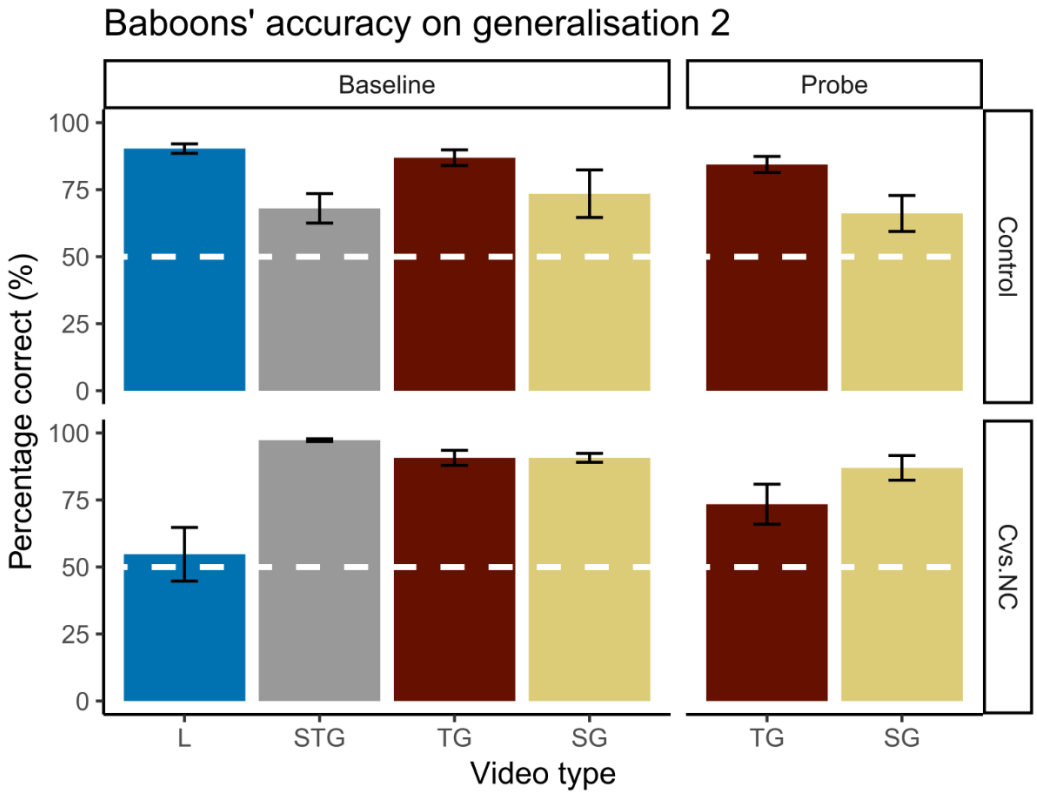


Fig 5. Baboons' performance on generalisation 2. The y-axis shows the percentage of correct classification, with the chance level of 50% correct marked in white. The different Michottean event types are displayed on the x-axis. On top the control condition and at the bottom the causal vs. non-causal condition.

Discussion

We introduced a new paradigm to assess whether individuals favour causality or spatiotemporal properties when presented with Michottean events. In this categorisation task, Guinea baboons were as successful to learn to categorize Michottean events following a causal division (causal vs. non-causal) as they were to categorize these events according to a control division that was equal in terms of spatiotemporal differences. By contrast, humans were faster in the causal categorisation condition. This showed that, even though both species demonstrated the ability to successfully distinguish and categorise Michottean events and generalise to novel events, only the humans made use of causality. Guinea baboons relied on spatiotemporal properties only to solve the categorisation task. Even though we cannot rule out that Guinea baboons have causal representations of Michottean events, we show that where humans apply causality to sort Michottean events, Guinea baboons rather use a feature-based discrimination.

One could argue that the 2D stimuli may be too distant from any natural experience of causality baboons routinely encounter, and therefore they may not distil a causal relation from the depicted events. There are two reasons to doubt that the use of artificial stimuli is the reason why the baboons did not categorise according to causality in our task. First, in infant studies, where it is common practice to employ simplified stimuli (cf. Kominsky

et al., 2022), causal perception has been evidenced in very young infants using similar 2D shapes (Cohen & Amsel, 1998; Leslie & Keeble, 1987; Oakes, 1994). Similarly to non-human animals, 6-month-old infants' experience with animated shapes is limited. As such, these stimuli are as unnatural and unfamiliar for baboons as for infants, suggesting that it is unlikely that experience with the stimuli may explain a difference in performance between these two populations. We believe that the baboons who participated in our study may actually have *more* experience engaging with on-screen stimuli than infants, because of their involvement in other studies using artificial stimuli in the same computerised set-up (e.g. Dautriche et al., 2022). Secondly, even when demonstrations with real-life 3D objects are employed, monkeys sometimes have difficulty to use contact-mechanical information over spatiotemporal features, as was shown for rhesus macaques (*Macaca mulatta*) who failed to search in the right location for food which had been stopped from rolling by a barrier (Santos, 2004). Instead, the rhesus performed better on tasks that centred around finding the food using spatiotemporal information, a point that we come back to below. Because of these reasons, we think that the use of 2D stimuli is appropriate for baboons and not responsible for their absence of use of causality.

We note, however, that our categorisation task was challenging for the baboons to learn as only 7 of 11 baboons were able to reach our learning criteria in phase 3, where they had to learn to categorise spatial gap events and temporal gap events, suggesting difficulties in the task at this point. Note that the other baboons of the group who did not pass phase 1 (learning to distinguish a direct launch (L) from a spatial and temporal gap event (STG)), may have done so for various reasons, such as lack of interest, not necessarily because they did not or could not understand the task. The difficulties during phase 3, can presumably be explained by the discrepancy of correct responses between the two comparison stimuli. In phase 3, baboons had to learn to answer one response stimulus in a quarter of the trials, and the other response stimulus in the remaining three quarters of the trials. This discrepancy has driven some baboons towards a response strategy where one response stimulus is touched all the time regardless of the Michottean event stimulus, leading to a 75% reward pay-off with minimal effort. This is also visible in the subsequent generalisation phase: the baboons' performance to set apart the causal event in the causal vs. non-causal condition worsened to chance level and was even lower than their ability to set apart a spatial and temporal gap event in the control condition. This suggests that the causal nature of the launching event is not an outstanding feature for the baboons. In contrast, the humans perform well despite this difficulty, indicating that causality is something that they pick up on with ease. Overall, it is clear that the causal/non-causal division has not been salient enough for the baboon participants to override this bias towards one response stimulus.

Besides causality, what other strategies for successful classification were at the baboons' disposal? Certainly, baboons could not have used superficial features of our stimuli, such as the speed of the objects or the duration of the events, since our choice of stimuli controlled for those. To learn the classifications of the two conditions equally fast, the only possible way is to rely on the presence or absence of the temporal and spatial gap. Such a categorisation strategy where focussing on a combination of features instead

of a concept can account for successful classification, is in line with the feature model for discrimination (e.g. in pigeons, Von Fersen & Lea, 1990). In our task, this points to using the addition of spatial and temporal properties instead of using causality (based on these features). The deployment of a feature-based strategy finds support in previous work which shows that, in other tasks, Guinea baboons tend to prioritise local features over global aspects of presented stimuli (e.g. Barbet & Fagot, 2011; Dépy et al., 1997; Fagot & Deruelle, 1997). Other species, such as rhesus macaques, similarly demonstrate a preference to rely on spatiotemporal features over contact information when parsing (dynamic) mechanical relations (Santos, 2004). In sum, our results point to Guinea baboons employing a feature-based strategy, based on the presence or absence of a gap, to categorise Michottean events.

Other non-human animal species have also demonstrated discriminatory abilities between different Michottean events in previous studies (newly-hatched chicks: Mascalzoni et al., 2010; dogs: Völter & Huber, 2021), but, contrary to our design, these studies did not include ways to specifically test whether the animals employed causality or only spatiotemporal cues for their discrimination. The evidenced discrimination by the chicks as shown by a preference for the launcher in causal launches and no preference in temporal gap events, appeared to be driven primarily by the self-propelled motion of the first object (Mascalzoni et al., 2010), leaving open the question whether causality or the presence or absence of the temporal gap had played a role as well. Also the dogs might preferentially look towards the non-causal event (Völter & Huber, 2021), because of differences in the spatiotemporal properties of the stimuli, instead of causality *per se*. These studies thus show discriminatory abilities similar to what we have demonstrated in the first two phases of our experiment for eleven baboons, but we additionally implemented the specific test for sensitivity to causality. Similarly, Young *et al.* (2006) trained pigeons with a go/no-go paradigm to peck on one of four events displayed on a screen (L, SG, TG or STG), with the expectation that if pigeons discriminate the events based causality rather than spatiotemporal features, the individual trained to peck for the causal launching event should find the task easiest. However, even after extensive training, just one of four pigeons was able to generalise the event they learnt to peck on, to novel stimuli, whereas the other pigeons were not able to pick up on the relevant spatiotemporal properties nor on causality. Consequently, for all these non-human animals, evidence favouring the use of causality over low-level properties is still lacking.

Currently, chimpanzees appear to be the only species that is sensitive to causality in Michottean-like events (Matsuno & Tomonaga, 2017). How can we interpret the different results obtained between chimpanzees and, here, Guinea baboons? One possibility is that this difference stems from the use of different paradigms (causal capture effect for the chimpanzees vs. categorisation task for the baboons). Another possibility is that causal perception has evolved sometime after the separation between the Cercopithecoidea and the Hominoidea ~30Mya. This could explain why causal perception can be observed in humans and chimpanzees, but not on other species. Nonetheless, our study demonstrated a sensitivity to the specific spatiotemporal properties necessary for causality detection in

Guinea baboons, suggesting that this may be an evolutionary trait that is shared between Guinea baboons and humans.

Our findings for Guinea baboons clearly differ from those observed for the human adults, who were much faster to learn the task and who showed a clear advantage to learn the causal vs. non-causal division. Indeed, our findings support the idea that, for human adults, the distinction between causal and non-causal events is a central and prominent feature of perception. It has been reported numerous times that humans possess a capacity to quickly detect causality from visual displays (Scholl & Tremoulet, 2000) and we show this again with a novel paradigm. The causal vs. non-causal distinction could be seen as a mental default and has even been argued to underlie the occurrence of causative structures cross-linguistically (Strickland, 2017).

Contrary to human adults' ability to use causality when presented with Michottean launches, we believe the baboons' findings to be more in line with how human infants younger than six months of age react to Michottean launches. Cohen and Amsel (1998) showed with a habituation-dishabituation paradigm that it was only six-, but not four- and five-month olds, who did not dishabituate upon seeing a new non-causal event when they had been habituated to another. This suggest that infants younger than six months of age do not make use of causality yet, even though they are able to detect (spatiotemporal) differences between the stimuli from birth (Mascalzoni et al., 2013). It is clear, however, that the looking time methodology employed for human babies is different from the categorisation used in our study. It may be the case that an implicit looking time experiment is more sensitive to test a perceptual phenomenon than a categorisation task. Future studies could exploit eye-tracking or habituation paradigms to try and corroborate our hypothesis that Guinea baboons react similarly to infants younger than 6 months of age when presented with Michottean events.

Even though our findings indicate that Guinea baboons may not perceive causality (or at least do not use it for categorization) after being probed with Michottean launches, this does not necessarily imply that they lack the ability to detect causal structures altogether. Certainly, various animal species have been shown causal reasoning abilities, at least to a certain extent (Penn & Povinelli, 2007; Schloegl & Fischer, 2017; Völter & Call, 2017), and it is possible that these abilities are more likely to be detected by monkeys in actions/events that involve animate agents (Rochat et al., 2008; Wood et al., 2008). However, our studies point to a critical distinction between humans and Guinea baboons in their ability to transform visual sequences of motions with certain spatiotemporal parameters into a causal impression. To conclude, causality detection is needed to understand events (Zacks & Tversky, 2001) and in particular the interrelatedness of the entities involved (Elman, 2009). Extracting these event elements and how they relate to one another from ongoing experiences is crucial in understanding the world around us (Kuperberg, 2021). Investigating whether non-human animals process events similarly to humans provides insights into the evolutionary roots of event cognition (Brocard et al., 2024; Wilson et al., 2022). Understanding how animals perceive causality is an important starting point to figure out whether non-human animals comprehend events similarly to

humans. For causality perception in Michottean events, our study hints at a different manner of processing for Guinea baboons and human adults.

Acknowledgments

This work was supported by the Agence Nationale de la Recherche (ANR-20-CE28-0005) and the Institute for Language, Communication and the Brain (ILCB) through the Agence Nationale de la Recherche (ANR-16-CONV-0002). We thank Julie Gullstrand, Sebastien Barniaud and the staff at the Station de Primatologie for their help with the experiments with the baboons.

References

Barbet, I., & Fagot, J. (2011). Processing of contour closure by baboons (*Papio papio*). Journal of Experimental Psychology: Animal Behavior Processes, 37(4), 407–419. https://doi.org/10.1037/a0025365

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software, 67(1). https://doi.org/10.18637/jss.v067.i01

Brocard, S., Wilson, V. A. D., Berton, C., Zuberbühler, K., & Bickel, B. (2024). A universal preference for animate agents in hominids. iScience, 27(6). https://doi.org/10.1016/j.isci.2024.109996

Cohen, L. B., & Amsel, G. (1998). Precursors to infants' perception of the causality of a simple event. Infant Behavior and Development, 21(4), 713–731. https://doi.org/10.1016/S0163-6383(98)90040-6

Dautriche, I., Buccola, B., Berthet, M., Fagot, J., & Chemla, E. (2022). Evidence for compositionality in baboons (Papio papio) through the test case of negation. Scientific Reports, 12(1), 19181. https://doi.org/10.1038/s41598-022-21143-1

Davidson, G., Miller, R., Loissel, E., Cheke, L. G., & Clayton, N. S. (2017). The development of support intuitions and object causality in juvenile Eurasian jays (*Garrulus glandarius*). Scientific Reports, 7(1), 40062. https://doi.org/10.1038/srep40062

Dépy, D., Fagot, J., & Vauclair, J. (1997). Categorisation of three-dimensional stimuli by humans and baboons: Search for prototype effects. Behavioural Processes, 39(3), 299–306. https://doi.org/10.1016/S0376-6357(96)00757-7

Elman, J. L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. Cognitive Science, 33(4), 547–582. https://doi.org/10.1111/j.1551-6709.2009.01023.x

Fagot, J., Boë, L.-J., Berthomier, F., Claidière, N., Malassis, R., Meguerditchian, A., Rey, A., & Montant, M. (2019). The baboon: A model for the study of language evolution. Journal of Human Evolution, 126, 39–50. https://doi.org/10.1016/j.jhevol.2018.10.006

Fagot, J., & Deruelle, C. (1997). Processing of global and local visual information and hemispheric specialization in humans (Homo sapiens) and baboons (*Papio papio*). Journal of Experimental Psychology: Human Perception and Performance, 23, 429–442. https://doi.org/10.1037/0096-1523.23.2.429

Fagot, J., Gullstrand, J., Kemp, C., Defilles, C., & Mekaouche, M. (2014). Effects of freely accessible computerized test systems on the spontaneous behaviors and stress level of Guinea baboons (*Papio papio*): Computerized testing and welfare in baboons. American Journal of Primatology, 76(1), 56–64. https://doi.org/10.1002/ajp.22193

Fagot, J., & Paleressompoulle, D. (2009). Automatic testing of cognitive performance in baboons maintained in social groups. Behavior Research Methods, 41(2), 396–404. https://doi.org/10.3758/BRM.41.2.396

Finger, H., Goeke, C., Diekamp, D., Standvoß, K., & König, P. (2017). LabVanced: A unified JavaScript framework for online studies. International Conference on Computational Social Science.

Fischer, J., Higham, J. P., Alberts, S. C., Barrett, L., Beehner, J. C., Bergman, T. J., Carter, A. J., Collins, A., Elton, S., Fagot, J., Ferreira da Silva, M. J., Hammerschmidt, K., Henzi, P., Jolly, C. J., Knauf, S., Kopp, G. H., Rogers, J., Roos, C., Ross, C., … Zinner, D. (2019). Insights into the evolution of social systems and species from baboon studies. eLife, 8, e50989. https://doi.org/10.7554/eLife.50989

Fischer, J., Kopp, G. H., Dal Pesco, F., Goffe, A., Hammerschmidt, K., Kalbitzer, U., Klapproth, M., Maciej, P., Ndao, I., Patzelt, A., & Zinner, D. (2017). Charting the neglected West: The social system of Guinea baboons. American Journal of Physical Anthropology, 162(S63), 15–31. https://doi.org/10.1002/ajpa.23144

Hartig, F. (2022). DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. (Version R package version 0.4.6). https://CRAN.R-project.org/package=DHARMa

Kominsky, J. F., Lucca, K., Thomas, A. J., Frank, M. C., & Hamlin, J. K. (2022). Simplicity and validity in infant research. Cognitive Development, 63, 101213. https://doi.org/10.1016/j.cogdev.2022.101213

Kominsky, J. F., & Scholl, B. J. (2020). Retinotopic adaptation reveals distinct categories of causal perception. Cognition, 203, 104339. https://doi.org/10.1016/j.cognition.2020.104339

Kuperberg, G. R. (2021). Tea with milk? A hierarchical generative framework of sequential event comprehension. Topics in Cognitive Science, 13(1), 256–298. https://doi.org/10.1111/tops.12518

Lenth, R. V. (2022). emmeans: Estimated Marginal Means, aka Least-Squares Means (Version 1.9.0). https://CRAN.R-project.org/package=emmeans

Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? Cognition, 25(3), 265–288. https://doi.org/10.1016/S0010-0277(87)80006-9

Mascalzoni, E., Regolin, L., & Vallortigara, G. (2010). Innate sensitivity for self-propelled causal agency in newly hatched chicks. Proceedings of the National Academy of Sciences, 107(9), 4483–4485. https://doi.org/10.1073/pnas.0908792107

Mascalzoni, E., Regolin, L., Vallortigara, G., & Simion, F. (2013). The cradle of causal reasoning: Newborns' preference for physical causality. Developmental Science, 16(3), 327–335. https://doi.org/10.1111/desc.12018

Matsuno, T., & Tomonaga, M. (2017). Causal capture effects in chimpanzees (*Pan troglodytes*). Cognition, 158, 153–164. https://doi.org/10.1016/j.cognition.2016.10.023

Michotte, A. (1963). The Perception of Causality. Routledge. https://doi.org/10.4324/9781315519050

Moors, P., Wagemans, J., & de-Wit, L. (2017). Causal events enter awareness faster than non-causal events. PeerJ, 5, e2932. https://doi.org/10.7717/peerj.2932

Oakes, L. M. (1994). Development of infants' use of continuity cues in their perception of causality. Developmental Psychology, 30, 869–879. https://doi.org/10.1037/0012-1649.30.6.869

Oakes, L. M., & Cohen, L. B. (1990). Infant perception of a causal event. Cognitive Development, 5(2), 193–207. https://doi.org/10.1016/0885-2014(90)90026-P

O'Connell, S., & Dunbar, R. I. M. (2005). The perception of causality in chimpanzees (*Pan spp.*). Animal Cognition, 8(1), 60–66. https://doi.org/10.1007/s10071-004-0231-1

Penn, D. C., & Povinelli, D. J. (2007). Causal cognition in human and nonhuman animals: A comparative, critical review. Annual Review of Psychology, 58(1), 97–118. https://doi.org/10.1146/annurev.psych.58.110405.085555

R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/.

Radvansky, G. A., & Zacks, J. M. (2011). Event perception. WIREs Cognitive Science, 2(6), 608–620. https://doi.org/10.1002/wcs.133

Rakison, D. H., & Krogh, L. (2012). Does causal action facilitate causal perception in infants younger than 6 months of age? Developmental Science, 15(1), 43–53. https://doi.org/10.1111/j.1467-7687.2011.01096.x

Rips, L. J. (2011). Causation from perception. Perspectives on Psychological Science, 6(1), 77–97. https://doi.org/10.1177/1745691610393525

Rochat, M. J., Serra, E., Fadiga, L., & Gallese, V. (2008). The evolution of social cognition: Goal familiarity shapes monkeys' action understanding. Current Biology, 18(3), 227–232. https://doi.org/10.1016/j.cub.2007.12.021

Rolfs, M., Dambacher, M., & Cavanagh, P. (2013). Visual adaptation of the perception of causality. Current Biology, 23(3), 250–254. https://doi.org/10.1016/j.cub.2012.12.017

Santos, L. R. (2004). 'Core Knowledges': A dissociation between spatiotemporal knowledge and contact-mechanics in a non-human primate? Developmental Science, 7(2), 167–174. https://doi.org/10.1111/j.1467-7687.2004.00335.x

Schloegl, C., & Fischer, J. (2017). Causal reasoning in non-human animals (M. R. Waldmann, Ed.; Vol. 1). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199399550.013.36

Schlottmann, A., & Shanks, D. R. (1992). Evidence for a distinction between judged and perceived causality. The Quarterly Journal of Experimental Psychology, 44(2), 321–342. https://doi.org/10.1080/02724989243000055

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. Trends in Cognitive Sciences, 4(8), 299–309. https://doi.org/10.1016/S1364-6613(00)01506-0

Strickland, B. (2017). Language reflects "core" cognition: A new theory about the origin of cross-linguistic regularities. Cognitive Science, 41(1), 70–101. https://doi.org/10.1111/cogs.12332

Völter, C. J., & Call, J. (2017). Causal and inferential reasoning in animals. In APA handbook of comparative psychology: Perception, learning, and cognition, Vol. 2 (pp. 643–671). American Psychological Association. https://doi.org/10.1037/0000012-029

Völter, C. J., & Huber, L. (2021). Dogs' looking times and pupil dilation response reveal expectations about contact causality. Biology Letters, 17(12), 20210465. https://doi.org/10.1098/rsbl.2021.0465

Von Fersen, L., & Lea, S. E. G. (1990). Category discrimination by pigeons using five polymorphous features. Journal of the Experimental Analysis of Behavior, 54(2), 69–84. https://doi.org/10.1901/jeab.1990.54-69

Wilson, V. A. D., Zuberbühler, K., & Bickel, B. (2022). The evolutionary origins of syntax: Event cognition in nonhuman primates. Science Advances, 8(25), eabn8464. https://doi.org/10.1126/sciadv.abn8464

Wood, J. N., Glynn, D. D., & Hauser, M. D. (2008). Rhesus monkeys' understanding of actions and goals. Social Neuroscience, 3(1), 60–68. https://doi.org/10.1080/17470910701563442

Young, M. E., Beckmann, J. S., & Wasserman, E. A. (2006). Pigeons's discrimination of Michotte's launching effect. Journal of the Experimental Analysis of Behavior, 86(2), 223–237. https://doi.org/10.1901/jeab.2006.60-05

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. Psychological Bulletin, 127(1), 3–21.

Supporting information

S1 Table. Baboon details. The following table shows details on the baboons participating in our experiment. Part refers to the first and second period in which we subjected the participants to the Michottean stimuli (within-subjects design). Red cells indicate that the participant did not reach the learning criteria. Green cells indicate that the participant reached the learning criteria and the number refers to the number of blocks of 60 trials needed to.

Name	Sex	Date of birth	Experiment						
			Part	Phase 1a	Phase 1b	Phase 2	Phase 3		Phase 4
Angele	F	14-05-05	1						
			2						
Arielle	F	25-10-05	1	24	11		Cvs.NC	9	
			2	107	50		Control		
Articho	M	20-10-05	1						
			2						
Atmosphere	F	08-03-98	1						
			2						
Bobo	M	09-08-06	1						
			2						
Dora	F	28-07-08	1	26	21		Control	1	
			2						
Dream	F	28-05-08	1						
			2	40	29		Control	5	
Ewine	F	27-07-09	1	56	1		Control	4	
			2	10	10		Cvs.NC		
Fana	F	20-02-10	1						
			2	33	127		Cvs.NC		
Felipe	M	22-05-10	1						
			2						
Feya	F	05-08-10	1	15					
			2	55	17		Control		
Flute	F	18-09-10	1						
			2						
Harlem	M	16-07-12	1						
			2						
Hermine	F	05-03-12	1						
			2						
Lips	F	24-11-15	1						
			2	166	66		Control		
Lome	M	29-08-15	1						
			2						
Mako	M	24-07-16	1	109	34		Cvs.NC		
			2	99	29		Control		
Mali	F	08-02-16	1	43	6		Cvs.NC		
			2	38	51		Control	4	
Muse	F	21-06-16	1	2	23		Control	10	
			2	17	16		Cvs.NC	4	
Nekke	F	24-10-17	1	53					
			2						
Petoulette	F	14-03-99	1						
			2						
Pipo	M	28-08-99	1						
			2						
Violette	F	20-12-04	1	3	1		Cvs.NC	30	
			2	60	40		Control	23	

S2 Tables. Stimuli details. In the tables below you can find the descriptions of the videos which were used as stimuli. The trajectory and duration in ms of object A is shown in orange and of object B in purple. The shown trajectories are in scale. Object A and B always have the same speed. The spatial gap is always ¼ of the distance from the initial position of A to the end position of B, the temporal gap is always 1 second.

S2A Table. Phase 1: L vs. STG, part 1 Videos 1 and 2.

No.	Event type	Duration (ms) and distance of movements and pause	Total duration (ms)
1	L	0 1200 400	1600
2	STG	1000 200 400	1600

S2B Table. Phase 1: L vs. STG, part 2 Videos 1, 2, 3 and 4.

No.	Event type	Duration (ms) and distance of movements and pause	Total duration (ms)
3	L	0 200 600	800
4	STG	1000 800 400	2200

S2C Table. Phase 2: Generalisation 1 Videos 1, 2, 3 and 4 as baseline trials and 5, 6, 7 and 8 as probe trials.

No.	Event type	Duration (ms) and distance of movements and pause	Total duration (ms)
5	L	0 1067 533	1600
6	STG	1000 900 300	2200
7	L	0 900 300	1200
8	STG	1000 300 150	1450

S2D Table: Phase 3: L/STG vs. rest Videos 1, 2, 3, 4, 9 and 10.

No.	Event type	Duration (ms) and distance of movements and pause	Total duration (ms)
9	SG	0 800 400	1200
10	TG	1000 1200 400	2600

S2E Table: Phase 4: Generalisation 2 Videos 1, 2, 3, 4, 9 and 10 as baseline trials and 11, 12, 13 and 14 as probe trials.

No.	Event type	Duration (ms) and distance of movements and pause	Total duration (ms)
11	SG	0 1000 500	1500
12	SG	0 225 675	900
13	TG	1000 750 250	2000
14	TG	1000 167 333	1500

S3 Table. Categorisation strategies. We specifically chose the videos to control for irrelevant cues that could lead to correct classification. The following table summarises the cues which can be used for successful categorisation and the cues that can be ruled out because they are controlled for in the videos. The videos which make sure that the cues are ruled out are mentioned between the brackets (see numbered videos in S2 Tables).

	Cues which can be used for successful categorisation	**Ruled out cues**
Phase 1: L vs. STG, part 1	- Causality - Presence / absence of gap (spatial and temporal) - Presence / absence of spatial gap - Presence / absence of temporal gap - Speed - Movement duration object A - Distance trajectory object A - Distance trajectory object B - Initial position of object B - End position of object A - Ratio of movement duration object A to B	- Total duration (1,2) - Duration before object B starts moving (1,2) - Movement duration object B (1,2)
Phase 1: L vs. STG, part 2	- Causality - Presence / absence of gap (spatial and temporal) - Presence / absence of spatial gap - Presence / absence of temporal gap - Ratio of movement duration object A to B	+ - Speed (1,4 & 2,3) - Movement duration object A (2,3) - Distance trajectory object A (2,3) - Movement duration object B (1,2,4) - Distance trajectory object B (1,4) - Initial position object B (1,4) - End position object A (2,3)
Phase 2: Generalisation 1	- Causality - Presence / absence of gap (spatial and temporal) - Presence / absence of spatial gap - Presence / absence of temporal gap	+ - Ratio of movement duration object A to B (5,2,4 & 6,1,3)
Phase 3: L/STG vs. rest	- Causality - Presence / absence of gap (spatial and temporal)	+ - Presence / absence of spatial gap - Presence / absence of temporal gap
Phase 4: Generalisation 2	- Causality - Presence / absence of gap (spatial and temporal)	

3. Article 2

No evidence for agent-patient role attribution in human infants, human adults and Guinea baboons (*Papio papio*)

Floor Meewis[1,2*], Iris Barezzi[1], Marielle Hababou-Bernson[1], Joël Fagot[1,2], Nicolas Claidière[1,2], Isabelle Dautriche[1]

1. Centre de Recherche en Psychologie et Neurosciences, UMR7077, Aix-Marseille University, CNRS, Marseille, France
2. Station de Primatologie-Celphedia, CNRS UAR846, Rousset, France

Abstract

Languages describe 'who is doing what to whom' by distinguishing the event roles of agent (doer) and patient (undergoer), but it is debated whether they result from non-linguistic representations that may already exist in preverbal infants and non-human animals. The phenomenon of causal perception, where the subsequent movements of two objects A and B evoke the impression of A launching B, is a simple depiction of an agent-patient relation. The seminal study by Leslie and Keeble (1987) proposed that infants of six months old may be able to attribute agent and patient roles to such causal displays, after they demonstrated the infants' dishabituation upon seeing a launching event that was reversed. They introduced the idea that a role reversal had taken place upon reversing the direction of the launching event (launcher becoming launchee), but not in a non-causal temporal gap event where the agent and patient roles were not present. The present study tested this hypothesis in three different populations: 6-month-old human infants, human adults and Guinea baboons (*Papio papio*). For the human infants we applied a habituation-dishabituation design and for the human adults and baboons a match-to-sample task. We were unable to replicate the findings of Leslie and Keeble in the human infants. Similarly, we did not find evidence for an effect specific to reversing launching events in human adults and baboons. The present study is in line with two other non-replications in human infants and calls into question the role reversal paradigm for Michottean launches to study event role attribution.

Keywords: event role; causal perception; Michotte; replication; role reversal; event cognition

Introduction

When communicating about the world, people tend to describe *who is doing what to whom*. Most languages achieve this by distinguishing between the doer (agent) and undergoer (patient) of the action, typically as the subject and object of a sentence. This distinction is often highlighted through word order or case marking (Bickel et al., 2015; Dryer, 2013; Napoli & Sutton-Spence, 2014). The origin of these thematic roles – whether they arise from abstract cognitive representations that precede language – remains a frequently debated topic (Dowty, 1991; Ferretti et al., 2001; Fillmore, 1968; Rissman & Majid, 2019; Strickland, 2017; Wilson et al., 2022; Zuberbühler, 2019, 2020, 2022; Zuberbühler & Bickel, 2022).

Insights into whether the mind represents thematic roles in the absence of language come from studies with human adults on non-linguistic tasks. There is now rich evidence that adults are expert in extracting roles from perceptual input alone: they can grasp agent and patient roles from very brief visual scenes (Dobel et al., 2007; Hafri et al., 2013) and do so unconsciously (Hafri et al., 2018). Additionally, it has become clear that adults discriminate agent from patient by displaying differential attentional behaviour for these two roles. Across a wide range of modalities, whether gestures, events or visual narratives, agents attract more visual attention than patients (Cohn & Paczynski, 2013; Isasi-Isasmendi et al., 2023; Webb et al., 2010). Agents are also easier to identify than patients (Dobel et al., 2007; Isasi-Isasmendi et al., 2023; Wilson et al., 2011; but see: Ünal et al., 2024) and are placed first when described with non-linguistic devices, such as gestures (Goldin-Meadow et al., 2008). While these studies make it clear that the adult mind has non-linguistic representations of agent and patient, they cannot tell if language knowledge is necessary for these representations to exist.

Studies with human infants and non-human animals can reveal whether conceptual knowledge about event roles is in place before full command of natural language or even without any command of language at all. In these populations, investigations into role assignment have often been intertwined with research on causal perception. In these studies, participants are presented with 2D visual displays of geometric shapes in which an object A moves towards a stationary object B, and B starts moving as soon as A stops and touches B (see Figure 1A), which human adults perceive as A causing B to move (Michotte, 1963; for review: Scholl & Tremoulet, 2000). This impression of causality is thought to be perceptually processed (Kominsky & Scholl, 2020; Moors et al., 2017; Rolfs et al., 2013). When a temporal gap and/or a spatial gap is inserted in between the movements of the two objects (see Figure 1A), such a causal impression is not evoked. From birth, infants look more towards causal launches than towards non-causal events (Mascalzoni et al., 2013), showing that they differentiate launching from other, similar events. From six months old onwards, infants view direct launching events as categorically distinct from event with a temporal gap and with a spatial gap in between the objects' movements (Cohen & Amsel, 1998; Oakes, 1994). However, even though these studies show a sensitivity to the spatiotemporal properties related to causality and simultaneously suggest sensitivity to causality at this early age, they do not provide

evidence that infants can represent the agent and patient roles in such displays. Similarly, non-human animals can discriminate between different Michottean events (newly-hatched chicks (*Gallus gallus*): Mascalzoni et al., 2010; Guinea baboons (*Papio papio*): Meewis et al., 2024; pet dogs (*Canis familiaris*): Völter & Huber, 2021; pigeons (*Columba livia*): Young et al., 2006) and some may even perceive similar types of events as causal (chimpanzees (*Pan troglodytes*): Matsuno & Tomonaga, 2017). However, whether non-humans also attribute agent and patient roles to such events is unknown. A recent study that tested whether different ape species have a bias towards agents in videos depicting real life events, showed variable results that were largely dependent on animacy cues (Brocard et al., 2024). In sum, while the perception of causality has been examined through Michottean launches in human adults, infants, and non-human animals, merely being sensitive to causality or its spatiotemporal properties does not indicate whether agent and patient roles are recognized.

In infancy, the capacity to represent agent and patient roles has been argued to be present at six months of age (Leslie & Keeble, 1987). In this study, Leslie and Keeble (1987) aimed to test whether Michottean launching events are perceived holistically or as having an internal structure. The authors habituated one group of infants to a launching event (A launching B) and another group to a similar event with a temporal gap (A moves, stays still, then B moves), with both events involving movement from left to right. In the test phase, infants were shown the same events but in reverse temporal order: B started to move first, followed by A, with motion from the right to the left side of the screen (see Figure 1B). Both groups of infants were thus habituated to a left to right movement and then tested on right to left movement. The authors predicted that if the infants perceive the events as causal, they may find the reversal of the launching event most surprising, as it involved a change in spatiotemporal properties as well as a change in causal direction. Indeed, the reversal of the launching event resulted in more surprise for 6-month-old infants than the reversal of the temporal gap event. Whilst confirming that the infants treat the launching event as different from the temporal gap event at this age, the authors additionally interpreted the stronger dishabituation for the launching event as being caused by a role reversal. By reversing the launch, the launcher (agent) had become the launchee (patient) and the other way around, whereas such a role reversal does not occur when reversing the temporal gap event, since this event does not involve agent and patient roles. Subsequently, this study has been widely cited as evidence of sensitivity to agent-patient roles in 6-month-old infants (Kominsky & Carey, 2024; Mascalzoni et al., 2013; Muentener & Carey, 2010; Ünal, Ji, et al., 2021). However, it is important to note that this conclusion was developed to explain the observed pattern of results, rather than being directly validated by the findings. Critically, a direct replication study conducted by Bélanger and Desrochers (2001) failed to demonstrate longer looking for the reversal of the launching event compared to the temporal gap event. However, they did show that the surprisal for the reversed launching event was greater than an additional condition in which they reversed a spatial gap event (another type of non-causal event), which is why this study is often cited as a replication of Leslie and Keeble's effect (e.g. Kominsky & Carey, 2024; Muentener & Bonawitz, 2017) even though the original comparison

(reversal of launching vs. reversal of temporal gap event) did not replicate. Another recent study applied the role reversal method to asymmetrical Michottean events in 8-month-old infants (Kominsky & Carey, 2024). Such an asymmetrical launching event, in which the speed of A is three times larger than the speed of B, is not perceived as different from a symmetrical launching event (Kominsky et al., 2017). During the test condition, the infants who were habituated to asymmetrical Michottean events, did not look longer to the reversal of the launching compared to the temporal gap event (Kominsky & Carey, 2024), indicating that reversing these highly similar events again does not lead to the same findings as in Leslie and Keeble's study. In sum, while it is widely accepted that 6-month-old infants represent agent and patient roles while viewing Michottean events, the evidence supporting this conclusion is limited.

Beyond the typical Michottean events, research on whether infants represent agent-patient roles has frequently employed role-reversal methods or similar approaches. However, these studies do not allow for definitive conclusions about event role attribution at six months of age, because they either lack appropriate controls (such as a non-causal condition: Golinkoff, 1975; Rochat et al., 2004) or because they show results that are not expected if infants below one year of age represent agent-patient roles. For example, one study investigated 10-to-12-month-old infants' reaction upon replacing the toys portraying object A and B in a launching and in a temporal gap event. The infants showed dishabituation for a change in object A and not B regardless of whether it was a launching or a temporal gap event (Cohen & Oakes, 1993), thus not supporting the idea that infants assign agent and patient roles to a launching event. Another study suggests that this ability may perhaps appear later in life. After habituating infants to a 2D animated launching event, 14-month-olds (but not 10-month-olds) were surprised to see a test event in which the patient engaged in a self-propelled motion but not the agent, which was not the case for the objects for the temporal gap event (Cicchino et al., 2011). This study shows that the infants made inferences about the abilities of the objects based on the role they fulfilled in the events, but only at 14 months. In short, since the seminal study of Leslie and Keeble (1987), there has been little evidence to support the common belief that 6-month-old infants can represent agent and patient roles in causal displays.

The present study thus aimed to replicate and extend the seminal study by Leslie and Keeble (1987). This is important for several reasons:

1) Leslie and Keeble (1987) is widely cited as evidence for early representations of the thematic roles of agent and patient.

2) The theory posited by the original study was developed post-hoc as an interpretation of the observed results, and therefore requires confirmatory hypothesis testing.

3) Subsequent studies by Bélanger and Desrochers (2001) and Kominsky and Carey (2024) failed to replicate the original findings using the same or highly similar events.

4) These failed replications raise questions about the robustness of the role reversal method.

In this study, we tested agent and patient role attribution to Michottean launching events using the role reversal method. We extended the original study in two key aspects.

First, we added a control condition. Previous studies have shown that even newborn infants treat launching events differently from temporal gap events (Mascalzoni et al., 2013). Therefore, it is possible that the longer looking times observed by Leslie and Keeble might be due to differences between the two events rather than specifically the role reversal. To address this, we implemented a "mirror control condition" where a reversal of direction occurs whilst the order of movements (roles in the launch) remains the same (see Figure 1C). This aims to determine whether the role reversal is indeed responsible for the previously reported longer looking times. Second, we tested not only 6-month-old infants, as in the original study, but also human adults and Guinea baboons (*Papio papio*). Studying human adults for whom it is well-established that they view Michottean events in terms of causality and quickly attribute roles to actors in interactions, will further validate the role reversal method for studying agent and patient role attribution. Additionally, findings in non-humans could provide insights into whether event role attribution is shared with non-human species as has been recently proposed (Wilson et al., 2022).

For the infants, we implemented a habituation study replicating Leslie and Keeble's original study. Six-month-old infants were shown a launching event or a temporal gap event during habituation and were subsequently tested on the reversed and mirrored control versions during the test phase. For the human adults and the baboons, we used a match-to-sample classification task, which was designed to test whether the participants would perceive the Michottean launches in terms of causality (Meewis et al., 2024). During this task, the participants were presented with a video of one of the Michottean events which they had to classify into one of two categories: launching event or spatial and temporal gap event. The two categories were represented by two distinct response buttons and during training participants had to learn the appropriate response through trial and error. After learning the categories, the participants went into the test phase in which they were presented with the reversed and mirrored control versions of the training events.

Since the assignment of agent-patient roles appears to be a cognitive mechanism that operates rapidly and automatically (e.g. Hafri et al., 2013, 2018) with its origin in core cognition (Strickland, 2017) and/or beyond the human lineage (Wilson et al., 2022), we preregistered the expectation that the reversal of the causal launch would be perceived as different from the reversal of the non-causal event and different from the mirrored versions of the causal and non-causal events as follows from Leslie and Keeble's proposed theory (Leslie & Keeble, 1987). For the infants, we used looking time as measure and expected a longer looking time for the reversal of the causal event. For the human adults and baboons, we expected more classification mistakes for the reversal of the causal event, as the role reversal would be seen as most different from the others.

Methods

The methods presented in this article have been preregistered. The preregistration for the infant study can be found here and for the human adults and baboons here, in the

paragraphs called “reversed”. All materials have been grouped together in the repository of the infant preregistration and can thus be found here.

Ethics statement

Human infants

The study on infants has received approval from the Comité De Protection Des Personnes Sud Meditérranee III (#2021.05.11quatre_21.05081.210511). Parents signed an informed consent for their child’s participation. They received compensation for travel expenses and were gifted an “infant researcher diploma” and a baby body suit.

Baboons

The study on baboons was carried out in accordance with French and EU standards and received approval from the French Ministère de l’Education Nationale et de la Recherche (#APAFIS-2717-2015111708173794-V3). Procedures used in the present study were also consistent with the guidelines of the Association for the Study of Animal Behaviour.

Human adults

The study on human adults has received approval from the CEEI of INSERM (#20-733). We provided information about the task to the human participants and they consented to participate voluntarily in the study after having read the instructions.

Participants

Human infants

Data of 53 healthy full-term infants (31 female; mean age: 28.0 ± 0.3(SEM) weeks old, age range: 24-34 weeks) were used for the analyses; 27 were habituated to the launching event (of which 13 were tested with mirrored first and 14 with reversed first) and 26 to the temporal gap event (of which 13 were tested with mirrored first and 13 with reversed first). A total 68 infants were tested, but 15 infants were excluded because they were preterm (4), did not habituate in under 18 trials (7) or were fussy during the experiment (4). The sample size was chosen based on Leslie and Keeble’s original study which tested 12 infants per condition and the recommendations for habituation-dishabituation designs by Oakes (2017).

Baboons

We analysed data of 11 Guinea baboons (*Papio papio*) (10 female; mean age: 11.1 ± 1.3(SEM) years old, age range: 5.8-17.4 years) of which 6 participated two times (see below). The task was made available to a total of 23 individuals who were housed at the Station de Primatologie in Rousset-sur-Arc (France) in two groups of eighteen and five individuals. The baboons were tested with the automatic learning device for monkeys (ALDM) developed by (Fagot & Paleressompoulle, 2009). With this system, the baboons had ad libitum access to testing booths with touch screens where they participated voluntarily in computerised tasks employing an operant conditioning method. We were unable to use data from 12 baboons because they did not participate in the task and/or

did not meet the learning criteria. The data we present here are part of a series of experiments with Michottean events that took place in May and July 2022, of which other parts have been described elsewhere (Meewis et al., 2024).

Human adults

We analysed the data from 39 participants (gender: 21 female, 18 male; mean age = 27.5 ± 1.4(SEM) years old, age range = 20-59 years) recruited through Prolific. We chose participants who reported to be fluent in English, such that they could understand the instructions. We excluded one additional participant because they did not complete the experiment in one go but paused for several hours in the middle. Data collection took place at the end of May 2023 and this experiment was also part of a series with Michottean events, of which other parts have been described elsewhere (Meewis et al., 2024). Participants were paid £5 for their time.

Stimuli

Stimuli types

The stimuli used in this study are based on Michotte's launching paradigm (Michotte, 1946/1963). These are simple 2D animations in which an object A moves first from one side to the middle of the screen and an object B moves second from the middle to the other side of the screen. In the *launching event (L)*, A moves towards B, which immediately starts moving in the same direction with the same speed upon being touched by A, see Figure 1A. Contrasting the L event, we used either a *temporal gap event (TG)* in which A moves towards B until they touch, then they both stay still, and B starts moving in the same direction with the same speed as A did, or *a spatial and temporal gap event (STG)* in which A moves towards B, stops before reaching B, they both stay still, after which B starts moving in the same direction with the same speed as A did, see Figure 1A. Since L, but not TG and STG, is perceived as causal by human adults (Scholl & Tremoulet, 2000), only the L has an agent-patient structure, and the TG and STG have two separately moving agents (Leslie & Keeble, 1987; Mascalzoni et al., 2010).

In the test versions of the events, the movements of the object are reversed (starting from the opposite side of the screen). In the *reversed* event, besides the directional change, first object B moves and then object A, see Figure 1B. In the *mirror control* version, the directional change takes place, but the object A is still the first to move followed by object B, see Figure 1C. Importantly, a role reversal takes place only in the reversed version of the L event: when B launches A, the launcher has become the launchee, thus reversing the original agent-patient roles. In all the other conditions, either the agent-patient structure remains (mirror control L event) or there are no agent-patient roles to be reversed (reversed TG/STG event and mirror control TG/STG event).

Infant stimuli

For the infants, we recreated the stimuli used in Leslie and Keeble (1987): an L event and a TG event. The direction of movement was from left to right during habituation (and from right to left during test). Both objects were square-shaped, object A was red and

object B green. The background was black. Leslie and Keeble reported using objects of 3.8 cm in width and 3.1 in height and a travelling distance for both of the objects of 18.1 cm on a screen of 44x33 cm. We kept the same ratio of stimuli to background and displayed them on a 27'' screen. The different durations of the two videos were balanced by adding a few extra still frames at the beginning and end of the direct launch video. The total durations and durations of separate movements remained the same as in the original study (see Supplementary Table 1). All videos were made using the animation features of Microsoft PowerPoint and subsequently exported in mp4 format. We used a resolution of 2880x2160 pixels and a framerate of 60 frames/s (24 frames/s in Leslie and Keeble).

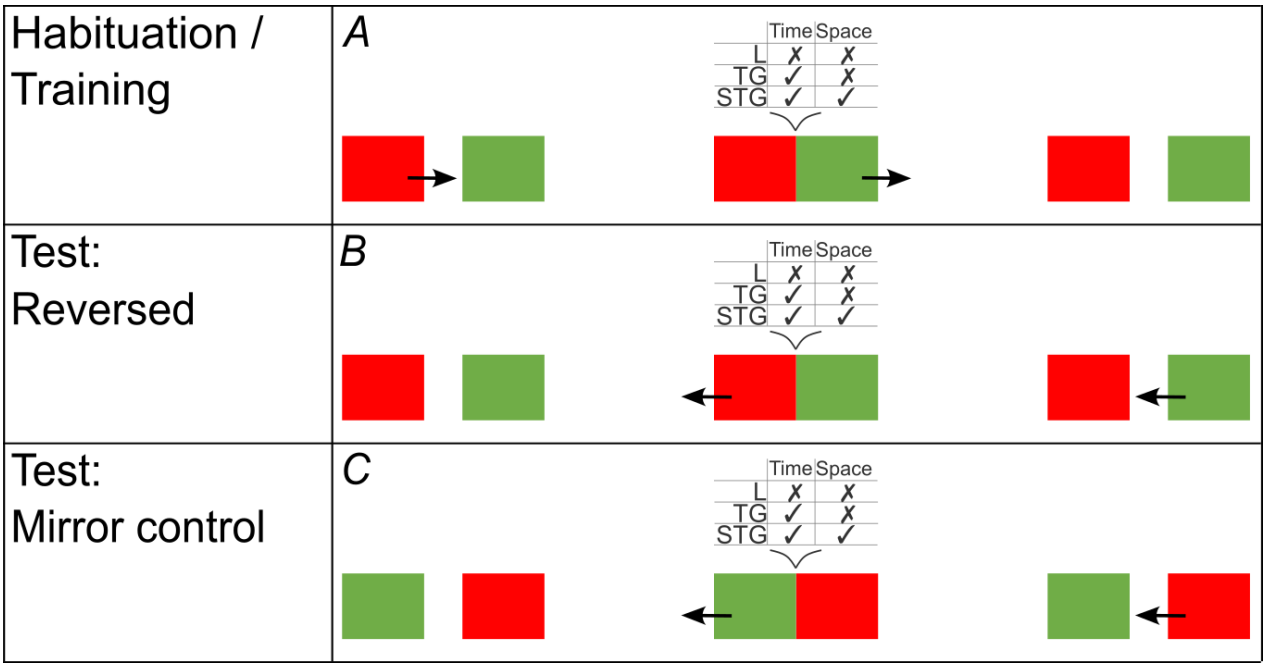


Figure 1: Schematical examples of the Michottean events used as stimuli. During a launching event (L), there is no temporal gap (Time = ×) and no spatial gap (Space = ×) added between the movements of the objects, resulting in an event that is perceived as causal by human adults. During a temporal gap event (TG), there is a temporal gap (Time = ✓), but no spatial gap (Space = ×) added between the movements of the objects, resulting in display perceived as non-causal. During a spatial and temporal gap event (STG), there is temporal gap (Time = ✓) and a spatial gap (Space = ✓) added between the movements of the two objects, which is perceived as non-causal relation. A: Events used in the habituation phase for the human infants and for training for the baboons and human adults. B: Reversed events used in the test phase. C: Mirror control events used in the test phase.

Baboon and human adult stimuli

We used the same stimuli for the baboons and the human adults: L and STG. We decided to use STG instead of TG events to maximize the difference from the L event, in order to facilitate discrimination training for the baboons. We created two stimulus sets. In the first set object A was an orange circle and object B a purple circle, with a movement direction from left to right during training (and right to left during test). In the second

stimulus set, object A was a blue square and object B a yellow square as with a movement direction from right to left during training and from left to right during test. Stimuli were presented on a black background. The two sets used the same durations (see Supplementary Table 2). All videos were made using the animation features of Microsoft PowerPoint and subsequently exported in mp4 format. We used videos with a resolution of 944x720 pixels and a framerate of 30 frames/s.

Procedure and design

Human infants

We used a habituation-dishabituation design just like Leslie and Keeble's (1987). Half of the participants was habituated to the L event, and the other half of the participants was habituated to a TG event. Following habituation, two test trials were displayed, one with the reversed and one with the mirror control version of the habituation video, in random order, see Figure 1.

During stimulus presentation, the infant sat on their parent's lap, facing a screen. The caregiver wore opaque glasses or averted their eyes to avoid influencing the infant. During the experiment, the experimenter was situated outside the view of the infant.

The stimuli were presented with the help of Habit2, Version 2.2.1 (Oakes et al., 2019). During habituation, a loop of the event (L or TG) was played. The habituation criterion was put to a decrease of 50% of looking time based on the sliding window average of three consecutive trials compared to the average looking time of the first three trials (Oakes, 2010). The maximum number of trials for habituation was set at eighteen (Leslie & Keeble, 1987). A trial was defined as looking at the screen and ended when the infant looked away for more than a second (Leslie & Keeble, 1987). Looking away for more than a second prompted the appearance of an attention getter (brightly coloured bubbles with sound) on the screen, when looking was re-established the next trial was presented.

During the experiment, the infant was videotaped whilst looking at the videos. The experimenter used the projection of the video to determine whether the infant was looking at the screen. Both during habituation and test, looking times were recorded in real time in Habit2 by pressing a key on the keyboard of the computer. For the participants included in the analysis, the experiment took on average 2.6 minutes.

Baboons and human adults

The human adults and baboons participated in a match-to-sample (MTS) task in which they were trained to discriminate L from STG events and were subsequently tested on the classification of the reversed and mirrored events.

The humans received minimal instructions before the start of the experiment. We explained that they would see videos which they had to classify by pressing one of two buttons. We also told them they would receive feedback in some of the trials (as a green or red screen) and in some other trials no feedback in the form of a purple screen. Lastly, we told them that the task's duration would increase with wrong answers, to give them

an incentive to try to answer correctly. Causality or agent-patient roles were not mentioned.

Videos and response stimuli were presented to the baboons using E-prime (V2) and to the humans using Labvanced (Finger et al., 2017). During a trial, a video played as a sample, which was followed by two distinctive coloured shapes as comparison stimuli of which the participant had to choose one, see Figure 2. For the first stimulus set, these were a yellow leaf and a blue molecule, and for the second stimulus set, these were a cyan chef's hat and a purple swan. The sample video was displayed in the centre of the screen, and the comparison stimuli were displayed for the monkeys at the bottom and for the humans in the middle of the screen, with their right-left location counterbalanced across trials. Which response stimulus was assigned to which event category was counterbalanced across participants.

The human adults and baboons underwent training to discriminate the L from STG before being subjected to the reversed and mirror control events. Training took place by giving feedback to the participants. For the monkeys, we rewarded correct responses with some wheat grains, whereas incorrect responses were punished with a 3s green time-out screen. Touching the screen during the video interrupted the trial and also led to the 3s green time-out screen. The humans received a 1s green screen with the word 'Yes!' to correct responses and were punished with a 3s red screen with the word 'No!' to incorrect responses, see Figure 2.

In the training phase, participants learnt to discriminate one L from one STG video. These two events were matched for their total duration (i.e., 1600 ms). Participants were presented with blocks of randomly ordered trials, such that, within a block, there were half of L trials and half of STG trials. The baboons had blocks of sixty trials and the humans blocks of eight trials in the training phases. To continue to the next phase of the experiment, the baboons had to reach 80% correct classification for both events within the block (60 trials) and the humans had to reach 100% correct classification for all the trials in the block (8 trials). After reaching these pre-set criteria, two extra events were added, one L and one STG. We intended that the participants would classify the events according to the causality of the event (although paying attention to the presence or absence of the gaps was another successful strategy). For both species, each block consisted of 25% of previously learnt L events, 25% of previously learnt STG events, 25% new L events and 25% new STG events. When the baboons reached the criteria of 70% correct on all four events, they continued to the next phase. For the humans we again placed a 100% correct criterion for all of the trials. Next, we tested whether the participants had not just memorised the four training videos and how they associate with the response stimuli, by measuring their capability to generalise to unseen L and STG events (see: Meewis et al., 2024).

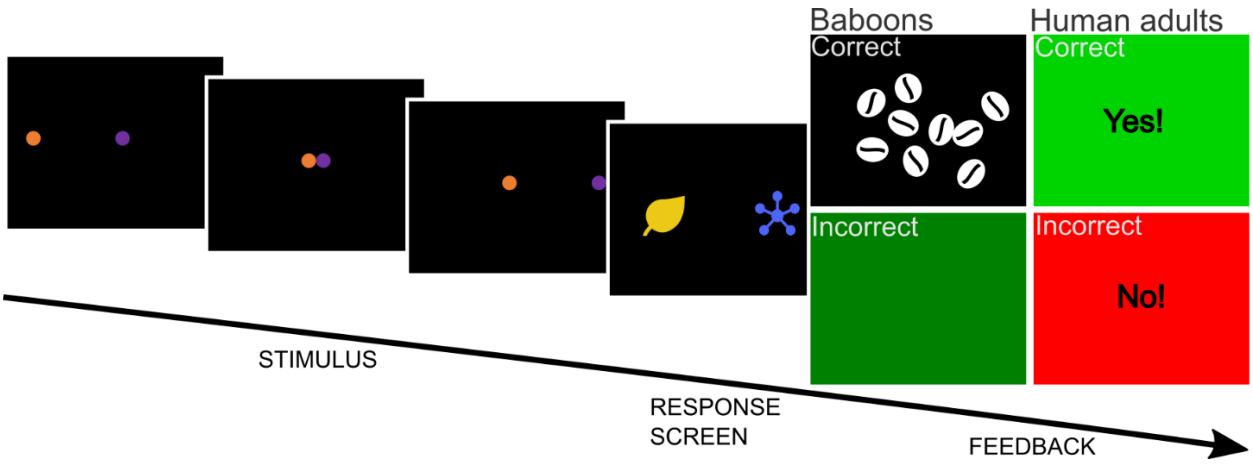


Figure 2: Match-to-sample task for the baboons and the human adults. First, a video is played portraying one of the Michottean events (either a direct launch (L) or a spatial and temporal gap event (STG)). Next, a screen appears with two response buttons. For correct choices the baboons receive a food reward and the humans see a green screen with the word 'Yes!', for incorrect choices the baboons see a green screen for three seconds (time-out) and the humans a red screen with the word 'No!' for three seconds (time-out).

Subsequently, we tested how baboons and humans assigned the reversed and mirror control Michottean events into these previously learnt categories. For the baboons, we used blocks of 60 trials, including 8 probe trials (2 L reversed; 2 L mirror control; 2 STG reversed; 2 STG mirror control) and 52 previously learnt events (26 L and 26 STG), and repeated these blocks sixteen times. The probes were randomly rewarded in 80% of cases regardless of the response given to avoid any learning for these trials. For the humans, we presented three times a block of the four probe events. We did not give the humans feedback on these probe trials, but instead showed a purple screen for two seconds after they made their choice, regardless of the response given, which was explained in the instructions as indicative of no feedback.

Analyses

All analyses were performed using RStudio (R Core Team, 2021).

Human infants

Since looking time data often shows a right-skewed distribution, we checked the distributions and performed logarithmic transformations where needed (Csibra et al., 2016) following our pre-registration.

First, we monitored whether there existed a preference for either of the habituation videos. To be certain that there were no differences related to habituation carrying over in the test phase, we implemented three two-way ANOVAs, one for each habituation property, checking whether there were differences regarding habituation between the habituation events (L or TG) and the test groups (Reversed first or Mirror control first) or an interaction between those. We tested as habituation properties 1) the number of trials

needed to reach the habituation criterion, 2) the mean looking time over the first three habituation trials and 3) the mean looking over the last three habituation trials. We did not detect any differences between groups of infants habituating for the two event types (L or TG) and their following test order (order of reversed and mirror) regarding the number of trials needed for habituation (all $p > .652$), the mean duration of the first three habituation trials (all $p > 0.143$) and the mean duration of the last habituation trials (all $p > .134$).

Since the habituation was equal between the groups, we analysed the test trial looking times (and not recovery time) as preregistered. For the analysis of looking times in the test trials, we used the following general linear mixed model from the lme4-package (Bates et al., 2015): log(looking time) ~ event type (L/TG) * test type (Mirror/Reversed) + order + (1 | participant). P-values for main fixed effects were based on likelihood ratio tests, simple effects were reported from the summary table of the model. Since there was no effect of order of test events on the logarithmically transformed looking times ($\chi 2 = 3.69$, $p = .055$), we removed this term from the model.

We added an extra, not preregistered, test to verify whether the infants dishabituated upon seeing the test events. Here, we calculated the recovery times by subtracting the looking time for the last habituation trial from the first test trials after habituation (disregarding the second test trial) and tested whether these were significantly above zero with one-sided one sample t-tests.

Baboons

To test the response accuracy for mirrored and reversed events, we implemented a generalised linear mixed model with binomial error structure using the lme4-package (Bates et al., 2015): accuracy (0/1) ~ event type (L/STG) * test type (Mirror/Reversed) + stimulus set (first/second) + (1 | participant). We kept stimulus set in the model because it had a significant effect on the response accuracy, such that they were more accurate on the first set during the first period than on the second set during the second period (62.6 ± 3.9(SEM) vs. 57.9 ± 3.1(SEM)% correct; $\chi^2 = 5.41$, $p = .020$). We decided to deviate from our preregistered methods and remove the baseline (accuracy on L baseline was 73.7 ± 2.1(SEM)% correct and accuracy on STG baseline 73.8 ± 2.3(SEM)% correct) from the analysis and only use the test events, since their accuracy on baseline resulted from training and thus only the accuracy on the mirrored and reversed events was of interest.

Human adults

To test the response accuracy for mirrored and reversed events, we implemented a generalised linear mixed model with binomial error structure using the lme4-package (Bates et al., 2015): accuracy (0/1) ~ event type (L/STG) * test type (Mirror/Reversed) + (1 | participant).

Results

Human infants

The infants looked on average 8.5 ± 1.0 (SEM) seconds to the reversed launching event, 12.2 ± 1.6 (SEM) seconds to the mirrored launching event, 10.2 ± 0.9 (SEM) seconds to the reversed temporal gap event and 11.0 ± 1.4 (SEM) seconds to the mirrored temporal gap event. There was no interaction between event type and test type ($\chi^2 = 2.08$, $p = .149$), and no main effect of event type ($\chi^2 = 0.54$, p .461) nor of test type ($\chi^2 = 1.79$, $p = .181$) on the infants' looking times. Thus, contrary to Leslie and Keeble's results, infants did not dishabituate more to the reversed launching event than to reversed temporal gap event, neither to the mirrored events, suggesting no effect of the role reversal.

Using the recovery times (looking time first test trial minus looking time last habituation trial), we found that the infants dishabituated for all of the test events: the reversed launching event (2.0 ± 1.1 s recovery; $t(13) = 1.81$, $p = .047$), the mirrored launching event (6.7 ± 2.0 s recovery; $t(12) = 3.29$, $p = .003$), reversed temporal gap event (3.6 ± 1.4 s recovery; $t(12) = 2.65$, $p = .011$) and the mirrored temporal gap event (6.4 ± 1.6 s recovery; $t(12) = 4.09$, $p < .001$).

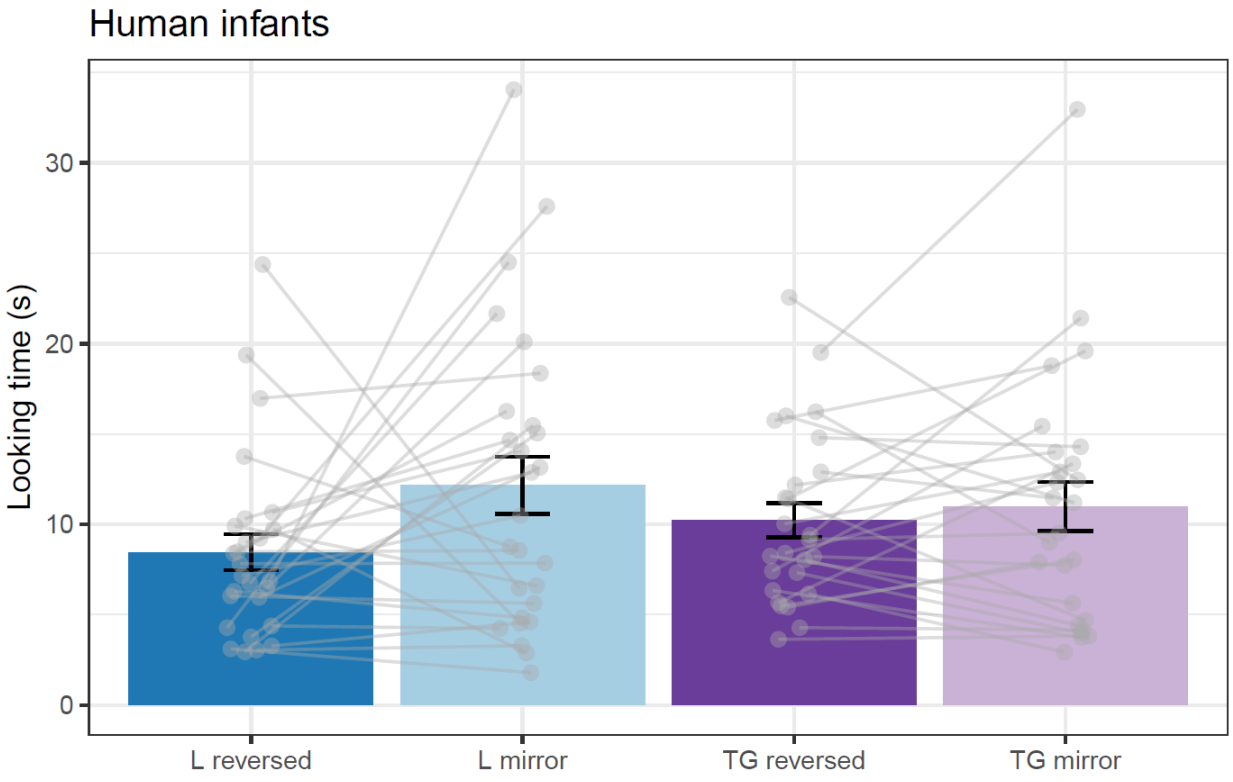


Figure 3: The human infants' looking times during the test events L reversed, L mirror, TG reversed and TG mirror. Error bars reflect standard error of the mean. Grey dots and connecting lines show individual data.

Baboons

For the baboons, we found an interaction between test type and event type (χ^2 = 9.72, p = .002). Post-hoc comparisons showed that the baboons were equally good at classifying the reversed L compared to the reversed STG (60.7 ± 4.2(SEM)% correct vs. 66.4 ± 3.9(SEM)% correct; z = -1.96, p = .050). They were significantly better at classifying the mirrored L compared to the mirrored STG (60.1 ± 4.7(SEM)% correct vs. 52.8 ± 6.1(SEM)% correct; z = 2.45, p = .014). The accuracy scores for the reversed L and mirrored L did not differ from each other (z = 0.19, p = .952), whereas they were less accurate for the mirrored STG compared to the mirrored L (z = 4.57, p < .001). In sum, the baboons did not perform worst on the reversed launching event, suggesting no effect of the role reversal.

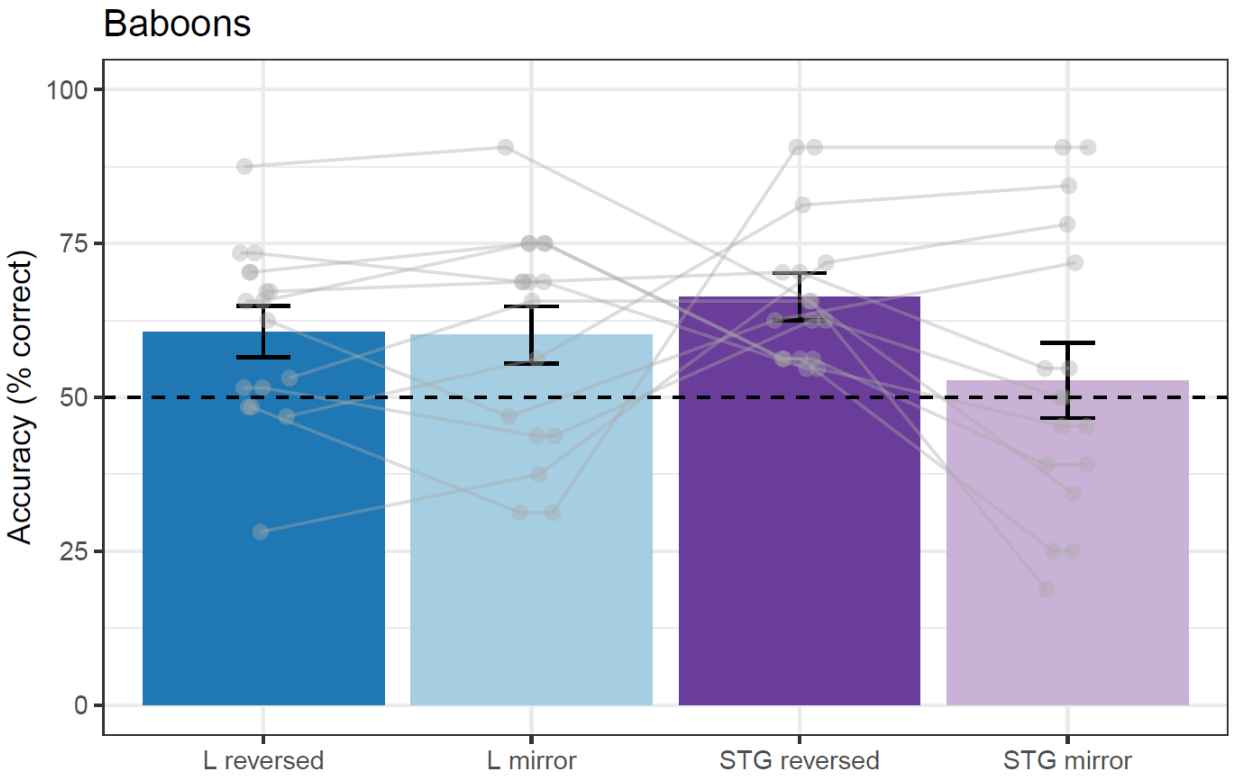


Figure 4: The Guinea baboons' accuracy scores in percentage correct for the events L reversed, L mirror, STG reversed and STG mirror. Error bars reflect standard error of the mean. Grey dots and connecting lines show individual data. The black dotted line shows the chance level of 50% correct.

Human adults

The model on the human responses did not reveal an interaction effect between test type and event type on accuracy (χ^2 = 2.12, p = .145). There was a main effect of test type indicating that the human adults performed better at classifying the mirrored than the reversed events (91.0 ± 2.4(SEM) vs. 73.9 ± 4.6(SEM)% correct; χ^2 =36.05, p < .001). There

was no effect of event type, showing that the human adults performed equally on the L and the STG events (83.3 ± 3.6(SEM) vs. 81.6 ± 4.0(SEM)% correct; χ^2 = 0.39, p = .53). Hence, the human adults did not perform worst on the reversed launching event, suggesting no effect of the role reversal.

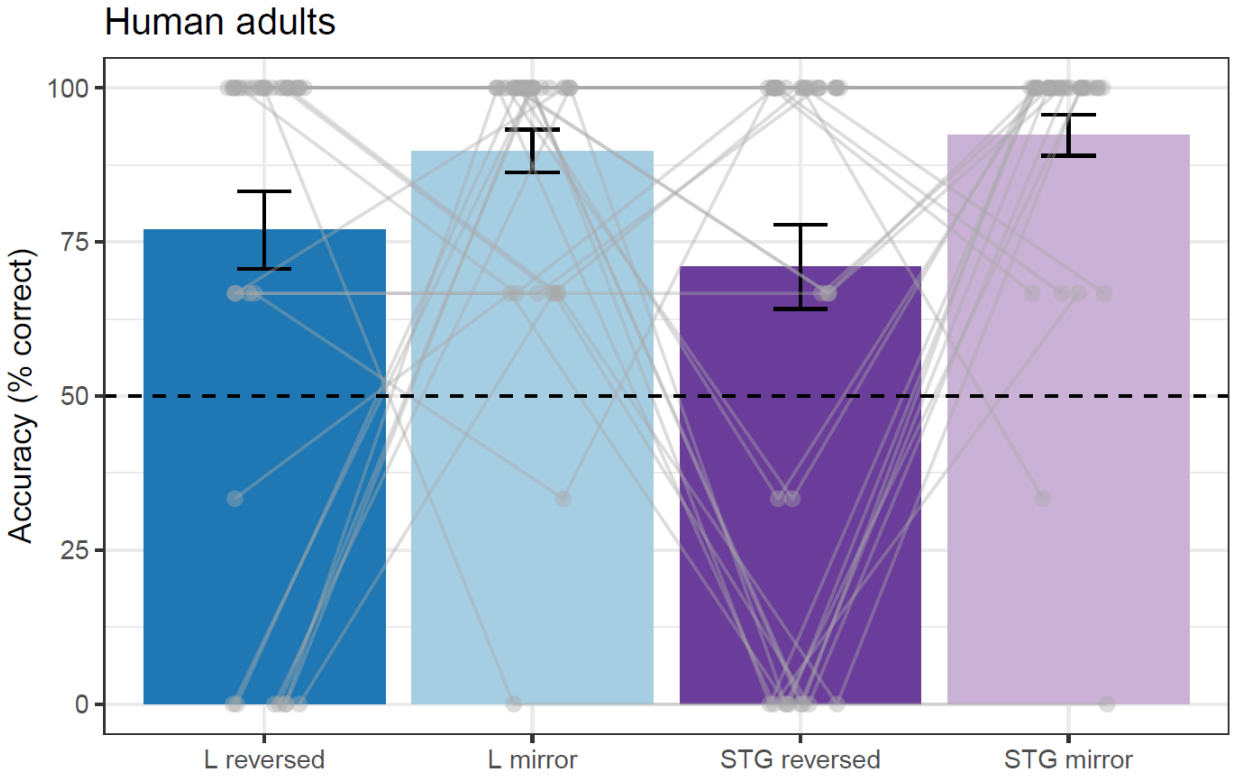


Figure 5: The human adults' accuracy scores in percentage correct for the events L reversed, L mirror, STG reversed and STG mirror. Error bars reflect standard error of the mean. Grey dots and connecting lines show individual data. The black dotted line shows the chance level of 50% correct.

Discussion

This study showed no evidence of agent and patient role attribution to Michottean launching events by means of the role reversal method in three different populations: human infants, Guinea baboons and human adults. Importantly, we did not replicate the findings by Leslie and Keeble (1987), after which the notion that preverbal infants of six months represent agent and patient began gaining momentum. Our results thus cast doubt on the robustness of the role reversal paradigm and/or the assumed ability of 6-months-olds to assign agent-patient roles to causal displays.

For infants, our results are in contrast with Leslie and Keeble's study, but in line with two other, more recent, non-replications (Bélanger & Desrochers, 2001; Kominsky & Carey, 2024). One may argue that our study employed a slightly different procedure

(inclusion of mirror control condition leading to two test trials) which could have resulted in these null-findings. However, an additional analysis focusing on the subset of infants that saw the reversed test trial first, exactly copying the method of Leslie and Keeble, shows the same pattern of results (see Supplementary analysis). Altogether, three studies have failed to replicate the results of Leslie and Keeble, calling into question the robustness of the original findings. One possible explanation for this non-replication is that 6-month-old infants may not, in fact, attribute agent and patient roles to Michottean launches. Since causality perception is assumed to be present at this age (e.g. Cohen & Amsel, 1998), this would point to a potentially problematic dichotomy between sensitivity to causality and the ability to represent agent-patient roles. One possibility is that the apparent sensitivity to causality in previous studies actually reflects a sensitivity to continuous spatiotemporal properties rather than to causality itself. Cohen and Amsel (1998) and Oakes (1994) showed that 6 and 7-month-old infants who habituated to a launching event dishabituated upon seeing a temporal gap or a spatial gap event, but infants habituated to a temporal gap event did not dishabituate to a spatial gap event (and vice versa). This result is consistent with the idea that infants group events based on perceived causality, but it is also compatible with infants grouping together events based on constant, sustained movement (both spatial and temporal events as interrupted movements). This alternative interpretation would be in line with infants not attributing roles to launching displays.

A second possibility is that 6-month-old infants are indeed sensitive to causality and do extract agent-patient roles, but the role reversal method with Michottean launches may not reliably capture this ability. Using 2D chasing interactions, another study successfully demonstrated that 6-month-old infants possess a sensitivity to event roles. Similar to human adults, infants prefer to look towards the agent of a chasing interaction (Galazka & Nyström, 2016), indicating that they discriminate between the two event roles at this age. Another recent study showed that 7-month-old infants are able to generalise agent and patient roles after seeing a sequence of different images depicting one person acting on another (Papeo et al., 2024). The possibility that the role-reversal paradigm is ill-suited for testing event role attribution is further supported by our results showing no effect of the role reversal in human adults. Irrespective of causality, the human adults were worse at classifying the reversed compared to the mirrored events, presumably due to low-level features, such that their performance was more disrupted by the two changes in the mirrored control (direction plus colour swap) compared to the single change in the reversed condition (direction). This means that the humans did not perceive a role reversal in the reversed launching event at all, or at least that the role reversal was less salient than the low-level changes. This finding was unexpected given that it is well-established that human adults perceive Michottean events in terms of causality (Scholl & Tremoulet, 2000), a result that we further confirmed for subsequent steps of our task (reported in: Meewis et al., 2024). In those subsequent steps, we demonstrated that human adults learnt faster to categorise two additional non-causal events (one with a spatial gap and one with a temporal gap) together with the previously learnt STG category, forming a distinction based on causality, than with the L event, confirming that human

adults readily extract causality from these events. Similarly, other studies have demonstrated that human adults extract agent-patient roles from various other events (Cohn & Paczynski, 2013; Dobel et al., 2007; Hafri et al., 2013, 2018a; Webb et al., 2010; Wilson et al., 2011). In sum, using the role-reversal method, we found no indication that human adults attribute agent and patient roles to events which they perceive as causal. This finding raises questions about the suitability of the role-reversal method in general.

For the Guinea baboons, we again found no indications that they have extracted the agent and patient roles. Since the baboons were equally accurate to classify the reversed launching event and the reversed spatial and temporal gap event, there is no effect of the hypothesised role reversal. Further experiments following the present ones reported no evidence that baboons perceive the Michottean launches in terms of causality in this task (Meewis et al., 2024). That is, the baboons learnt equally fast to categorise temporal gap events and spatial gap events according to a causal as to a random division, showing that they relied on lower-level features instead of the causal dimension to classify the events. Without perceiving the Michottean event as causal, it is impossible to attribute agent and patient roles. Therefore, we cannot draw firm conclusions about role attribution based on the present findings. Importantly, another experiment from our lab using a different paradigm, has demonstrated that baboons have an attentional preference for the agent in a 2D chasing interaction (Meewis et al., submitted), revealing that, under the right circumstances, Guinea baboons appear to discriminate agent and patient roles. The present null-findings combined with positive findings from a different paradigm lead us to question whether applying role reversals to Michottean launches is an appropriate way to study event role attribution.

It should be noted that the measures used for the three populations were not exactly the same, warranting some caution when comparing the results and criticising the role reversal method. The infant findings are based on looking time, an implicit measure, whereas the adult and baboon findings are based on accuracy, a more explicit measure. Since this same explicit measure was sufficient to show the use of causality in human adults (Meewis et al., 2024), we assumed that it could also be used for the present question as well. However, to ensure that the role reversal task itself results in an inability to demonstrate agent and patient role attribution, future studies should corroborate this using an implicit measure in human adults.

In sum, the findings of our experiments raise doubt about the ability of 6-month-old infants to attribute agent and patient roles to Michottean launches and/or the effectiveness of the role reversal method to address this question. Given the lack of effect of the role reversal across all our populations, including human adults, it seems more likely that there is an issue with the role reversal method itself. Consequently, the use of the role reversal method for testing the agent and patient role attribution in Michottean launches needs reconsideration. Such reconsideration may have wide-ranging consequences as Leslie and Keeble's study is a cornerstone in the field of cognitive development and has inspired many subsequent studies over a period of more than thirty years. It thus necessitates a revaluation of a broad body of work applying role reversals to study agent and patient role attributions.

From this study alone, we cannot draw definitive conclusions about the existence of cognitive representations of agent and patient in the absence of language. This thus remains a crucial question with potentially far-reaching implications for the development and evolution of language. Verification of previous, influential studies is paramount to build a solid research foundation to bring us closer to understanding how causal events and their event roles are processed.

Acknowledgements

This work was supported by the Agence Nationale de la Recherche (ANR-20-CE28-0005) and the Institute for Language, Communication and the Brain (ILCB) through the Agence Nationale de la Recherche (ANR-16-CONV-0002). We thank Julie Gullstrand, Sebastien Barniaud and the staff at the Primatology Station for their help with the experiments with the baboons.

References

Bélanger, N. D., & Desrochers, S. (2001). Can 6-month-old infants process causality in different types of causal events? British Journal of Developmental Psychology, 19(1), 11–21. https://doi.org/10.1348/026151001165930

Bickel, B., Witzlack-Makarevich, A., Choudhary, K. K., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2015). The neurophysiology of language processing shapes the evolution of grammar: Evidence from case marking. PLOS ONE, 10(8), e0132819. https://doi.org/10.1371/journal.pone.0132819

Brocard, S., Wilson, V. A. D., Berton, C., Zuberbühler, K., & Bickel, B. (2024). A universal preference for animate agents in hominids. iScience, 27(6). https://doi.org/10.1016/j.isci.2024.109996

Cicchino, J. B., Aslin, R. N., & Rakison, D. H. (2011). Correspondences between what infants see and know about causal and self-propelled motion. Cognition, 118(2), 171–192. https://doi.org/10.1016/j.cognition.2010.11.005

Cohen, L. B., & Amsel, G. (1998). Precursors to infants' perception of the causality of a simple event. Infant Behavior and Development, 21(4), 713–731. https://doi.org/10.1016/s0163-6383(98)90040-6

Cohen, L. B., & Oakes, L. M. (1993). How infants perceive a simple causal event. Developmental Psychology, 29(3), 421–433. https://doi.org/10.1037/0012-1649.29.3.421

Cohn, N., & Paczynski, M. (2013). Prediction, events, and the advantage of Agents: The processing of semantic roles in visual narrative. Cognitive Psychology, 67(3), 73–97. https://doi.org/10.1016/j.cogpsych.2013.07.002

Csibra, G., Hernik, M., Mascaro, O., Tatone, D., & Lengyel, M. (2016). Statistical treatment of looking-time data. Developmental Psychology, 52(4), 521–536. http://dx.doi.org/10.1037/dev0000083

Dobel, C., Gumnior, H., Bölte, J., & Zwitserlood, P. (2007). Describing scenes hardly seen. Acta Psychologica, 125(2), 129–143. https://doi.org/10.1016/j.actpsy.2006.07.004

Dowty, D. (1991). Thematic Proto-Roles and Argument Selection. Language, 67(3), 547. https://doi.org/10.2307/415037

Dryer, M. S. (2013). Order of Subject, Object and Verb. In M. S. Dryer & M. Haspelmath (Eds.), The World Atlas of Language Structures Online. Zenodo. https://doi.org/10.5281/zenodo.7385533

Fagot, J., & Paleressompoulle, D. (2009). Automatic testing of cognitive performance in baboons maintained in social groups. Behavior Research Methods, 41(2), 396–404. https://doi.org/10.3758/BRM.41.2.396

Ferretti, T. R., McRae, K., & Hatherell, A. (2001). Integrating Verbs, Situation Schemas, and Thematic Role Concepts. Journal of Memory and Language, 44(4), 516–547. https://doi.org/10.1006/jmla.2000.2728

Fillmore, C. J. (1968). Lexical Entries for Verbs. Foundations of Language, 4(4), 373–393.

Finger, H., Goeke, C., Diekamp, D., Standvoß, K., & König, P. (2017). LabVanced: A unified JavaScript framework for online studies. International Conference on Computational Social Science.

Galazka, M., & Nyström, P. (2016). Infants' preference for individual agents within chasing interactions. Journal of Experimental Child Psychology, 147, 53–70. https://doi.org/10.1016/j.jecp.2016.02.010

Goldin-Meadow, S., So, W. C., Özyürek, A., & Mylander, C. (2008). The natural order of events: How speakers of different languages represent events nonverbally. Proceedings of the National Academy of Sciences, 105(27), 9163–9168. https://doi.org/10.1073/pnas.0710060105

Golinkoff, R. M. (1975). Semantic development in infants: The concepts of agent and recipient. Merrill-Palmer Quarterly of Behavior and Development, 21(3), 181–193.

Hafri, A., Papafragou, A., & Trueswell, J. C. (2013). Getting the gist of events: Recognition of two-participant actions from brief displays. Journal of Experimental Psychology: General, 142(3), 880–905. https://doi.org/10.1037/a0030045.Getting

Hafri, A., Trueswell, J. C., & Strickland, B. (2018). Encoding of event roles from visual scenes is rapid, spontaneous, and interacts with higher-level visual processing. Cognition, 175(February 2017), 36–52. https://doi.org/10.1016/j.cognition.2018.02.011

Isasi-Isasmendi, A., Andrews, C., Flecken, M., Laka, I., Daum, M. M., Meyer, M., Bickel, B., & Sauppe, S. (2023). The Agent Preference in Visual Event Apprehension. Open Mind, 7, 240–282. https://doi.org/10.1162/opmi_a_00083

Kominsky, J. F., & Carey, S. (2024). Infants' representations of michottean triggering events. Cognition, 250, 105844. https://doi.org/10.1016/j.cognition.2024.105844

Kominsky, J. F., & Scholl, B. J. (2020). Retinotopic adaptation reveals distinct categories of causal perception. Cognition, 203, 104339. https://doi.org/10.1016/j.cognition.2020.104339

Kominsky, J. F., Strickland, B., Wertz, A. E., Elsner, C., Wynn, K., & Keil, F. C. (2017). Categories and Constraints in Causal Perception. Psychological Science, 28(11), 1649–1662. https://doi.org/10.1177/0956797617719930

Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? Cognitive Development, 25, 265–288.

Mascalzoni, E., Regolin, L., & Vallortigara, G. (2010). Innate sensitivity for self-propelled causal agency in newly hatched chicks. Proceedings of the National Academy of Sciences of the United States of America, 107(9), 4483–4485. https://doi.org/10.1073/pnas.0908792107

Mascalzoni, E., Regolin, L., Vallortigara, G., & Simion, F. (2013). The cradle of causal reasoning: Newborns' preference for physical causality. Developmental Science, 16(3), 327–335. https://doi.org/10.1111/desc.12018

Matsuno, T., & Tomonaga, M. (2017). Causal capture effects in chimpanzees (Pan troglodytes). Cognition, 158, 153–164. https://doi.org/10.1016/j.cognition.2016.10.023

Meewis, F., Barezzi, I., Joël, F., Claidière, N., & Dautriche, I. (2024). Preprint: A comparative study of causal perception in Guinea baboons (Papio papio) and human adults. OSF. https://doi.org/10.31219/osf.io/mejwp

Michotte, A. (2017). The Perception of Causality. Routledge. https://doi.org/10.4324/9781315519050

Moors, P., Wagemans, J., & de-Wit, L. (2017). Causal events enter awareness faster than non-causal events. PeerJ, 5, e2932. https://doi.org/10.7717/peerj.2932

Muentener, P., & Bonawitz, E. (2017). The Development of Causal Reasoning. In M. R. Waldmann (Ed.), The Oxford Handbook of Causal Reasoning (p. 0). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199399550.013.40

Muentener, P., & Carey, S. (2010). Infants' causal representations of state change events. Cognitive Psychology, 61(2), 63–86. https://doi.org/10.1016/j.cogpsych.2010.02.001

Napoli, D. J., & Sutton-Spence, R. (2014). Order of the major constituents in sign languages: Implications for all language. Frontiers in Psychology, 5. https://doi.org/10.3389/fpsyg.2014.00376

Oakes, L. M. (1994). Development of infants' use of continuity cues in their perception of causality. Developmental Psychology, 30, 869–879. https://doi.org/10.1037/0012-1649.30.6.869

Oakes, L. M. (2017). Sample size, statistical power, and false conclusions in infant looking-time research. Infancy, 22(4), 436–469. https://doi.org/10.1111/infa.12186

Papeo, L., Vettori, S., Serraille, E., Odin, C., Rostami, F., & Hochmann, J.-R. (2024). Abstract thematic roles in infants' representation of social events. Current Biology. https://doi.org/10.1016/j.cub.2024.07.081

R Core Team. (2021). R: A language and environment for statistical computing . R Foundation for Statistical Computing. https://www.R-project.org/.

Rissman, L., & Majid, A. (2019). Thematic roles: Core knowledge or linguistic construct? Psychonomic Bulletin and Review, 26(6), 1850–1869. https://doi.org/10.3758/s13423-019-01634-5

Rochat, P., Striano, T., & Morgan, R. (2004). Who is doing what to whom? Young infants' developing sense of social causality in animated displays. Perception, 33(3), 355–369. https://doi.org/10.1068/p3389

Rolfs, M., Dambacher, M., & Cavanagh, P. (2013). Visual Adaptation of the Perception of Causality. Current Biology, 23(3), 250–254. https://doi.org/10.1016/j.cub.2012.12.017

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. Trends in Cognitive Sciences, 4(8), 299–309. https://doi.org/10.1016/S1364-6613(00)01506-0

Strickland, B. (2017). Language reflects "core" cognition: A new theory about the origin of cross-linguistic regularities. Cognitive Science, 41(1), 70–101. https://doi.org/10.1111/cogs.12332

Ünal, E., Ji, Y., & Papafragou, A. (2021). From event representation to linguistic meaning. Topics in Cognitive Science, 13(1), 224–242. https://doi.org/10.1111/tops.12475

Ünal, E., Wilson, F., Trueswell, J., & Papafragou, A. (2024). Asymmetries in encoding event roles: Evidence from language and cognition. Cognition, 250, 105868. https://doi.org/10.1016/j.cognition.2024.105868

Völter, C. J., & Huber, L. (2021). Dogs' looking times and pupil dilation response reveal expectations about contact causality. Biology Letters, 17(12), 20210465. https://doi.org/10.1098/rsbl.2021.0465

Webb, A., Knott, A., & MacAskill, M. R. (2010). Eye movements during transitive action observation have sequential structure. Acta Psychologica, 133(1), 51–56. https://doi.org/10.1016/j.actpsy.2009.09.001

Wilson, F., Papafragou, A., Bunger, A., & Trueswell, J. (2011). Rapid extraction of event participants in caused motion events. Proceedings of the Annual Meeting of the Cognitive Science Society, 7.

Wilson, V. A. D., Zuberbühler, K., & Bickel, B. (2022). The evolutionary origins of syntax: Event cognition in nonhuman primates. Science Advances, 8(25), eabn8464. https://doi.org/10.1126/sciadv.abn8464

Young, M. E., Beckmann, J. S., & Wasserman, E. A. (2006). Pigeons's discrimination of Michotte's launching effect. Journal of the Experimental Analysis of Behavior, 86(2), 223–237. https://doi.org/10.1901/jeab.2006.60-05

Zuberbühler, K. (2019). Evolutionary roads to syntax. Animal Behaviour, 151, 259–265. https://doi.org/10.1016/j.anbehav.2019.03.006

Zuberbühler, K. (2020). Syntax and compositionality in animal communication. Philosophical Transactions of the Royal Society B: Biological Sciences, 375(1789), 20190062. https://doi.org/10.1098/rstb.2019.0062

Zuberbühler, K. (2022). Event parsing and the origins of grammar. WIREs Cognitive Science, 13(3). https://doi.org/10.1002/wcs.1587

Zuberbühler, K., & Bickel, B. (2022). Transition to language: From agent perception to event representation. WIREs Cognitive Science, 13(6). https://doi.org/10.1002/wcs.1594

Supplementary information

Supplementary table 1. Durations of the different parts of the visual display for the stimuli used for the infants. They are the same as in Leslie and Keeble, 1987.

	Beginning still	Movement object A	Object A & B still	Movement object B	End still	Total video	Black screen
L	500 ms	1000 ms ½ screen	42 ms	1000 ms ½ screen	500 ms	3042 ms	333 ms
TG	333 ms	1000 ms ½ screen	500 ms	1000 ms ½ screen	333 ms	3167 ms	333 ms

Supplementary table 2: Durations of the different parts of the visual display for the stimuli used for the Guinea baboons and human adults.

	Beginning still	Movement object A	Object A & B still	Movement object B	End still	Total video	Black screen
L	0 ms	1200 ms ¾ screen	1 ms	400 ms ¼ screen	0 ms	1600 ms	0 ms
STG	0 ms	200 ms ¼ screen	1000 ms ¼ screen distance between A&B	400 ms ½ screen	0 ms	1600 ms	0 ms

Supplementary analysis: exact replication Leslie and Keeble

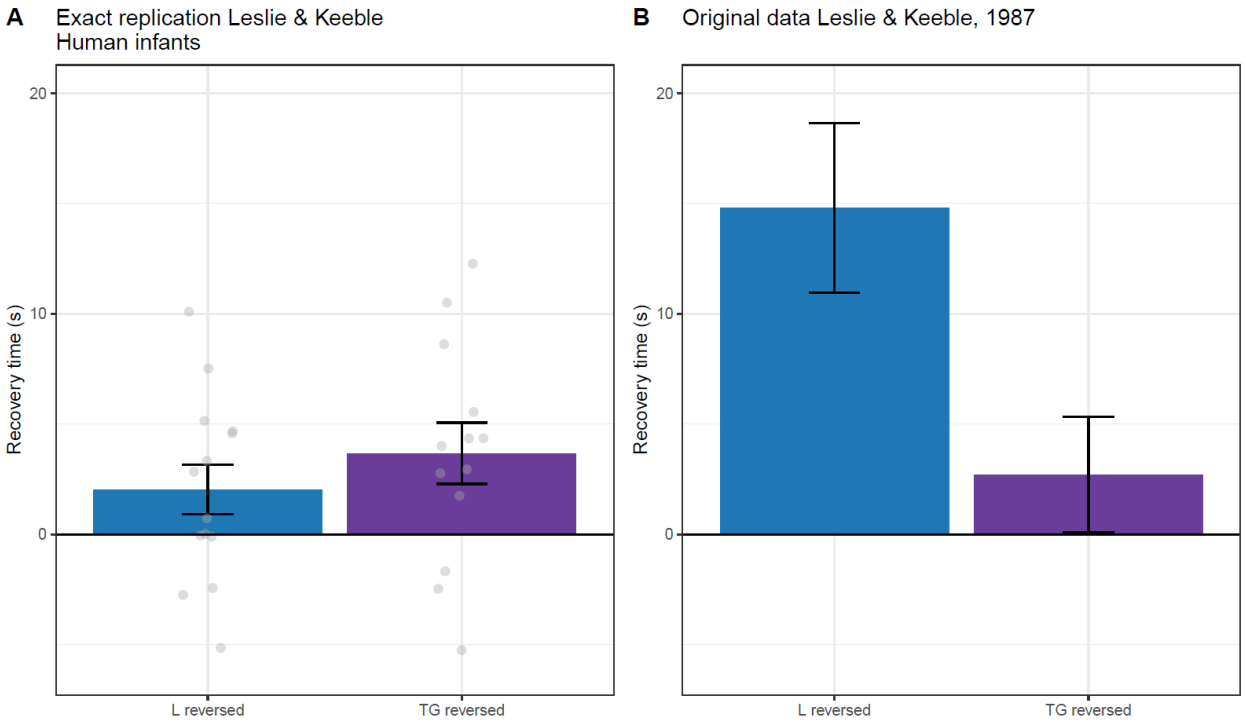
Methods:

From all the tested participants, we selected the participants reflecting the paradigm of Leslie and Keeble exactly. We thus picked those participants who were presented with the reversed version of the habituation trials as first test trial. In other words, we removed all the mirrored trials and the participants who saw the mirrored version before the reversed version. This resulted in 27 infants (14 female; mean age: 28.0 ± 0.5(SEM) weeks old, age range: 24-32 weeks) meeting these criteria, of whom 14 infants habituated to the launching event and 13 infants habituated to the temporal gap event. This sample is comparable to Leslie and Keeble's.

We calculated the recovery looking time as the looking time for the first test trial minus the looking time for the last habituation trial (cf. Leslie & Keeble, 1987). The recovery times of the reversed launching event and reversed temporal gap event were compared with a t-test.

Results:

The recovery looking time of the L group was 2.0 ± 1.1(SEM) s and the of the TG group 3.8 ± 1.4(SEM) s, see Supplementary figure 1A. These looking times were significantly different from zero in the L group ($t(13) = 1.81$, $p = .047$) and in the TG group ($t(12) = 2.65$, $p = .011$), confirming that the infants dishabituated upon seeing the new events. There was however no difference in recovery time between the infants habituated to L and those to TG ($t(23.5) = -0.2$, $p = .368$). The results are different from those reported by Leslie and Keeble, see Supplementary figure 1B.



Supplementary figure 1. The human infants' recovery looking times during the test events L reversed and TG reversed. Error bars reflect standard error of the mean. Grey dots show individual data. A. Data from the present study. B. Original data reproduced from Leslie & Keeble, 1987.

4. Article 3

Agent preference in chasing interactions in Guinea baboons (*Papio papio*): uncovering the roots of subject-object order in language

Floor Meewis[1,2*], Joël Fagot[1,2], Nicolas Claidière[1,2], Isabelle Dautriche[1]

1. Centre de Recherche en Psychologie et Neurosciences, UMR7077, Aix-Marseille University, CNRS, Marseille, France
2. Station de Primatologie-Celphedia, CNRS UAR846, Rousset, France

Abstract

Languages tend to describe 'who is doing what to whom' by placing subjects before objects. This bias for agents is reflected in event cognition: agents capture more attention than patients in human adults and infants. We investigated whether this agent preference is shared with non-human animals. We presented Guinea baboons (*Papio papio*, N = 13) with a change detection paradigm on chasing animations. The baboons were trained to respond to a color change which was applied to either the chaser/agent or the chasee/patient. They were faster to detect a change to the chaser than to the chasee, which cannot be explained by low-level features in our stimuli like the chaser's motion pattern or position. An agent preference may be an evolutionarily old mechanism that is shared between humans and other primates, which could have become externalized in language as a tendency to place the subject first.

Keywords: agent-patient; thematic roles; event cognition; language evolution; linguistic universal; syntax

Introduction

Human languages vary greatly at all levels of description, yet they also share an important number of commonalities (Greenberg, 1963). For instance, in 97% of languages with a dominant word order, the subject (S) precedes the object (O) (Dryer, 2013), e.g., "the girl (S) pushed the boy (O)" in English or "larki-ne (S) larke-ko (O) dhakka diya" in Hindi. These cross-linguistic statistical regularities have been the topic of many enquiries in the cognitive science of language, but their origin remains a topic of ongoing debates (Chomsky, 1957; Culbertson et al., 2020; Evans & Levinson, 2009; MacWhinney, 1977). Statistical universals may be the product of language contact and history (Dunn et al., 2011) and/or reflect evolutionary processes that are independent of cognition (Bybee, 2006, 2009; Gibson et al., 2019; Kirby et al., 2008). Another possibility is that they are the result of cognitive biases and mental representations that pre-exist language (Culbertson & Kirby, 2016; Martin et al., 2024; Strickland, 2017). For instance, the tendency for word orders to prioritize subjects before objects may originate from the way humans naturally process events around them, with a preference for attending to agents (typically mapped to subject positions) before patients (typically mapped to object positions) (Jackendoff, 1999; Ünal, Ji, et al., 2021). Such an agent preference, i.e. a prioritized attention towards agents when observing events, has been recently argued to have deep evolutionary roots beyond human language (Wilson et al., 2022; Zuberbühler, 2019, 2020, 2022; Zuberbühler & Bickel, 2022).

The strong link between linguistic event descriptions and event cognition provides support for this theory (Rissman & Majid, 2019; Ünal, Ji, et al., 2021). While reflected as specific arguments in active transitive sentences, the underlying conceptual structure is thought to be more general: the agent (doer) causes an action that affects the patient (undergoer) (Fillmore, 1968; Jackendoff, 2002). Furthermore, the agent and patient are often characterized by features that are associated with their role, such as the agent's volitional behaviour, awareness, motility and independence, while the patient is defined by its reliance on the agent, undergoing change, and being less dynamic (Dowty, 1991). It follows that attributes such as agency (intentionality and goal-directedness) and animacy (being alive) are applicable to both roles, but are more likely to be associated with the agent role. Even though speakers can use linguistic devices to highlight any argument by placing it first, e.g. the patient in passive structures, placing the agent in subject position appears to be the default (Bickel et al., 2015; Gertner & Fisher, 2012).

Evidence that agent and patient may be represented in an abstract sense, comes from experiments demonstrating the saliency of these roles, and in particular the agent, over a wide range of events, such as pictures, drawings and even animations with 2D-shapes. Not only do human adults spontaneously and unconsciously extract event roles from brief visual events (Hafri et al., 2013, 2018), but agents attract their attention more than patients. While watching events, human adults first direct their gaze toward the agent of an action (Webb et al., 2010). Similarly, in cartoons, agents are looked at longer than patients (Cohn & Paczynski, 2013) and from short displays of drawings, agents are systematically better identified compared to patients, instruments and actions (Dobel et

al., 2007). Furthermore, adults are faster to orient to agents when asked to detect agents than to orient to patients when asked to detect patients (Wilson et al., 2011).

Developmental studies suggest that this agent preference is not the result of linguistic experience with subject-object ordering. 7-month-old infants use postural and positional cues to assign agent and patient roles in an abstract way (Papeo et al., 2024). When presented with 2D chasing interactions, 5-month-old preverbal infants look more towards the chaser than the chasee (Galazka & Nyström, 2016) and thus show an attentional preference for agents similar to human adults (Meyerhoff et al., 2014b). In 9-month-old infants, looking at the chaser evokes neural responses related to social perception, suggesting that the chaser may be perceived as an animate entity (Galazka et al., 2016). This bias towards the agent appears to be reflected in word learning too; 14-month-olds learn more easily a label for a chaser than for a chasee (Yin & Csibra, 2015).

Although an agent preference is thus found under non-linguistic conditions, its evolutionary origins remain a mystery. One possibility is that an agent preference can be found in non-human animals. This would suggest that the structure of event representations has an old evolutionary history and may form the roots of one of the key characteristics of human language, its subject-object word order. Alternatively, an agent preference could be a typically human bias that has evolved concurrently with language (Wilson et al., 2022). The evidence for an agent preference in non-human animals so far is not conclusive. Although non-human animals can detect some cues indicative of animacy (Hauser, 1998; Rosa-Salva et al., 2016), less is known about using the relational aspect of an interaction to identify agents as compared to patients. Great apes (chimpanzees (*Pan troglodytes*), gorillas (*Gorilla gorilla*) and orangutans (*Pongo abelii*)) scan videos of dyadic event scenes similarly to humans with their gaze alternating between the agent and the patient, but in this study an agent preference was not observed in any of the species (including humans) (Wilson et al., 2023). At present, it thus remains unclear whether non-human animals also show an attentional bias for agents as compared to patients.

Here, we investigated the evolutionary origin of the agent preference in Guinea baboons (*Papio papio*). Baboons share a common ancestor with humans that lived 25-30 Mya. They have evolved under similar environmental conditions as humans (Fischer et al., 2019), making them a good model species for language evolution (Fagot et al., 2019), especially with regards to possible roots in social cognition (Seyfarth & Cheney, 2014). To test an agent preference in Guinea baboons, we studied their response to generated animations of chasing interactions in which the agent and the patient were represented by two identical, simple geometrical shapes. Although it is debated whether monkeys attribute mental states to 2D shapes (Schafroth et al., 2021), forms similar to ours have been repeatedly used in previous studies, which revealed that chasing can be discriminated from other motion patterns by human adults (Abdai et al., 2017; Atsumi et al., 2017; Barrett et al., 2005; Hofrichter & Rutherford, 2019; Meyerhoff et al., 2014b; Rochat et al., 1997), human infants (Frankenhuis et al., 2013; Rochat et al., 1997) and several non-human species, including dogs (*Canis familiaris*; Abdai et al., 2017; Abdai & Miklósi, 2022), cats (*Felis silvestris catus*; Abdai & Miklósi, 2022), rhesus macaques

(*Macaca mulatta*, Atsumi et al., 2017), squirrel monkeys (*Saimiri sciureus*; Atsumi & Nagasaka, 2015) and pigeons (*Columba livia*; Goto et al., 2002).

We presented baboons with a change detection task on a touch screen, in which the participants had to respond to a color change to either the agent or the patient of a chasing interaction, see Fig. 1. The rationale of this procedure is that areas that attract attention are prone to faster change detection than less-attended regions (New et al., 2007); a similar method using dot-probe tasks has been successfully applied to non-human primates (van Rooijen et al., 2017). We pre-registered the prediction that if baboons show an agent preference, such that the chaser captures their attention more than the chasee as in human infants and adults (Galazka et al., 2016; Meyerhoff et al., 2014b), they should be faster to detect the color change applied to the chaser than to the chasee. Aside from response time, we also tested accuracy with the same predictions. To control for a preference for the motion pattern regardless of the role, we implemented a random condition where we removed the interaction such that two objects were present that moved chaser-like and chasee-like, but non-contingently to each other. Additionally, to control for a preference for the object's position relative to the other, we implemented a condition where a leader, positioned in front, shows the way to a follower, positioned behind. In this case, the roles were reversed compared to the chasing condition: the agent (the leader) moves in front and the patient (the follower) behind.

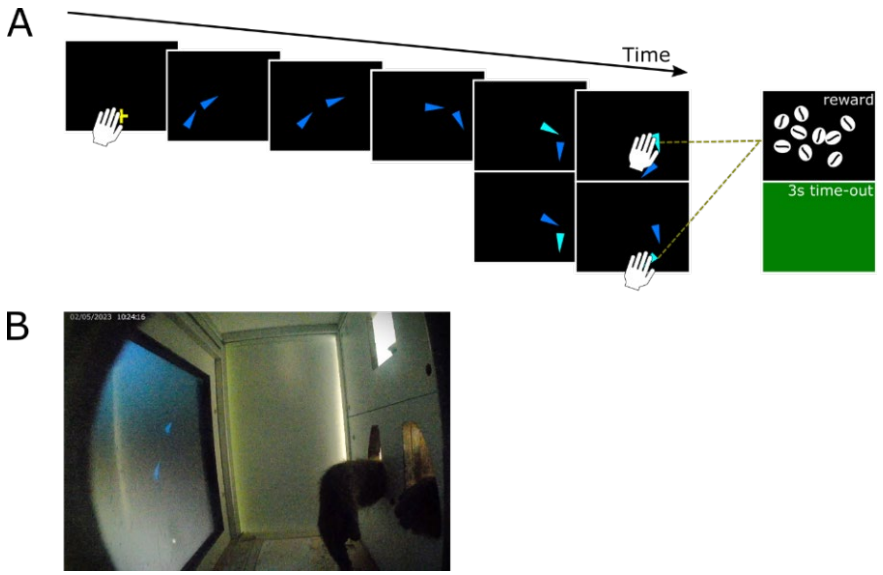


Fig. 1. Experimental set-up. Panel A shows an experimental trial of the chasing change detection task. First, a fixation cross appears that has to be touched by the participant. Next, an animation with moving objects is generated. Depicted here are two objects involved in a chasing interaction. After a certain period, one of the objects will change color. On the top row the agent changed color, on the bottom row the patient changed color. The changed object requires a touch response which subsequently results in a food reward as shown by the dotted line (top). Any other touches result in a three-second time-out screen (bottom). Panel B shows a participant taking part in the experiment.

Method

This study has been pre-registered here and materials used in the experiment, including examples of the generated animations, are available on OSF.

Ethics statement

The study on baboons was carried out in accordance with French and EU standards and received approval from the French Ministère de l'Education Nationale et de la Recherche (approval #APAFIS-2717-2015111708173794-V3). Procedures used in the present study were also consistent with the guidelines of the Association for the Study of Animal Behavior.

Participants

The study was made available to a group of 23 baboons who are housed at the "Station de Primatologie" in Rousset-sur-Arc (France) in two groups of eighteen and five individuals. 13 individuals (10 females, mean age = 12.1 ± 1.6(SEM) years, age range = 5.5-25.1 years) successfully learnt the task and were included in our analyses. The baboons were tested by using 14 automatic learning devices for monkeys (ALDM; Fagot & Paleressompoulle, 2009), equipped with touch screens and a food dispenser, which were freely accessible from their enclosure. The baboons participated voluntarily in these tasks employing an operant conditioning method.

Data collection started at the end of April 2023 and finished in the beginning of June 2023 for a period of forty days.

General procedure

We designed a change detection task. For humans, change detection paradigms have revealed a substantial role of attention in the detection of changes in scenes (Rensink et al., 1997). Importantly, areas that attract attention are prone to faster change detection than less-attended regions, such that, for example, changes to animate entities are noticed faster than to inanimate entities (Altman et al., 2016; New et al., 2007, 2010). A similar rationale is used for dot-probe tasks which show faster response times for touching the dot behind an attention-attracting picture and this type of paradigm has been successfully applied to non-human primates (van Rooijen et al., 2017).

In our adaptation of the change detection task, we used a color change which required a touch response. The color change was applied to an object in the shown animation, see Fig. 1. A trial consisted of a fixation cross, which had to be touched to start. Next, an animation was generated in which one object (during training) or two objects (during testing) were moving around on the screen. When one of them changed color, the participant had to touch it to earn a food reward. Touching either the background, the object that did not change color or any object before it changed color, resulted in a punishment of a three second green time-out screen, after which the trial ended. Without

any touching response, the animation stopped after six seconds, and no reward was distributed. The experiment was created in and presented to the baboons with OpenSesame (Mathôt et al., 2012).

Stimuli and conditions

The stimuli were generated using a Pygame script in OpenSesame. For each trial, a new animation was generated such that no trial was ever the same. The animations were shown with a framerate of 60 fps on a black 1024 x 768 screen. The objects were triangles of 70 pixels in size, oriented towards the moving direction. Compared to circular shapes, triangles are known for enhancing the chasing perception (Abdai et al., 2021; Gao et al., 2009). The position of each object on the screen was updated each frame to make it look like it was moving. The stimuli were created with the goal of making the objects in the animation appear animate by mimicking the ability to self-propel and perform speed and directional changes; features to which newborns (Di Giorgio et al., 2017, 2021) and newly-hatched chicks (Rosa-Salva et al., 2016) are sensitive.

The moving behaviors were based on Reynold's descriptions of how to program naturally moving autonomous agents as described by Shiffman (2012). The moving direction was defined by the current velocity vector and the 'desired' vector. Each frame, the object was updated from the current to the desired vector (the steering behavior), making the object appear to be moving into the direction of the desired vector. The desired vector had a combined direction and length based on a combination of different vectors that are acting on the object. Which forces made up the desired vector depended on the object. Details can be found in our Open Sesame script, which is available on OSF.

Several testing conditions will be presented below. For the chasing condition, we created a chasing interaction with configurations that are required for the impression of chasing to emerge for humans (Gao et al., 2009; Visch & Tan, 2009). During chasing, one object was the chaser that would always move towards a second object, the chasee. The chasee was wandering around and would accelerate and flee away when the chaser came close by (distance smaller than 200 pixels). The chaser is thus positioned behind and the chasee in front. In this interaction, the chaser is the agent and the chasee is the patient. We had two versions of chasing, one where the objects started further away from each other, creating a "heat-seek pursuit" from the chaser, and one where they would start close to one another. A heat-seek pursuit helps human adults and infants detect chasing (Galazka & Nyström, 2016; Gao et al., 2009).

For the following condition, we adjusted the behaviors of chasing such that we removed the fleeing behavior of the object positioned in front, making it seem as if it was leading the way for the object behind. Note that here the agent-patient roles are reversed compared to chasing and now the object positioned in front is the agent and object behind the patient. Again, we implemented two versions of following, one where the two objects started far away from one another and one where they started close by.

In the random condition, two objects were present that were moving exactly like a chaser and a chasee but not contingently. To achieve this, we generated two sets of

chasing interactions of which one showed only the chaser and the other only the chasee. In the first set, we made the object positioned in front black, leaving only the chaser visible, and in the second, we made the object behind black, leaving only the chasee visible.

Additionally, we had a final condition called clone. The clone condition was comparable to the one used by Atsumi & Nagasaka for squirrel monkeys (2015). Two objects moved side-by-side, one object moved as a chaser (chasing an invisible chasee) and a second object was placed alongside it. We implemented this condition because we wanted to exclude the possibility that potential differences between the chasing and the random condition could be explained by the fact that in chasing the objects were closer together than in random (regardless of the relation) making it easier to parse these stimuli. Baboons are known to have a local processing bias in visual tasks (Deruelle & Fagot, 1998), which may facilitate the processing of objects closer together compared to objects further apart. In the clone condition, we thus decided to keep the two objects close to each other, but not expressing an agent-patient relation, to test whether this would lead to shorter response times compared to random.

Example videos of our generated animations can be found on OSF.

Training

We implemented a five-step training phase to familiarize the baboons with our change detection paradigm. During the training phase, only one object moved across the screen. The participant had to touch the object after it changed color. If the participant touched the object that changed color (within a radius of 140 pixels), this counted as a correct response.

We incrementally increased the timing of the color change and we decreased the perceptual saliency of the color change. In step 1, the color change appeared 100ms after the onset of the video, in step 2 it appeared at a randomly chosen time between 500 and 1000ms, in step 3 it was between 1000 and 3000ms and in step 4 and 5 between 3000 and 4000ms. In steps 1-4 we used the color change from blue (RGB(0,0,255)) to cyan (RGB(0,255,255)), in step 5 we used dark blue (RGB(0,55,99)) to another dark blue (RGB(37,43,99)). The two dark blue colors were chosen because they are close, but baboons are still able to perceive them as different (Davidoff & Fagot, 2010). At the same time, this color difference is subtle, which requires the baboons' attention to detect it and respond appropriately.

During training, the generated animations were based on the chasing stimuli from the test phase. Half of the trials, we showed one object moving chaser-like and half of the trials, chasee-like. We presented these trials in blocks of 60 trials and when a criterion of 80% correct responses was reached, the participant continued to the next step. We measured which participants reached the criteria and in how many blocks they did so (see SI).

Test phase

During the test phase we showed our participants two-object animations. We implemented different conditions in the test phase: chasing (both with and without heat-seek pursuit), following (with and without heat-seek pursuit), random and clone (see SI).

The different conditions were displayed in random order in blocks of 60 trials. We thus had 10 trials per condition per block. The color change was applied at a randomly chosen time between 3000 and 4000ms from dark blue (RGB(0,55,99)) to another dark blue (RGB(37,43,99)). The color change was applied to one object in half of the trials and to the other in the remaining trials. Touching the object that changed color (within a radius of 140 pixels) counted as a correct response. We measured response time. We additionally measured accuracy, see SI for details. We collected 80 blocks of 60 trials, leading to 800 trials per condition per participant.

Analyses

Response time

To measure whether participants would be faster to respond to a color change applied to the agent compared to the patient in a chasing interaction, we tested whether there would be an interaction effect between the response times in chasing and the other conditions. We analyzed our data using a Bayesian generalized linear mixed model with ex-gaussian error structure implemented with the brms R package (Bürkner, 2017). Our preregistered model (RT ~ condition * target + (1+block|participant) turned out to be too complex, as it did not converge. To reduce complexity, we decided to 1) use target instead of block for the varying slopes, 2) focus on the most relevant conditions only: chasing, following and random, excluding the clone condition[a], 3) merge the chasing with and without heat-seek pursuit as well as the following with and without heat-seek pursuit (as we detected no differences between them, see SI). We continued our model with a chasing combined condition, a following combined condition and a random condition.

For all correct trials, we modelled the effect of condition (chasing combined, following combined and random) interacting with target (which of the two objects present changed color) on the response time after the color change, with intercepts varying per participant. We used the following model: RT ~ condition * target + (1+target|participant). We applied 6000 iterations and 4 chains.

For each parameter, we report estimates (B), estimated error (EE), and the 95% credible interval (CI). If zero lies outside the credible interval, we conclude there is sufficient evidence to suggest the estimate is different from zero.

Additionally, to determine whether there was an absolute difference between the response times of the two targets per condition, we conducted post-hoc pairwise

[a] The clone condition was implemented to verify whether it may be easier to parse an animation with two objects that move close together (like chasing and following) compared to far apart (like random), but this was not the case. We observed an average response time of 1024 ± 32(SEM) ms in the clone condition, which appears to be higher than the response times in the random condition: 947 ± 14(SEM)) ms averaged over the two objects, indicating that the clone condition with two objects close together was not easier, but in fact harder, to parse than the other conditions.

comparisons between the response times of the two objects within each condition on the computed estimated marginal means using the function 'contrast' of the package 'emmeans' (Lenth, 2022).

Accuracy

Even though participants in the test phase were selected for high accuracy during training (participants needed to reach 80% correct responses to proceed through the five steps of training), we also assessed whether participants would identify a color change applied to the agent in a chasing interaction more accurately than a color change applied to the patient. We used a Bayesian generalized linear mixed model with binomial error structure implemented with the brms R package (Bürkner, 2017). The preregistered model was too complex, so we adjusted similarly to the response time model, by 1) using target instead of block for the varying slopes, 2) excluding the clone condition and 3) merging chasing and chasing without heat-seek pursuit into chasing combined and merging following and following without heat-seek pursuit into following combined (see below).

We modelled participants' responses (1 for correct, 0 for incorrect) using a mixed logit model specified as: response ~ condition * target + (1+target|participant) with target being either of the two objects present that changed color. We used 4000 iterations and 4 chains.

Like for the response time analyses, we report estimates (B), estimated error (EE), and the 95% credible interval (CI) and conclude there is enough evidence for an estimate different from zero if zero lies outside the credible interval.

Results

Training phase

The 13 baboons who successfully passed the training phases took on average 2806 ± 301 trials. Of the 13 who succeeded training, 10 completed all 4800 test trials. Three others who completed 2417 test trials, 2763 test trials and 3568 test trials were included in the analyses. See SI Table 2 for individual data.

Test phase

Response time

Our response time analysis revealed an interaction between the targets in the chasing condition (i.e. agent and patient) and the targets random condition (i.e. agent-like and patient-like) (B = -34.16, EE = 5.88, 95% CI = [-45.79, -22.69]) and no interaction between the targets in the chasing condition and the following condition (i.e. agent and patient) (B = -8.51, EE = 4.73, 95% CI = [-17.96, 0.83), see Fig. 2.

In the chasing condition, the participants were significantly faster to detect the color change to the agent, the chaser, compared to the patient, the chasee (922 ± 22(SEM) ms vs. 951 ± 18.4(SEM) ms, B = -14.86; 95% CI = [-24.79, -5.23]), in line with the hypothesized agent preference. 11 of 13 baboons detected a color change faster for the agent than the patient in chasing. Even though the following condition aligned in effect direction with chasing, we did not detect an absolute difference between the response times to the agent, the leader, and the patient, the follower (954 ± 30(SEM) ms vs. 966 ± 35(SEM) ms, B = -6.35; 95% CI = [-16.18, -3.33]). 9 of 13 baboons had a faster response time for the agent compared to the patient. In the random condition, the baboons recognized a color change to the chaser-like object slower than to the chasee-like object (962 ± 22(SEM) ms vs. 932 ± 16(SEM) ms, B = 19.25, 95% CI = [7.43, 31.08]). This was true for 9 of 13 baboons.

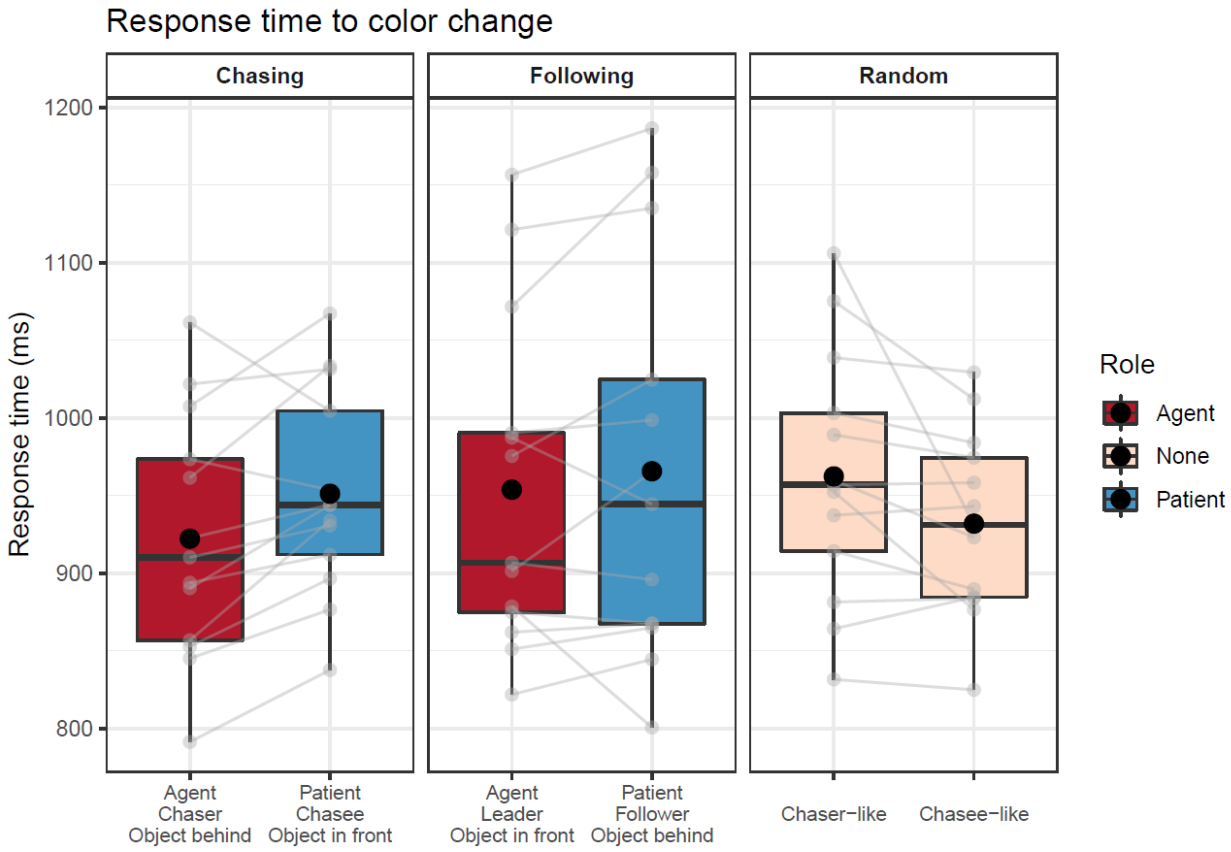


Fig. 2. Response times (in ms) to the color changes applied to the two moving objects in our different conditions. Average response times are displayed with a large black dot and the grey lines indicate individual data.

Accuracy

Accuracy in detecting color changes was consistently high for all conditions and targets because participants were trained to be accurate throughout training and testing by positive reinforcement. The accuracy analysis showed an interaction between the targets of the chasing condition and the following condition (B = 1.27, EE = 0.07, 95% CI = [1.12, 1.41]), and between the targets of chasing and the random condition (B = 0.60, EE = 0.09, 95% CI = [0.42, 0.77]), see Fig. 3.

Post-hoc, we tested for differences in accuracy to detect a color change between the two targets in each condition. In the random condition, the baboons were equally accurate to detect the color change to the chaser-like object as to the chasee-like object (91.1 ± 1.3(SEM) % correct vs. 90.7 ± 1.0(SEM) % correct, B = -0.06, 95% CI = [-0.08, 0.22]). For chasing, the baboons were more accurate for the agent, the chaser, the object behind, compared to the patient, the chasee, the object positioned in front (92.1 ± 1.0(SEM) % correct vs. 86.3 ± 1.1(SEM) % correct, B = 0.66, 95% CI = [0.55, 0.77]). During the following condition, the baboons identified the color change to the patient, the follower, the object positioned behind, more accurately than to the agent, the leader, the object in front (94.2 ± 0.8(SEM) % correct vs. 89.6 ± 0.9(SEM) % correct, B = -0.60, 95% CI = [-0.72, -0.48]).

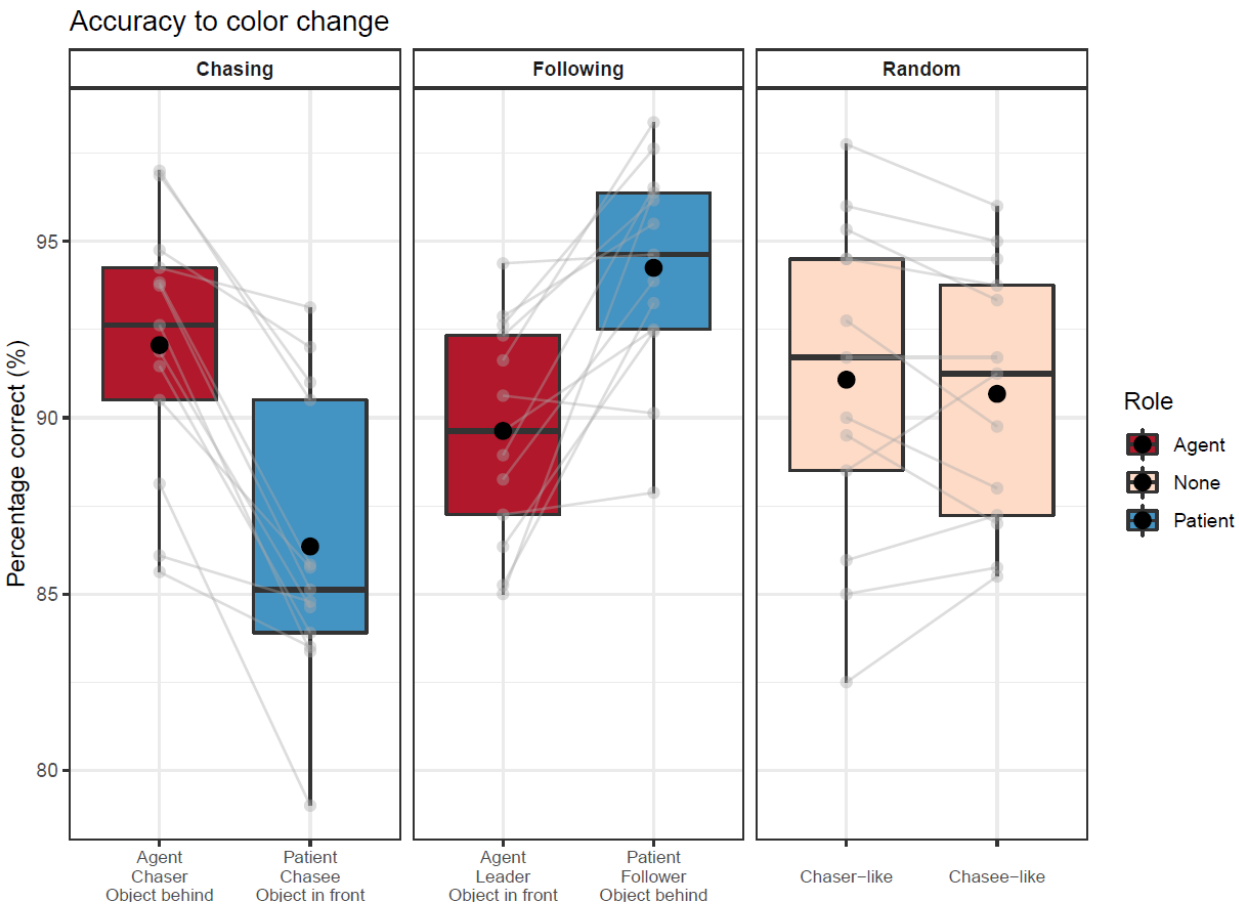


Fig. 3. Accuracy (in % correct) to the color changes applied to the two moving objects in our different conditions. Average accuracy scores are displayed with a large black dot and the grey lines indicate individual data.

So, in the random condition, where objects were distant, baboons were as accurate to detect a color change applied to the chaser-like object as to the chasee-like object. However, in both the chasing and following conditions, where the objects moved together in concert, color changes to the trailing object positioned behind were detected more accurately. Presumably this is due to the predictable movements of the second object

positioned behind (it would move in the direction of the one in front, whereas the one positioned in front had “freedom” to go across the screen either to flee the chaser or to lead the follower), making it simply easier to successfully tap it after the color change had been identified. An additional possibility is that response time is a more sensitive measure for attention allocation than accuracy; indeed previous studies focused on response time to test agent and patient role extraction from pictures (Hafri et al., 2018) and, importantly, this measure has been used to show an agent preference in chasing stimuli (Meyerhoff et al., 2014b).

Discussion

This study demonstrates that Guinea baboons have an agent preference in chasing interactions: they are faster to detect a color change that is applied to the agent compared to the patient of a chasing interaction. This suggests that, consistent with our hypothesis, baboons exhibit an attentional bias towards the chaser similarly to human adults (Meyerhoff et al., 2014b) and preverbal infants (Galazka & Nyström, 2016).

Importantly, faster responses for the chaser compared to the chasee cannot be attributed to the chaser’s specific motion pattern, because in the random condition, where the objects moved independently, baboons exhibited a faster response for the chasee-like object instead of the chaser-like object. We speculate that this facilitating effect stemmed from the chasee's more erratic movements. Similar preferences for looking at unpredictable movements have been documented in human adults and dogs (Abdai et al., 2017) and newly-hatched chicks (Lemaire et al., 2022). Crucially, when the same two objects move closer together and in contingency in the chasing condition, this preference for a separately moving chasee shifted in favor of the chaser. We view this as strong evidence for an agent preference, suggesting that the coordinated motion during chasing overcame the preference for the chasee's motion.

Similarly, faster responses for the chaser compared to the chasee cannot be the result of baboons anticipating the chaser's trajectory based on the motion of the object positioned in front. In the following condition, baboons equally fast to respond to the object positioned in front, whose trajectory is not predictable (the agent/leader) compared to the object positioned behind, whose trajectory could be predicted from the first one (the patient/follower). This suggests that motion predictability is not what makes a color change easier to detect for chasers compared to chasees. In contrast, motion predictability fully explained response accuracy, with baboons being more accurate to touch the trailing object compared to the one positioned in front in both the chasing and the following condition. Previous studies also use response time to argue in favor of an agent bias (e.g. Hafri et al., 2018; Meyerhoff et al., 2014b), presumably because response time is a more sensitive measure for involuntary attention allocation (Prinzmetal et al., 2005; Smith & Ratcliff, 2009).

In summary, our findings suggest that baboons possess an agent preference, i.e. a prioritized attention towards agents when observing events. This agent bias was visible during chasing events. In following events, even though the effect went into the same

direction as in chasing, the preference for the agent did not reach statistical significance. One possibility is that the agent and patient roles in this condition were less salient compared to chasing, because the movement patterns the two objects were very similar, making it harder to discriminate between them. This is consistent with data from 5-months-old human infants who looked longer to a chaser in a chasing interaction but not to a leader in a following interaction (Galazka & Nyström, 2016). Baboons' agent preference is thus so far restricted to chasing events. Although data on the agent preference in human adults is based on more varied event types, most infant studies used chasing patterns (Galazka et al., 2016, 2016; Yin & Csibra, 2015) making it impossible to judge whether they attend to the chaser because it is the chaser or because it portrays the agent role in an abstract sense (but see: Papeo et al., 2024). Further investigation is thus needed to determine whether the agent preference we uncovered for chasing, extends to other events.

The discovery that baboons do not process the chasing events holistically, but instead decompose them into agent and patient roles with a specific focus on the agent, alludes to the possibility that baboons cognitively represent events in a similar format to humans, akin to a language of thought. The agent preference indeed fulfills a key property of a language of thought: the presence of constituents that are discrete and structured (Quilty-Dunn et al., 2023).

The similarity between baboons' and humans' event processing has important implications. A fundamental cognitive capacity to represent events with a preference for the agent, may form the basis of event syntax in languages, which is proposed by Wilson and colleagues as the agency detection hypothesis (Wilson et al., 2022). This theory accounts for the tight link between event cognition and linguistic structure (Papafragou & Grigoroglou, 2019; Rissman & Majid, 2019; Ünal & Papafragou, 2016) and elaborates that cross-linguistic syntactic regularities in linguistic event descriptions are externalizations of how we mentally represent them (Strickland, 2017; Ünal, Ji, et al., 2021; Zuberbühler & Bickel, 2022). At minimum, our results suggest that the cognitive mechanisms involved in processing events are evolutionarily old and can possibly be traced back to at least the last common ancestor between baboons and humans who lived approximately 30 Mya.

Considering the relatively recent emergence of language in humans, we propose that it draws upon various cognitive functions that have evolved for diverse purposes. The ability to represent 'who is doing what to whom' may be a fundamental cognitive capacity shared across species. An agent preference serves as a possible explanation for why we observe a cross-linguistic tendency to emphasize the subject by placing it first. This word order may stem from an agent preference rooted in cognition, challenging the idea that word order patterns are solely the product of deep linguistic principles.

Acknowledgements

This work was supported by the Agence Nationale de la Recherche (ANR-20-CE28-0005) and the Institute for Language, Communication and the Brain (ILCB) (ANR-16-

CONV-0002). We thank Julie Gullstrand, Sebastien Barniaud and the staff at the Primatology Station for their help with the experiments with the baboons. Thanks to Jean-Baptiste Melmi for his advice in the early stages of creating the chasing stimuli.

References

Abdai, J., Ferdinandy, B., Lengyel, A., & Miklósi, Á. (2021). Animacy perception in dogs (Canis familiaris) and humans (Homo sapiens): Comparison may be perturbed by inherent differences in looking patterns. Journal of Comparative Psychology, 135(1), 82–88. https://doi.org/10.1037/com0000250

Abdai, J., Ferdinandy, B., Terencio, C. B., Pogány, Á., & Miklósi, Á. (2017). Perception of animacy in dogs and humans. Biology Letters, 13(6), 20170156. https://doi.org/10.1098/rsbl.2017.0156

Abdai, J., & Miklósi, Á. (2022). Selection for specific behavioural traits does not influence preference of chasing motion and visual strategy in dogs. Scientific Reports, 12(1), 2370. https://doi.org/10.1038/s41598-022-06382-6

Altman, M. N., Khislavsky, A. L., Coverdale, M. E., & Gilger, J. W. (2016). Adaptive attention: How preference for animacy impacts change detection. Evolution and Human Behavior, 37(4), 303–314. https://doi.org/10.1016/j.evolhumbehav.2016.01.006

Atsumi, T., Koda, H., & Masataka, N. (2017). Goal attribution to inanimate moving objects by Japanese macaques (Macaca fuscata). Scientific Reports, 7(1), Article 1. https://doi.org/10.1038/srep40033

Atsumi, T., & Nagasaka, Y. (2015). Perception of chasing in squirrel monkeys (Saimiri sciureus). Animal Cognition, 18(6), 1243–1253. https://doi.org/10.1007/s10071-015-0893-x

Barrett, H. C., Todd, P. M., Miller, G. F., & Blythe, P. W. (2005). Accurate judgments of intention from motion cues alone: A cross-cultural study. Evolution and Human Behavior, 26(4), 313–331. https://doi.org/10.1016/j.evolhumbehav.2004.08.015

Bickel, B., Witzlack-Makarevich, A., Choudhary, K. K., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2015). The neurophysiology of language processing shapes the evolution of grammar: Evidence from case marking. PLOS ONE, 10(8), e0132819. https://doi.org/10.1371/journal.pone.0132819

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. Journal of Statistical Software, 80, 1–28. https://doi.org/10.18637/jss.v080.i01

Bybee, J. (2006). Language change and universals. In Linguistic universals (pp. 179–212). Cambridge University Press.

Bybee, J. (2009). Language universals and usage-based theory. In Language universals (pp. 17–39). Oxford University Press.

Chomsky, N. (1957). Syntactic Structures. In Syntactic Structures. De Gruyter Mouton. https://doi.org/10.1515/9783110218329

Cohn, N., & Paczynski, M. (2013). Prediction, events, and the advantage of Agents: The processing of semantic roles in visual narrative. Cognitive Psychology, 67(3), 73–97. https://doi.org/10.1016/j.cogpsych.2013.07.002

Culbertson, J., & Kirby, S. (2016). Simplicity and Specificity in Language: Domain-General Biases Have Domain-Specific Effects. Frontiers in Psychology, 6. https://doi.org/10.3389/fpsyg.2015.01964

Culbertson, J., Schouwstra, M., & Kirby, S. (2020). From the world to word order: Deriving biases in noun phrase order from statistical properties of the world. Language, 96(3), 696–717.

Davidoff, J., & Fagot, J. (2010). Cross-species assessment of the linguistic origins of color categories. Comparative Cognition & Behavior Reviews, 5. https://comparative-cognition-and-behavior-reviews.org/vol5_davidoff_fagot/

Deruelle, C., & Fagot, J. (1998). Visual search for global/local stimulus features in humans and baboons. Psychonomic Bulletin & Review, 5(3), 476–481. https://doi.org/10.3758/BF03208825

Di Giorgio, E., Lunghi, M., Simion, F., & Vallortigara, G. (2017). Visual cues of motion that trigger animacy perception at birth: The case of self-propulsion. Developmental Science, 20(4), e12394. https://doi.org/10.1111/desc.12394

Di Giorgio, E., Lunghi, M., Vallortigara, G., & Simion, F. (2021). Newborns' sensitivity to speed changes as a building block for animacy perception. Scientific Reports, 11(1), 542. https://doi.org/10.1038/s41598-020-79451-3

Dobel, C., Gumnior, H., Bölte, J., & Zwitserlood, P. (2007). Describing scenes hardly seen. Acta Psychologica, 125(2), 129–143. https://doi.org/10.1016/j.actpsy.2006.07.004

Dowty, D. (1991). Thematic proto-roles and argument selection. Language, 67(3), 547. https://doi.org/10.2307/415037

Dryer, M. S. (2013). Order of Subject, Object and Verb. In M. S. Dryer & M. Haspelmath (Eds.), The World Atlas of Language Structures Online. Zenodo. https://doi.org/10.5281/zenodo.7385533

Dunn, M., Greenhill, S. J., Levinson, S. C., & Gray, R. D. (2011). Evolved structure of language shows lineage-specific trends in word-order universals. Nature, 473(7345), 79–82. https://doi.org/10.1038/nature09923

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. Behavioral and Brain Sciences, 32(5), 429–448. https://doi.org/10.1017/S0140525X0999094X

Fagot, J., Boë, L.-J., Berthomier, F., Claidière, N., Malassis, R., Meguerditchian, A., Rey, A., & Montant, M. (2019). The baboon: A model for the study of language evolution. Journal of Human Evolution, 126, 39–50. https://doi.org/10.1016/j.jhevol.2018.10.006

Fagot, J., & Paleressompoulle, D. (2009). Automatic testing of cognitive performance in baboons maintained in social groups. Behavior Research Methods, 41(2), 396–404. https://doi.org/10.3758/BRM.41.2.396

Fillmore, C. J. (1968). Lexical Entries for Verbs. Foundations of Language, 4(4), 373–393.

Fischer, J., Higham, J. P., Alberts, S. C., Barrett, L., Beehner, J. C., Bergman, T. J., Carter, A. J., Collins, A., Elton, S., Fagot, J., Ferreira da Silva, M. J., Hammerschmidt, K., Henzi, P., Jolly, C. J., Knauf, S., Kopp, G. H., Rogers, J., Roos, C., Ross, C., … Zinner, D. (2019). Insights into the evolution of social systems and species from baboon studies. eLife, 8, e50989. https://doi.org/10.7554/eLife.50989

Frankenhuis, W. E., House, B., Clark Barrett, H., & Johnson, S. P. (2013). Infants' perception of chasing. Cognition, 126(2), 224–233. https://doi.org/10.1016/j.cognition.2012.10.001

Galazka, M., Bakker, M., Gredebäck, G., & Nyström, P. (2016). How social is the chaser? Neural correlates of chasing perception in 9-month-old infants. Developmental Cognitive Neuroscience, 19, 270–278. https://doi.org/10.1016/j.dcn.2016.05.005

Galazka, M., & Nyström, P. (2016). Infants' preference for individual agents within chasing interactions. Journal of Experimental Child Psychology, 147, 53–70. https://doi.org/10.1016/j.jecp.2016.02.010

Gao, T., Newman, G. E., & Scholl, B. J. (2009). The psychophysics of chasing: A case study in the perception of animacy. Cognitive Psychology, 59(2), 154–179. https://doi.org/10.1016/j.cogpsych.2009.03.001

Gertner, Y., & Fisher, C. (2012). Predicted errors in children's early sentence comprehension. Cognition, 124(1), 85–94. https://doi.org/10.1016/j.cognition.2012.03.010

Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How Efficiency Shapes Human Language. Trends in Cognitive Sciences, 23(5), 389–407. https://doi.org/10.1016/j.tics.2019.02.003

Goto, K., Lea, S. E., & Dittrich, W. H. (2002). Discrimination of intentional and random motion paths by pigeons. Animal Cognition, 5(3), 119–127. https://doi.org/10.1007/s10071-002-0139-6

Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In Universals of language (pp. 40–70). MIT Press.

Hafri, A., Papafragou, A., & Trueswell, J. C. (2013). Getting the gist of events: Recognition of two-participant actions from brief displays. Journal of Experimental Psychology: General, 142(3), 880–905. https://doi.org/10.1037/a0030045

Hafri, A., Trueswell, J. C., & Strickland, B. (2018). Encoding of event roles from visual scenes is rapid, spontaneous, and interacts with higher-level visual processing. Cognition, 175, 36–52. https://doi.org/10.1016/j.cognition.2018.02.011

Hauser, M. D. (1998). A nonhuman primate's expectations about object motion and destination: The importance of self-propelled movement and animacy. Developmental Science, 1(1), 31–37. https://doi.org/10.1111/1467-7687.00009

Hofrichter, R., & Rutherford, M. D. (2019). Early attentional capture of animate motion: 4-Year-Olds show a pop-out effect for chasing stimuli. Perception, 48(3), 228–236. https://doi.org/10.1177/0301006619828256

Jackendoff, R. (1999). Possible stages in the evolution of the language capacity. Trends in Cognitive Sciences, 3(7), 272–279. https://doi.org/10.1016/S1364-6613(99)01333-9

Jackendoff, R. (2002). Foundations of Language: Brain, Meaning, Grammar, Evolution. Oxford University Press UK.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. Proceedings of the National Academy of Sciences, 105(31), 10681–10686. https://doi.org/10.1073/pnas.0707835105

Lemaire, B. S., Rosa-Salva, O., Fraja, M., Lorenzi, E., & Vallortigara, G. (2022). Spontaneous preference for unpredictability in the temporal contingencies between agents' motion in naïve domestic chicks. Proceedings of the Royal Society B: Biological Sciences, 289, 20221622. https://doi.org/10.1098/rspb.2022.1622

Lenth, R. V. (2022). emmeans: Estimated Marginal Means, aka Least-Squares Means (Version 1.9.0) . https://CRAN.R-project.org/package=emmeans

MacWhinney, B. (1977). Starting Points. Language, 53(1), 152–168. https://doi.org/10.2307/413059

Martin, A., Adger, D., Abels, K., Kanampiu, P., & Culbertson, J. (2024). A Universal Cognitive Bias in Word Order: Evidence From Speakers Whose Language Goes Against It. Psychological Science, 35(3), 304–311. https://doi.org/10.1177/09567976231222836

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. Behavior Research Methods, 44(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

Meyerhoff, H. S., Schwan, S., & Huff, M. (2014). Perceptual animacy: Visual search for chasing objects among distractors. Journal of Experimental Psychology: Human Perception and Performance, 40(2), 702–717. https://doi.org/10.1037/a0034846

New, J., Cosmides, L., & Tooby, J. (2007). Category-specific attention for animals reflects ancestral priorities, not expertise. Proceedings of the National Academy of Sciences, 104(42), 16598–16603. https://doi.org/10.1073/pnas.0703913104

New, J., Schultz, R. T., Wolf, J., Niehaus, J. L., Klin, A., German, T. C., & Scholl, B. J. (2010). The scope of social attention deficits in autism: Prioritized orienting to people and animals in static natural scenes. Neuropsychologia, 48(1), 51–59. https://doi.org/10.1016/j.neuropsychologia.2009.08.008

Papafragou, A., & Grigoroglou, M. (2019). The role of conceptualization during language production: Evidence from event encoding. Language, Cognition and Neuroscience, 34(9), 1117–1128. https://doi.org/10.1080/23273798.2019.1589540

Papeo, L., Vettori, S., Serraille, E., Odin, C., Rostami, F., & Hochmann, J.-R. (2024). Abstract thematic roles in infants' representation of social events. Current Biology. https://doi.org/10.1016/j.cub.2024.07.081

Prinzmetal, W., McCool, C., & Park, S. (2005). Attention: Reaction time and accuracy reveal different mechanisms. Journal of Experimental Psychology: General, 134(1), 73–92. https://doi.org/10.1037/0096-3445.134.1.73

Quilty-Dunn, J., Porot, N., & Mandelbaum, E. (2023). The best game in town: The reemergence of the language-of-thought hypothesis across the cognitive sciences. Behavioral and Brain Sciences, 46, e261. https://doi.org/10.1017/S0140525X22002849

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. Psychological Science, 8(5), 368–373. https://doi.org/10.1111/j.1467-9280.1997.tb00427.x

Rissman, L., & Majid, A. (2019). Thematic roles: Core knowledge or linguistic construct? Psychonomic Bulletin & Review, 26(6), 1850–1869. https://doi.org/10.3758/s13423-019-01634-5

Rochat, P., Morgan, R., & Carpenter, M. (1997). Young infants' sensitivity to movement information specifying social causality. Cognitive Development, 12, 537–561.

Rosa-Salva, O., Grassi, M., Lorenzi, E., Regolin, L., & Vallortigara, G. (2016). Spontaneous preference for visual cues of animacy in naïve domestic chicks: The case of speed changes. Cognition, 157, 49–60. https://doi.org/10.1016/j.cognition.2016.08.014

Schafroth, J. L., Basile, B. M., Martin, A., & Murray, E. A. (2021). No evidence that monkeys attribute mental states to animated shapes in the Heider–Simmel videos. Scientific Reports, 11(1), Article 1. https://doi.org/10.1038/s41598-021-82702-6

Seyfarth, R. M., & Cheney, D. L. (2014). The evolution of language from social cognition. Current Opinion in Neurobiology, 28, 5–9. https://doi.org/10.1016/j.conb.2014.04.003

Shiffman, D. (2012). Chapter 6: Autonomous agents. In The nature of code. https://natureofcode.com/autonomous-agents/

Smith, P. L., & Ratcliff, R. (2009). An integrated theory of attention and decision making in visual signal detection. Psychological Review, 116(2), 283–317. https://doi.org/10.1037/a0015156

Strickland, B. (2017). Language Reflects "Core" Cognition: A New Theory About the Origin of Cross-Linguistic Regularities. Cognitive Science, 41(1), 70–101. https://doi.org/10.1111/cogs.12332

Ünal, E., Ji, Y., & Papafragou, A. (2021). From Event Representation to Linguistic Meaning. Topics in Cognitive Science, 13(1), 224–242. https://doi.org/10.1111/tops.12475

Ünal, E., & Papafragou, A. (2016). Interactions Between Language and Mental Representations. Language Learning, 66(3), 554–580. https://doi.org/10.1111/lang.12188

van Rooijen, R., Ploeger, A., & Kret, M. E. (2017). The dot-probe task to measure emotional attention: A suitable measure in comparative studies? Psychonomic Bulletin & Review, 24(6), 1686–1717. https://doi.org/10.3758/s13423-016-1224-1

Visch, V. T., & Tan, E. S. (2009). Categorizing moving objects into film genres: The effect of animacy attribution, emotional response, and the deviation from non-fiction. Cognition, 110(2), 265–272. https://doi.org/10.1016/j.cognition.2008.10.018

Webb, A., Knott, A., & MacAskill, M. R. (2010). Eye movements during transitive action observation have sequential structure. Acta Psychologica, 133(1), 51–56. https://doi.org/10.1016/j.actpsy.2009.09.001

Wilson, F., Papafragou, A., Bunger, A., & Trueswell, J. (2011). Rapid extraction of event participants in caused motion events. Proceedings of the Annual Meeting of the Cognitive Science Society, 7.

Wilson, V. A. D., Sauppe, S., Brocard, S., Ringen, E., Daum, M. M., Wermelinger, S., Gu, N., Andrews, C., Isasi-Isasmendi, A., Bickel, B., & Zuberbühler, K. (2023). Primate origins of human event cognition (p. 2023.11.23.568086). bioRxiv. https://doi.org/10.1101/2023.11.23.568086

Wilson, V. A. D., Zuberbühler, K., & Bickel, B. (2022). The evolutionary origins of syntax: Event cognition in nonhuman primates. Science Advances, 8(25), eabn8464. https://doi.org/10.1126/sciadv.abn8464

Yin, J., & Csibra, G. (2015). Concept-based word learning in human infants. Psychological Science, 26(8), 1316–1324. https://doi.org/10.1177/0956797615588753

Zuberbühler, K. (2019). Evolutionary roads to syntax. Animal Behaviour, 151, 259–265. https://doi.org/10.1016/j.anbehav.2019.03.006

Zuberbühler, K. (2020). Syntax and compositionality in animal communication. Philosophical Transactions of the Royal Society B: Biological Sciences, 375(1789), 20190062. https://doi.org/10.1098/rstb.2019.0062

Zuberbühler, K. (2022). Event parsing and the origins of grammar. WIREs Cognitive Science, 13(3). https://doi.org/10.1002/wcs.1587

Zuberbühler, K., & Bickel, B. (2022). Transition to language: From agent perception to event representation. WIREs Cognitive Science, 13(6). https://doi.org/10.1002/wcs.1594

Supporting Information

Individual data

Table S2: For each participant, the number of blocks of 60 trials completed per phase are displayed. During training there was one object changing color. There are five training phases to increment the time before a color change took place as well as the subtlety of the color change. Not all participants made it to the final test phase. In the test phase, there are two objects of which one changed color. We have data from 13 individuals who participated in the test phase, displayed here in green.

Participant	training1	training2	training3	training4	training5	testing
Angele	1	55	NA	NA	NA	NA
Arielle	1	17	5	7	36	80
Articho	2	1	NA	NA	NA	NA
Atmosphere	1	10	7	2	32	46
Bobo	1	25	NA	NA	NA	NA
Dora	2	6	2	1	11	59
Dream	1	NA	NA	NA	NA	NA
Ewine	2	11	3	1	46	80
Fana	1	31	7	7	18	80
Felipe	1	23	4	1	6	40
Feya	1	49	4	3	2	NA
Flute	2	25	NA	NA	NA	NA
Harlem	2	67	5	NA	NA	NA
Hermine	1	19	0	NA	NA	NA
Lips	2	12	4	2	9	80
Lome	1	7	5	3	18	80
Mako	1	57	3	2	3	80
Mali	1	22	2	2	7	80
Muse	1	13	2	1	6	80
Nekke	1	33	7	15	16	80
Petoulette	1	14	NA	NA	NA	NA
Pipo	1	10	NA	NA	NA	NA
Violette	1	6	6	3	32	80

Heat-seek pursuit methods

To compare the conditions with and without heat-seek pursuit, we created four Bayesian generalized linear mixed models implemented with the brms R package (Bürkner, 2017).

Concerning response time, we took all correct trials and fitted two ex-gaussian models with the following structure: RT ~ condition * target + (1+target|participant). For the model including chasing and chasing without heat-seek pursuit, we applied 4000 iterations and 4 chains. For the model with following and following without heat-seek pursuit, we instead used 6000 iterations and 4 chains as the model was not converging with 4000 iterations.

Concerning accuracy, we took all trials and used two binomial models with the following structure: accuracy ~ condition * target + (1|participant). For the model including chasing and chasing without heat-seek pursuit, we applied 6000 iterations and 4 chains (as the model with 4000 iterations did not converge). For the model with following and following without heat-seek pursuit, we used 4000 iterations and 4 chains.

Heat-seek pursuit results

We decided to merge the conditions with and without heat-seek pursuit, because we did not detect differences between them. Neither the response time, nor the accuracy model showed an interaction between the targets of the chasing and chasing without heat-seek conditions (RT: B = 3.95, EE = 7.01, 95% CI = [-9,83, 17.73], accuracy: B = -0.14, EE = 0.10, 95% CI = [-0.34, 0.06]).

We also found no interaction effect in the accuracy and response time models between the two targets in the following and the following without heat-seek condition (RT: B = 3.83, EE = 6.13, 95% CI = [-8.10, 15.96], accuracy: B = 0.11, EE = 0.10, 95% CI = [-0.12, 0.302).

We did not detect an effect of the heat-seek pursuit for both chasing and following. A heat-seek pursuit helps human adults and infants detect chasing (Galazka & Nyström, 2016; Gao et al., 2009), but apparently this is not the case for baboons. This points to a difference in how human adults and infants process this feature compared to baboons. From personal observations, we know that the baboons do not focus on the screen for the full duration of the animations, but rather glance at the screen every now and then. It is possible that the baboons looked at the animation mostly around the time that the color change would appear (always between 3000 and 4000 ms) and therefore do not show any effect of the heat-seek pursuit which happened at the beginning of the animations only. Consistent with the idea that relations are perceived rapidly and automatically (Hafri et al., 2018; Hafri & Firestone, 2021), a heat-seek pursuit appears not to be a necessary feature to deduce agent-patient roles from a chasing interaction.

5. Discussion

This thesis has presented three studies that investigated whether event representations of non-human primates and preverbal infants contain linguistically relevant features, such as sensitivity to causality and to agent-patient roles. The first two experiments did not show the existence of such linguistically relevant features, the last experiment hints at it. After having summarised the main findings of these studies, I will discuss the implications below.

5.1. Summarising the main findings

The first study demonstrated that Guinea baboons do not use causality when categorising different Michottean events, where human adults do. Baboons learned equally fast to categorise events according to a division based on causality compared to a matched control condition, when humans were faster for the former. The results therefore confirm that human adults quickly extract and use causality from launching events (cf. Scholl & Tremoulet, 2000), but baboons use only the spatiotemporal features of Michottean launches.

The second study showed that when applying a role reversal method to Michottean events, no evidence for sensitivity to agent and patient roles was found in preverbal infants, human adults and Guinea baboons. This study included an attempt to replicate the seminal study by Leslie and Keeble (1987), but did not find the same results. Since there was no effect of the role reversal in all three populations, including human adults for whom it is well-established that they extract event roles from events (e.g. Hafri et al., 2018), we argued that the role reversal method may be not suitable to probe event role attribution.

The last study revealed that for chasing stimuli, Guinea baboons show an attentional bias towards the agent of the interaction. They were faster to respond to a colour change that was applied to the chaser compared to the chasee, an effect that could not be explained by low-level features such as the motion pattern or the relative positioning of the moving objects. Such an attentional bias towards agents is in line with how human adults and human infants process chasing events (Galazka & Nyström, 2016; Meyerhoff et al., 2014b). We concluded that this attentional bias towards the agent is in line with the tendency to prioritise agents by placing them first in the sentence in subject position.

5.1.1. Contributions and limitations

Since two of three papers present null-findings regarding the main research question, namely, whether linguistically relevant mental representations exist in the absence of language, I will discuss here the scientific contributions and limitations of the three articles presented in this thesis.

5.1.1.1. Article 1

The first study about causal perception in Guinea baboons and human adults presents two main contributions: a critical evaluation of the previous literature regarding causal perception in non-humans and the development of a novel method to test for causality attribution. We point out that several studies in non-human animals show discriminatory abilities between different Michottean events, but do not critically test the attribution of causality. Since demonstrating that discrimination abilities for different causal and non-causal events does not directly prove that this distinction is made on the basis of causality (because the same results can be obtained by focussing on spatiotemporal properties), we have developed a new method to tackle specifically this question. We have further validated this methodology in human adults. Our task can further be adapted to test other species to gain a wider understanding of causal perception from a phylogenetic perspective.

However, this study also has some limitations. We present a null-result for the attribution of causality by baboons in our task, and, given that null-results are notoriously difficult to interpret, our findings do not warrant the conclusion that Guinea baboons cannot perceive causality in Michottean launches *per se*. It may be possible that the baboons had perceived the causality, but did just not apply this in the categorisation task. While trying to probe causal perception, the task simultaneously taxes other cognitive mechanisms which may result in difficulties that are unrelated to causality (cf. Seed et al., 2012). Clearly, the task was demanding, as shown by the number of trials needed to learn the classification and the small number of baboons fulfilling the task's criteria. It may be that to solve the task, it was easier for the baboons to focus on spatiotemporal properties. That the baboons tend to gravitate towards solutions that are easy for them, was also observable in the critical phase of our experiment, where they had to set one type of Michottean events apart from the other three, leading to an unequal division of 25 and 75%. Evidently, as we pointed out in the manuscript's discussion, this unequal division has paved the way for some baboons to adopt a press-one-response-button-all-the-time method which delivers a reward 75% of times whilst preventing them from learning the actual rule. It was thus absolutely not easy for the baboons to overcome the easy solution and learn the actual rule. Apart from that, it was significantly harder to correctly classify the L event separately in the causal vs. non-causal condition compared to the STG event in the control condition (both represented 25% of occurrences). This clearly contrasts with the human adults who apparently find it easy to classify the causal launch as distinct from the other events – even though there is a discrepancy in responses (25-75%). An adjustment to the task to allow for more equally distributed responses may circumvent this problem.

Another possibility is that baboons really did not extract causality from the Michottean displays, but this does not mean that they possess no causal representations at all (since there is ample evidence that non-humans have some understanding of causality (Penn & Povinelli, 2007; Schloegl & Fischer, 2017; Völter & Call, 2017)), only that they do not perceive causality in Michottean events. Except for a demonstration of the causal capture

effect in chimpanzees (Matsuno & Tomonaga, 2017), there is so far no evidence for causality perception in Michottean launches in non-human animals.

We wanted to test for causal perception, because many studies report that the perception of causality concerns low-level perceptual processes (Kominsky & Scholl, 2020; Moors et al., 2017; Ohl & Rolfs, 2023; Rolfs et al., 2013), which made it a key candidate to have a long evolutionary history. For example, visual processing has revealed similarities between Guinea baboons and humans, such as the susceptibility to some visual illusions (Barbet & Fagot, 2002, 2007; Fagot et al., 2006). Given the limitations outlined above and the null-findings, we wonder if a categorisation task is truly appropriate to study a perceptual phenomenon. A more appropriate test for *perception* of causality would be to check whether retinotopically specific adaptation takes place upon seeing many repetitions in non-humans as well, but this is not (ethically) feasible as this requires the launching event to reach the same area of the retina over and over again. Other ways to test for causal perception, could be to pursue the methodology of the causal capture effect which was successful in chimpanzees (Matsuno & Tomonaga, 2017), although it has the drawback of not having the comparative data from human infants. The causal capture effect makes use of the contrast is between launching and overlapping events, whereas infant studies have been employing the contrast between launching and gap events.

5.1.1.2. Article 2

Study 2 puts forward as contribution a failed replication of Leslie and Keeble's study and subsequent reflections on the suitability of applying the role reversal method with Michottean launches to address the question of sensitivity to agent-patient roles. This is important since the original paper has had a major influence on the field of cognitive science generating subsequent interpretations that infants of 6 months old possess representations of agent and patient (i.e. cited 1334 times according to google scholar (03-10-2024)). Replication studies remain necessary in light of the replication crisis (see: Pashler & Wagenmakers, 2012), and the Leslie and Keeble study needs further scrutiny. Future studies should be aware that this cornerstone study may be less reliable than previously thought. While the idea that preverbal infants possess abstract representations of agent and patient was recently corroborated (Papeo et al., 2024), we demonstrated that the role reversal method with Michottean launches does not accurately test this capacity.

Another methodological contribution of the second study is the implementation of the mirror control condition. This control is crucial in determining whether a role reversal, rather than other differences in the two stimuli types, may be driving differences in looking time.

A limitation of study 2 is that we did not use exactly the same testing methods for babies as for human adults and baboons. The method for the human adults and baboons had not been applied before and it is possible that taking the measure of accuracy is not suitable to test a potential effect of the role reversal. Accuracy reflected the ability to differentiate a launching event from a spatial and temporal gap event. We made the

assumption that if the reversal of the launch is detected as a role reversal, this would generate a certain disturbance or surprisal leading to classification mistakes. Given the null-finding we cannot be certain that this assumption is correct. To counter this, we did observe that the mirror control conditions resulted in lower accuracy scores, giving credit to the idea that implementing changes (compared to the baseline trials) generate mistakes. All in all, together with the absence of role reversal effect in infants, we deem the theory that the role reversal in Michottean launches does not reliably capture agent-patient role attributions the most likely explanation of our findings.

5.1.1.3. Article 3

Study 3 provides a scientific contribution in demonstrating that chasing events are perceived in a similar way by Guinea baboons as human adults (Meyerhoff et al., 2014b) and infants (Galazka & Nyström, 2016). Additionally our study provides empirical evidence in favour of the agency detection hypothesis posited by Wilson *et al.* (2022), revealing that the cognitive capacity to process events as non-holistically is not language-dependent nor unique to humans.

The change detection task that we designed is comparable to tasks to determine the saliency of event roles that have been previously used in human adults and children (Ünal et al., 2024; Ünal, Richards, et al., 2021) and shows that this paradigm can successfully be applied to non-humans.

A limitation to the present study is that currently the attentional bias was only demonstrated for chasing. Even though each stimulus presentation was different, indicating that such bias is not restricted to just one specific video of chasing, a clear agent bias was only observable in chasing. To go from a bias observed in chasing to abstract representations, there remains a need to prove that an attentional bias exists regardless of the event at hand. Such more generalised biases have been shown for human adults and infants based on postural information and were consequently taken as indications of abstract mental representations (Hafri et al., 2018; Papeo et al., 2024) and we are therefore currently working on a novel task to test the same in Guinea baboons.

Another limitation which was briefly addressed in the manuscript, concerns the difference in results regarding accuracy and response time. While I hypothesised that they would be equally impacted by the attentional bias, this turned out to be not the case. Accuracy and response time were differently affected by the shapes' movements and colour change. Accuracy was trained for throughout training and testing by positive feedback, it may therefore better reflect how difficult or easy it was to tap the object and the position of the second object was presumably more predictable. Additionally, it could be the case that response time may better reflect involuntary attention (Prinzmetal et al., 2005; Smith & Ratcliff, 2009).

5.2. Implications

This thesis asked whether event representations of non-human primates and preverbal infants contain linguistically relevant features. Can this question be answered based on the three studies conducted during this thesis project?

At a first glance, these studies appear to contradict each other; how is it possible that based on the first two studies the answer is suggestive of a 'no' and based on the third study, a tentative 'yes' can be given? In fact, there are few reasons indicating that these three studies may not be contradicting each other. Since study 1 and 2 portray null effects, there is no direct proof that mental representations of events that are similar to human adults are *not* present. From study 3, a first glimpse into agent-patient perception in non-human primates points to similar processing. A second extrapolation from this finding, is that Guinea baboons seem to have caught up on the causal structure from the chasing relation. When agent and patient roles are extracted, this implies that the relative functioning of the two roles are grasped, and thus their relation which is causal.

What differences in the studies made it possible to show agent-patient differentiation (and by extension, causality) in study 3, but not in study 1 and 2? Besides potential methodological limitations in study 1 and 2 as described above, it could be that this concerns a difference between processing of mechanical and social relations, thus differentiating between agent-patient relations involving only objects or also involving an animate-like entity. For human infants, those two areas (objects vs. agents) have been proposed to be differently processed according to the core cognition theory (Spelke & Kinzler, 2007). Possibly certain cues related to agency facilitate the perception of the event in terms of causality and agent-patient roles. For example, the chasing interactions were specifically programmed to portray motion that are perceived as animate, such as speed changes and directional changes (Gao et al., 2009, 2019). This divide could be important, because, for example, infants of 5 months old hold different expectations for what is likely subsequent behaviour when an object is inert or has been shown to have an internal energy source (Luo et al., 2009). In the Michottean events that we used for the baboons and human adults, the agent did not show any cues of having an internal energy source. Possibly this prevented the participants to view the event as an agent-patient relation in the launching event, in contrast to the chasing event which did involve such features.

Based on these three studies, a tentative conclusion is that non-human primates and preverbal infants possess event representations that have some linguistically relevant elements (i.e. extracting agent and patient and their causal relation), but actually accessing those experimentally is not a trivial task. Given that at least one study revealed positive evidence for agent and patient role sensitivity, this shows that such cognitive capacities can be demonstrated in a non-human primate and do not depend on language. Next, it will be discussed how the finding of such an attentional bias relates to the evolution of language.

5.2.1. Linking the current findings to language evolution

This thesis is built around the theory that some linguistically relevant features may be present in the event cognition of non-human species. The agency detection hypothesis proposed that the processing of events as component parts with a specific focus on agents may form the basis of syntax (Wilson et al., 2022). Demonstrations for such a way of processing in non-human primates would thus nourish the idea that the cognitive origins of some properties of language are present beyond the human lineage.

In Article 3 we present findings in favour of the agency detection hypothesis: Guinea baboons process chasing events in component parts, since they differentiate the agent and patient by demonstrating a specific attentional bias towards the agents. This is reminiscent of how human adults and human infants process chasing events (Galazka & Nyström, 2016; Meyerhoff et al., 2014b). As mentioned above, establishing that such an attentional bias extends beyond chasing is necessary to fully support the agency detection hypothesis. We can thus only very tentatively infer that the common ancestor of humans and Guinea baboons may have also possessed such an attentional bias, catapulting this capacity to be present as long ago as 25 to 30 Mya. Ideally, other non-human primate species should be tested on the same task, to make the cladistic inferences about the capacities of common ancestors more confidently.

How can an attentional bias towards agents present in Guinea baboons have consequences for human language? Since language is such a recently acquired ability, it is not an uncommon idea that it must exploit pre-existing cognitive machinery which has evolved for very different purposes. One such cognitive system may be event cognition, which likely evolved to help organisms make sense of the world. Event cognition allows us to parse and understand sequences of actions. That such a capacity could be already found in a common ancestor of Guinea baboons and humans, also indicates that this ability was present in early humans, possibly skewing evolution of their linguistic capacities. From very early on, young infants show adaptive biases towards social stimuli, such as faces and biological motion (Simion et al., 2011). These biases may get refined over time to specifically tune to our (social) environment and culminate in a more general event processing mechanism. Throughout the first year of life an agent bias may develop because it is useful to understand how an event unfolds from the beginning (with the agent's act) to the end. The attention being drawn to agents as a result of how event cognition operates, may make it easier to form a coherent concept about agents and subsequently to learn to map symbols correctly onto these concepts while learning language. For example, it is easier for children to learn a novel word for the agent in chasing interaction (Yin & Csibra, 2015), presumably because this role had attracted the attention. How attention is distributed across the event parallels the unrolling of it, which makes it straightforward to express it in this same way (cf. Schlenker et al., 2024). This relationship between event cognition and language learning likely extends to language evolution. As (early) humans naturally observed and categorized events in the world, the languages they developed would reflect these patterns of attention. The natural way of

observing events may have facilitated production and processing of language that follows this sequential order. The human predisposition to focus on agents likely influenced the structure of languages over time, leading to a tendency to place the subject first. In such a way, simple attentional biases may have impressive consequences that are visible in language universals nowadays.

Even though such an evolutionary account may not be directly testable, it generates testable predictions. For example, that it should be harder to acquire linguistic structures that deviate from the default that is inherited from our primate ancestors. An interesting road to follow might be, first, to map out the acquisition trajectory of multiword utterances in languages which follow OS, and compare this to SO languages to see if OS might be acquired later in life. Another idea would be to adapt the paradigms that measure how a first noun phrase is perceived in children (e.g. Abbot-Smith et al., 2017; Gertner & Fisher, 2012) and apply this to learners of an OS language, would they also expect the first occurring noun phrase to be the agent or would they overextend their linguistic rule and interpret it as object? These are just some ideas to further extend our understanding of how an agent bias may affect language acquisition and possibly evolution.

5.3. Outstanding questions and reflections

5.3.1. What constitutes an agent in the eyes and mind of a Guinea baboon or a human infant?

From Article 3 it appears that Guinea baboons have an attentional preference towards agents, but this prompts the question of how this agent was detected. Based on the results from the chasing events, it appears that contingent movements in which one influences the behaviour of the other, are used by Guinea baboons to extract an agent from an ongoing event. Only when the two objects moved in such a way that it was indicative of a relation, the agent preference emerged. Importantly, besides movement, there were no other cues present to discriminate the agent from the patient. So far, we do not know which other features could aid non-human primates in delineating agents from patients, because other studies in great apes used more ecologically valid, but less controlled stimuli resulting in ambiguous findings (Brocard et al., 2024; Wilson et al., 2023).

For human infants, there is more information to go by. Similarly to Guinea baboons, contingent relational motion cues were also what drove the agent bias in the same chasing stimuli (Galazka & Nyström, 2016). However, postural and positional cues can also be used to denote agents and patients (Papeo et al., 2024).

Humans are able to extract agent and patient from a wide diversity of events. We see that the same cues, posture and contingent relational movement, are of importance for human adults to discriminate agent from patient (Hafri et al., 2013, 2018; Meyerhoff et al., 2014b), indicating that there seems to be a continuity between young infants. For human adults it has further been established that objects should move close to another to elicit a discrimination between chasing and random movements (Meyerhoff et al., 2014a).

What the concept of agent might look like in the mind of a baboon or a preverbal infant is hard to tell. For example, these representations may be in an iconic format (as has already been proposed for infants: e.g. Mandler, 1992). Also baboons appear to represent some things, such as shapes, in an iconic sense (Vauclair et al., 1993). Another possibility is that they represent agent and patient in a format comparable to a language of thought (LOT). This has been proposed to be the case for baboons because of their intricate social knowledge (Seyfarth & Cheney, 2008). Recent evidence suggests that preverbal human infants may have a language of thought based on the fact that they appear to represent agent and patient in an abstract way (Papeo et al., 2024). Our findings in baboons show too that events are decomposed and might have an inherent structure (agents over patients), which seems to go beyond an iconic representation of the event. This opens up the possibility to apply the same line of reasoning as for infants to baboons. The presence of constituents that are discrete and structured is indeed one of the proposed key criteria of a language of thought (Quilty-Dunn et al., 2023). However, critically, we have not yet shown that these representations are truly abstract and show so-called role-filler independence, which means that the relation remains regardless of the entities in it and

the entities can exist outside of the relation. Other criteria indicative of a language of thought also need to be further examined to fully support this hypothesis.

5.3.2. Beyond agent and patient event roles

Besides agents and patients there are multiple other event roles, such as recipients, sources, goals and instruments, that we have not directly studied in this thesis project. A review article studied multiple of them and posited that event roles other than agent and patient can be left out of event descriptions and are presumably less cognitively robust than the agent and patient (Rissman & Majid, 2019). Still, it is worth investigating whether they are represented in a similar way in other species.

Take for example the role of goal. This event role is often contrasted with the role of source (much like the distinction between agent and patient). Using this contrast, it has become clear that goal is more likely to be encoded linguistically than source in children and adults who were asked to describe a video which depicted both source and goal (e.g. Do et al., 2022; Lakusta & Landau, 2005). A study sampling languages using allative and ablative markers showed that these markers can be applied rather widely, but that the goal-source dichotomy is also reflected in their separate uses (Kabata, 2013). So, even though these elements are not always expressed, if they are, goals are more likely to be placed in the sentence than sources. This is reflected in cognition, as, for example, goals are more easily retrieved from memory than sources (Chen et al., 2024).

Goals may have a special place in our cognitive system, as it has become clear that infants are specifically sensitive to goal-directed behaviour. For example, young infants of 3 months old are surprised to see an actor reaching for a new goal after having being habituated on another goal, but only after having played with the toys themselves (Sommerville et al., 2005). Goals are not just attributed to humans. For example, 12-month-old infants who were habituated to a circle jumping over a barrier to reach a second circle (i.e. the goal), are surprised to see that same circle perform the same jump as before when the barrier has been removed, but not if it performs a new action of moving towards the second circle directly (Gergely et al., 1995). This was replicated for 9-, but not 6-month-old infants (Csibra et al., 1999). In additional experiments, it was found that movement cues indicative of agency, such as self-propulsion, were not necessary to instantiate the goal attribution, leading the authors to suggest that movement cues are not a precondition for goal-assignment (Csibra et al., 1999). Infants may thus interpret behaviour assuming that goals are pursued efficiently without necessarily assigning intentional mental states to entity behaving goal-directed (Gergely & Csibra, 2003). Another study also looked at the relative saliency of goals relative to sources in infants and found that 12 months old infants look longer at events that had a goal change at test compared to the same events showing a source change, indicating that they show a bias towards goals (Lakusta et al., 2007). It thus appears that young infants have representations of goals and sources, and that goals take a specific salient cognitive space.

Non-human primates also attribute goals. One study adapted the previously mentioned experiment in which an object jumped over a barrier and found that young chimpanzees

are less surprised when the object moved straight to its goal after the barrier had been removed (Uller, 2004). Several macaques species (*Macaca nemestrina, Macaca fascicularis* and *Macaca mulatta*) and chimpanzees (*Pan troglodytes*) attribute goals to a human hand reaching for a target (Kano & Call, 2014; Rochat et al., 2008). Another showed that chimpanzees produce similar looking patterns towards goals as human adults, although this was not yet visible in human infants (Myowa-Yamakoshi et al., 2012).

There are thus various event representations related to (frequently) expressed linguistic structures that could be further explored in non-human primates. As illustrated here with the goal role, there are results are less straightforward than for agents and patients. In particular in non-human primates, the relative grasp of goal versus source has not yet been explored. A broader view into other event elements would strengthen the theory that event representations in general, and thus beyond agent and patient, are of importance for language evolution.

5.3.3. A brief examination of static topological events

The present thesis looked specifically at the cognitive roots of dynamic events with an agent-patient structure, but there are also other types of events, such as those denoting a state. The crucial difference with the events discussed in this thesis is that no causal change takes place. We can ask whether the theories regarding event representations in dynamic events may also apply more broadly? An important area of static events revolve around topological relations: spatial relations that include a figure which is located relatively to a ground (Levinson & Wilkins, 2006), such as "The painting hangs on the wall". Relations like these are described using prepositions in English and denote event states such as containment and support. It has been proposed that the linguistic expression in the domain of static spatial relations parallels the underlying spatial representations (Landau & Jackendoff, 1993). Moreover, Gentner and Bowerman (2010) proposed the typological prevalence hypothesis, stating that cross-linguistically frequent ways of categorising spatial relations correspond to the natural ways humans conceptualise them, which in turn makes them easier for children to learn. This theory thus makes explicit, for topological relations specifically, what is also at the core of this thesis regarding the link between language universals and (human) cognition in the domain of two actor dynamic event representations.

Support for the claim that there may be cognitively salient spatial relations, comes from research showing that infants discriminate among a variety of them from very early on (e.g. Casasola, 2008; Clark, 2004). For example, already at 4.5 months old, infants look longer at events in which a support is lifted, thereby leaving a box floating mid-air compared to one that is sustained by a hand, indicating that they have some notion of support (Needham & Baillargeon, 1993). During acquisition of spatial terms, infants must be able to generalise their representation irrespective of the specific objects involved, indicating that they represent those relations in an abstract fashion to be able to map a specific relation word (Casasola & Cohen, 2002). For example, infants generalise from 6

months onwards for the difference between above and below (Quinn et al., 1996) and for between relations (Quinn et al., 1999). Interestingly, there may be spatial relations that infants are sensitive to at a young age, but they will lose this discrimination ability over time if they learn a language that does not encode such a distinction, such as for the distinction of tight fit vs. loose fit which is encoded in Korean, but not in English (Hespos & Spelke, 2004; McDonough et al., 2003).

Whether such distinctions are shared with other species paints a more varied picture. Data from non-human primates indicate that they too can discriminate between certain spatial relations, but not all, and may do so with different processing mechanisms compared to humans. For example, Guinea baboons (*Papio papio*)and tufted capuchins (*Cebus apella*) can be trained to discriminate above and below relations (resp. Dépy et al., 1999; Spinozzi et al., 2004). Two other studies showed that chimpanzees and Japanese macaques (*Macaca fuscata*) discriminate impossible from possible support relations displayed in videos, but do so based on amount of contact between the two objects, which appears to be different from the more abstract support category in humans (Cacchione & Krist, 2004; Murai et al., 2011). Moreover, knowledge about support relations does not always translate into appropriate searching behaviour. For example, in a food search task, rhesus macaques preferentially search in incorrect places as if a solid object could move through another (Hauser, 1998; Santos, 2004). The same was found for cotton-top tamarins (*Saguinus oedipus*) even though it was confirmed that they can detect the violations with looking time measures (Santos et al., 2006). Based on these studies, it seems that non-human primates may process spatial relations differently to humans, which make it more difficult to have this knowledge guide behaviour.

Without clear evidence from our primate relatives, it could still be possible that in humans specific spatial relations are more cognitively natural and therefore widely expressed in languages across the world. However, the typological evidence that indeed certain spatial categories are expressed more frequently than others is currently not strong enough.

Typological data on how static relations are expressed is not available for many languages. This type of data is typically generated from studies that ask people to describe a variety of topological relations depicted on pictures (cf. Bowerman & Pederson, 1992) and it is subsequently determined which pictures are described using the same term to define semantic clusters. This way, it can be observed that some closely related languages show differences in how they express relations: English describes horizontal and vertical support using a single term ("*on*"), whereas Dutch employs a finer-grained distinction and uses two different terms ("*op*" and "*aan*", respectively) (Bowerman & Choi, 2001). Based on this method, there is some suggestive evidence that languages tend to cluster attachment, containment, superposition, and a category combining "near" and "under" relations (Levinson et al., 2003). However, this generalisation is based on a sample that contained only nine, albeit diverse, languages. Studying more languages is thus essential, but that would clearly be a difficult and time-consuming endeavour. In a side-project in which I was involved, Tanguy Tiran explored whether automatic translation might be a way to circumvent this problem and obtain spatial categorisation data on a larger sample

of languages to validate whether languages universally encode relations in the same way (Tiran et al., 2023, see Annex A). We elicited descriptions for a variety of depicted topological relations in English speakers. These descriptions were translated to about a hundred different languages using Google Translate. Then, we extracted the spatial terms by aligning the translated sentences to their originals. Lastly, we implemented a clustering analysis to determine which pictures were described by the same spatial terms. We identified five clusters corresponding to containment, under-relations, proximity, support, and a combination of on and attachment. These results thus only partly overlap with Levinson's (2003) clusters (e.g. attachment, containment). Even though the clustering results were encouraging, the developed method is in its current form not reliable enough (especially with regards to the extracted spatial terms from the translated sentences) to produce robust conclusions regarding which categorisations are typologically frequently encoded and which are not.

Because of this scarcity in typological data, there is currently no consensus as to which spatial relations to study with regards to the potential underlying cognitive distinctions that might drive the linguistic description. For example, another study decided to focus on a the distinction of mechanical object relationships (in and on) vs. non-mechanical object relations (behind, above) in human adults, and showed that words denoting the first category are cross-linguistically significantly shorter, suggesting that these are frequently used words (Strickland & Chemla, 2018). They had thus adopted broader relation categories than other studies, but without all typological data it is difficult to decide which categories are worth studying. All in all, it is insufficiently clear whether different languages carve up spatial relationships in the same way to determine whether humans do so cognitively in the same way.

Besides spatial relation representations or event representations, there are presumably other event domains in which to ask these same questions (e.g. dynamic events with one actor or even multiple actors that may not be acting on one another ("The woman is running"/"The women are running"). However, an important first step is thus to establish that there are indeed linguistic universals before determining whether cognitive factors are likely to be implicated in them. This calls for interdisciplinary work between linguists, typologists, and psychologists. Lastly, only if there is a clear link, it may be fruitful to probe whether similar event representations may be present in other species, to gain more insights from a language evolution perspective.

5.3.4. Why are many mental representations not expressed in non-human communication?

It will remain an open question why non-humans do not express the event elements that they have mental access to in their communication. This question is not restricted to the domain of event representations; there is a wide range of non-human primates' cognitive activities that are not directly accessible through their communication (cf. Fitch, 2019).

What stages in the evolution of language facilitated this expression of mental representations in human communication? Zuberbühler and Bickel (2022) note that one could judge some primate vocalisations as denoting agents and that specifically the ability to express agents *and* patients asks for an explanation. They then propose that human's ability to understand that current events are embedded in past and future events, with the potential of patients becoming future agents, may be out of reach for non-human primates, making this one of the key steps in human evolution leading to the capacity to express agent-patient events. This argument is reminiscent of the idea that displacement, the ability to talk about objects and events that are not there in the present moment, is one of the uniquely human features of language (Hockett, 1960). However, since great apes actually show some signs of mentally representing displaced events, for example they plan for the future (Janmaat et al., 2014; Mulcahy & Call, 2006), but do not address this through their communication (although they have been shown to be able to implement a pointing gesture to refer to a displaced object (Lyn et al., 2014)), the proposal that this is one of the critical steps is not the whole story. In fact, it circles back to the question of why non-human primates may possess a mental representation (e.g. about a future state) and then not express it.

A key point that may prevent non-human primates from having more extended expressibility, is a certain flexibility to coupling signals to mental representations. For example, if we go along with the point mentioned above that non-human primates express the agent, but not the patient, this may be because they are unable to assign it to a signal. Humans on the other hand do have the ability to entertain new representations and subsequently assign a new symbol to it (e.g. for novel inventions, such as the internet). Perhaps this inability in non-human primates can just be attributed to their relatively small number of signals. In the vocal domain, for instance, non-human primates are constrained by a limited set of vocalizations, which restricts the range of messages they can convey and the number of mental representations they can utter as a consequence. The levels of vocal learning are limited to slightly modifying already existing calls (e.g. Lemasson et al., 2011) or their use (Wegdell et al., 2019). There is no evidence for learning entirely new calls (but see learning of whistling (Lameira et al., 2013)), which might explain why they would only have calls for the utmost necessary expressions. However, the small repertoire is not the defining limitation. In the gestural domain, we see the opposite problem. Even though there is a larger repertoire of gestures, they still convey only few different meanings (Byrne et al., 2017). So, the problem of coupling mental representations and signals is not because there are insufficient signals. In the gestural domain, one could argue that there are even surplus signals. This means that there is a curious lack of flexibility in how animals can couple available signals to different mental representations. I think this coupling followed by an increased ability to learn entirely new signals, would have been a crucial step in the evolution of human language.

Of course, there are many different reasons as to why non-human primates do not express many of their mental representations. Rather than asking this question, it may be more appropriate to ask why humans do just that. Theorising further about the evolution

of declarative language, however, extends beyond the scope of this current work and remains a topic open for future research and exploration.

5.4. Conclusion

Although languages differ according to their unique conventions, certain statistical patterns nevertheless emerge, some of which are due to cognitive constraints. In the domain of linguistic event descriptions, there is a large body of research linking them to underlying abstract mental representations in human adults. Some of those appear to be already present in preverbal infants, even before they possess the ability to express them linguistically. Also in non-human species that will never express them in a similar vein, the conceptual apparatus may share some important commonalities with that of humans. The present thesis studied such event representations and demonstrated in three experiments differences and similarities in how non-human primates, human infants and human adults process dynamic two-actor events. We have demonstrated that Guinea baboons may not process launching events as portraying a causal relation, that some established methodological paradigms appear to be less suitable to tap into the attribution of event roles in causal relations than was thought, and that Guinea baboons have attentional preferences towards agents mirroring a crosslinguistic trend. Currently, we stand only at the beginning of uncovering to what extent event representations in non-human primates contain linguistically relevant features. Based on these initial findings we can speculate that the ability to transform the outside world into mental representations is one of the important building blocks of human language that may not be unique to humans. Human language has presumably evolved from this pre-existing representational system into a way of communicating that.

References

Abbot-Smith, K., Chang, F., Rowland, C., Ferguson, H., & Pine, J. (2017). Do two and three year old children use an incremental first-NP-as-agent bias to process active transitive and passive sentences?: A permutation analysis. *PLOS ONE*, *12*(10), e0186129. https://doi.org/10.1371/journal.pone.0186129

Abdai, J., Ferdinandy, B., Lengyel, A., & Miklósi, Á. (2021). Animacy perception in dogs (*Canis familiaris*) and humans (*Homo sapiens*): Comparison may be perturbed by inherent differences in looking patterns. *Journal of Comparative Psychology*, *135*(1), 82–88. https://doi.org/10.1037/com0000250

Abdai, J., Ferdinandy, B., Terencio, C. B., Pogány, Á., & Miklósi, Á. (2017). Perception of animacy in dogs and humans. *Biology Letters*, *13*(6), 20170156. https://doi.org/10.1098/rsbl.2017.0156

Abdai, J., & Miklósi, Á. (2022). Selection for specific behavioural traits does not influence preference of chasing motion and visual strategy in dogs. *Scientific Reports*, *12*(1), 2370. https://doi.org/10.1038/s41598-022-06382-6

Adibpour, P., & Hochmann, J.-R. (2023). Infants' understanding of the causal power of agents and tools. *Proceedings of the National Academy of Sciences*, *120*(50), e2309669120. https://doi.org/10.1073/pnas.2309669120

Altman, M. N., Khislavsky, A. L., Coverdale, M. E., & Gilger, J. W. (2016). Adaptive attention: How preference for animacy impacts change detection. *Evolution and Human Behavior*, *37*(4), 303–314. https://doi.org/10.1016/j.evolhumbehav.2016.01.006

Altmann, G. T. M., & Ekves, Z. (2019). Events as intersecting object histories: A new theory of event representation. *Psychological Review*, *126*(6), 817–840. https://doi.org/10.1037/rev0000154

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

Arnold, K., & Zuberbühler, K. (2006). Semantic combinations in primate calls. *Nature*, *441*(7091), 303–303. https://doi.org/10.1038/441303a

Aslin, R. N. (2007). What's in a look? *Developmental Science*, *10*(1), 48–53. https://doi.org/10.1111/j.1467-7687.2007.00563.x

Atsumi, T., Koda, H., & Masataka, N. (2017). Goal attribution to inanimate moving objects by Japanese macaques (*Macaca fuscata*). *Scientific Reports*, *7*(1), Article 1. https://doi.org/10.1038/srep40033

Atsumi, T., & Nagasaka, Y. (2015). Perception of chasing in squirrel monkeys (*Saimiri sciureus*). *Animal Cognition*, *18*(6), 1243–1253. https://doi.org/10.1007/s10071-015-0893-x

Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, M. A. (2001). Infants parse dynamic action. *Child Development*, *72*(3), 708–717. https://doi.org/10.1111/1467-8624.00310

Baldwin, D. A., & Kosie, J. E. (2021). How does the mind render streaming experience as events? *Topics in Cognitive Science*, *13*(1), 79–105. https://doi.org/10.1111/tops.12502

Barbet, I., & Fagot, J. (2002). Perception of the corridor illusion by baboons (*Papio papio*). *Behavioural Brain Research*, *132*(1), 111–115. https://doi.org/10.1016/S0166-4328(01)00393-X

Barbet, I., & Fagot, J. (2007). Control of the corridor illusion in baboons (*Papio Papio*) by gradient and linear-perspective depth cues. *Perception*, *36*(3), 391–402. https://doi.org/10.1068/p5108

Barbet, I., & Fagot, J. (2011). Processing of contour closure by baboons (*Papio papio*). *Journal of Experimental Psychology: Animal Behavior Processes*, *37*(4), 407–419. https://doi.org/10.1037/a0025365

Barner, D., Wood, J., Hauser, M., & Carey, S. (2008). Evidence for a non-linguistic distinction between singular and plural sets in rhesus monkeys. *Cognition*, *107*(2), 603–622. https://doi.org/10.1016/j.cognition.2007.11.010

Barrett, H. C., Todd, P. M., Miller, G. F., & Blythe, P. W. (2005). Accurate judgments of intention from motion cues alone: A cross-cultural study. *Evolution and Human Behavior*, *26*(4), 313–331. https://doi.org/10.1016/j.evolhumbehav.2004.08.015

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using **lme4**. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bélanger, N. D., & Desrochers, S. (2001). Can 6-month-old infants process causality in different types of causal events? *British Journal of Developmental Psychology*, *19*(1), 11–21. https://doi.org/10.1348/026151001165930

Berent, I., Lennertz, T., Jun, J., Moreno, M. A., & Smolensky, P. (2008). Language universals in human brains. *Proceedings of the National Academy of Sciences*, *105*(14), 5321–5325. https://doi.org/10.1073/pnas.0801469105

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, *109*(9), 3253–3258. https://doi.org/10.1073/pnas.1113380109

Bergman, T. J., Beehner, J. C., Cheney, D. L., & Seyfarth, R. M. (2003). Hierarchical classification by rank and kinship in baboons. *Science*, *302*(5648), 1234–1236. https://doi.org/10.1126/science.1087513

Bickel, B., Witzlack-Makarevich, A., Choudhary, K. K., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2015). The neurophysiology of language processing shapes the evolution of grammar: Evidence from case marking. *PLOS ONE*, *10*(8), e0132819. https://doi.org/10.1371/journal.pone.0132819

Bowerman, M., & Choi, S. (2001). Shaping meanings for language: Universal and language-specific in the acquisition of spatial semantic categories. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 475–511). Cambridge University Press.

Bowerman, M., & Pederson, E. (1992). Topological relations picture series. In S. C. Levinson (Ed.), *Space stimuli kit 1.2* (p. 51). Max Planck Institute for Psycholinguistics. https://doi.org/10.17617/2.883589

Brocard, S., Wilson, V. A. D., Berton, C., Zuberbühler, K., & Bickel, B. (2024). A universal preference for animate agents in hominids. *iScience*, *27*(6). https://doi.org/10.1016/j.isci.2024.109996

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01116

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*, 1–28. https://doi.org/10.18637/jss.v080.i01

Bybee, J. (2006). Language change and universals. In *Linguistic universals* (pp. 179–212). Cambridge University Press.

Bybee, J. (2009). Language universals and usage-based theory. In *Language universals* (pp. 17–39). Oxford University Press.

Byrne, R. W. (2000). Evolution of primate cognition. *Cognitive Science*, *24*(3), 543–570. https://doi.org/10.1207/s15516709cog2403_8

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

Cacchione, T., & Krist, H. (2004). Recognizing impossible object relations: Intuitions about support in chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology*, *118*(2), 140–148. https://doi.org/10.1037/0735-7036.118.2.140

Casasola, M. (2008). The development of infants' spatial categories. *Current Directions in Psychological Science*, *17*(1), 21–25. https://doi.org/10.1111/j.1467-8721.2008.00541.x

Casasola, M., & Cohen, L. B. (2002). Infant categorization of containment, support and tight-fit spatial relationships. *Developmental Science*, *5*(2), 247–264.

Chemla, E., Dautriche, I., Buccola, B., & Fagot, J. (2019). Constraints on the lexicons of human languages have cognitive roots present in baboons (*Papio papio*). *Proceedings of the National Academy of Sciences*, *116*(30), 14926–14930. https://doi.org/10.1073/pnas.1907023116

Chen, Y., Trueswell, J., & Papafragou, A. (2024). Sources and goals in memory and language: Fragility and robustness in event representation. *Journal of Memory and Language*, *135*, 104475. https://doi.org/10.1016/j.jml.2023.104475

Cheney, D. L., & Seyfarth, R. M. (2005). Constraints and preadaptations in the earliest stages of language evolution. *The Linguistic Review*, *22*, 135–159. https://doi.org/10.1515/tlir.2005.22.2-4.135

Chomsky, N. (1957). *Syntactic structures*. De Gruyter Mouton. https://doi.org/10.1515/9783110218329

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, *31*(5), 489–509. https://doi.org/10.1017/S0140525X08004998

Christiansen, M. H., Collins, C., & Edelman, S. (Eds.). (2009). *Language universals*. Oxford University Press.

Cicchino, J. B., Aslin, R. N., & Rakison, D. H. (2011). Correspondences between what infants see and know about causal and self-propelled motion. *Cognition*, *118*(2), 171–192. https://doi.org/10.1016/j.cognition.2010.11.005

Clark, E. V. (2003). Languages and representations. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in Mind: Advances in the Study of Language and Thought* (pp. 17–24). The MIT Press. https://direct.mit.edu/books/edited-volume/1917/chapter/52668/Languages-and-Representations

Clark, E. V. (2004). How language acquisition builds on cognitive development. *Trends in Cognitive Sciences*, *8*(10), 472–478. https://doi.org/10.1016/j.tics.2004.08.012

Cohen, L. B., & Amsel, G. (1998). Precursors to infants' perception of the causality of a simple event. *Infant Behavior and Development*, *21*(4), 713–731. https://doi.org/10.1016/S0163-6383(98)90040-6

Cohen, L. B., & Oakes, L. M. (1993). How infants perceive a simple causal event. *Developmental Psychology*, *29*(3), 421–433. https://doi.org/10.1037/0012-1649.29.3.421

Cohn, N., & Paczynski, M. (2013). Prediction, events, and the advantage of Agents: The processing of semantic roles in visual narrative. *Cognitive Psychology*, *67*(3), 73–97. https://doi.org/10.1016/j.cogpsych.2013.07.002

Coppola, M., & Newport, E. L. (2005). Grammatical Subjects in home sign: Abstract linguistic structure in adult primary gesture systems without linguistic input. *Proceedings of the National Academy of Sciences*, *102*(52), 19249–19253. https://doi.org/10.1073/pnas.0509306102

Csibra, G., Gergely, G., Bíró, S., Koós, O., & Brockbank, M. (1999). Goal attribution without agency cues: The perception of 'pure reason' in infancy. *Cognition*, *72*(3), 237–267. https://doi.org/10.1016/S0010-0277(99)00039-6

Csibra, G., Hernik, M., Mascaro, O., Tatone, D., & Lengyel, M. (2016). Statistical treatment of looking-time data. *Developmental Psychology*, *52*(4), 521–536. http://dx.doi.org/10.1037/dev0000083

Culbertson, J., & Kirby, S. (2016). Simplicity and specificity in language: Domain-general biases have domain-specific effects. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01964

Culbertson, J., Schouwstra, M., & Kirby, S. (2020). From the world to word order: Deriving biases in noun phrase order from statistical properties of the world. *Language*, *96*(3), 696–717.

Dal Pesco, F., & Fischer, J. (2018). Greetings in male Guinea baboons and the function of rituals in complex social groups. *Journal of Human Evolution*, *125*, 87–98. https://doi.org/10.1016/j.jhevol.2018.10.007

Dautriche, I., Buccola, B., Berthet, M., Fagot, J., & Chemla, E. (2022). Evidence for compositionality in baboons (*Papio papio*) through the test case of negation. *Scientific Reports*, *12*(1), 19181. https://doi.org/10.1038/s41598-022-21143-1

Davidoff, J., & Fagot, J. (2010). Cross-species assessment of the linguistic origins of color categories. *Comparative Cognition & Behavior Reviews*, *5*. https://comparative-cognition-and-behavior-reviews.org/vol5_davidoff_fagot/

Davidson, G., Miller, R., Loissel, E., Cheke, L. G., & Clayton, N. S. (2017). The development of support intuitions and object causality in juvenile Eurasian jays (*Garrulus glandarius*). *Scientific Reports*, *7*(1), 40062. https://doi.org/10.1038/srep40062

Dépy, D., Fagot, J., & Vauclair, J. (1997). Categorisation of three-dimensional stimuli by humans and baboons: Search for prototype effects. *Behavioural Processes*, *39*(3), 299–306. https://doi.org/10.1016/S0376-6357(96)00757-7

Dépy, D., Fagot, J., & Vauclair, J. (1999). Processing of above/below categorical spatial relations by baboons (Papio papio). *Behavioural Processes*, *48*(1–2), 1–9. https://doi.org/10.1016/S0376-6357(99)00055-8

Deruelle, C., & Fagot, J. (1998). Visual search for global/local stimulus features in humans and baboons. *Psychonomic Bulletin & Review*, *5*(3), 476–481. https://doi.org/10.3758/BF03208825

Di Giorgio, E., Lunghi, M., Simion, F., & Vallortigara, G. (2017). Visual cues of motion that trigger animacy perception at birth: The case of self-propulsion. *Developmental Science*, *20*(4), e12394. https://doi.org/10.1111/desc.12394

Di Giorgio, E., Lunghi, M., Vallortigara, G., & Simion, F. (2021). Newborns' sensitivity to speed changes as a building block for animacy perception. *Scientific Reports*, *11*(1), 542. https://doi.org/10.1038/s41598-020-79451-3

Do, M. L., Papafragou, A., & Trueswell, J. (2022). Encoding motion events during language production: Effects of audience design and conceptual salience. *Cognitive Science*, *46*(1), e13077. https://doi.org/10.1111/cogs.13077

Dobel, C., Gumnior, H., Bölte, J., & Zwitserlood, P. (2007). Describing scenes hardly seen. *Acta Psychologica*, *125*(2), 129–143. https://doi.org/10.1016/j.actpsy.2006.07.004

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, *67*(3), 547. https://doi.org/10.2307/415037

Dryer, M. S. (2013). Order of Subject, Object and Verb. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Zenodo. https://doi.org/10.5281/zenodo.7385533

Dunn, M., Greenhill, S. J., Levinson, S. C., & Gray, R. D. (2011). Evolved structure of language shows lineage-specific trends in word-order universals. *Nature*, *473*(7345), 79–82. https://doi.org/10.1038/nature09923

Elman, J. L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science*, *33*(4), 547–582. https://doi.org/10.1111/j.1551-6709.2009.01023.x

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*(5), 429–448. https://doi.org/10.1017/S0140525X0999094X

Fagot, J., Barbet, I., Parron, C., & Deruelle, C. (2006). Amodal completion by baboons (*Papio papio*): Contribution of background depth cues. *Primates*, *47*(2), 145–150. https://doi.org/10.1007/s10329-005-0165-5

Fagot, J., Boë, L.-J., Berthomier, F., Claidière, N., Malassis, R., Meguerditchian, A., Rey, A., & Montant, M. (2019). The baboon: A model for the study of language evolution. *Journal of Human Evolution*, *126*, 39–50. https://doi.org/10.1016/j.jhevol.2018.10.006

Fagot, J., & Deruelle, C. (1997). Processing of global and local visual information and hemispheric specialization in humans (*Homo sapiens*) and baboons (*Papio papio*). *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 429–442. https://doi.org/10.1037/0096-1523.23.2.429

Fagot, J., Gullstrand, J., Kemp, C., Defilles, C., & Mekaouche, M. (2014). Effects of freely accessible computerized test systems on the spontaneous behaviors and stress level of Guinea baboons (*Papio papio*): Computerized testing and welfare in baboons. *American Journal of Primatology*, *76*(1), 56–64. https://doi.org/10.1002/ajp.22193

Fagot, J., Malassis, R., Medam, T., & Montant, M. (2017). Comparing human and nonhuman animal performance on domain-general functions: Towards a multiple bottleneck scenario of language evolution. In *Origins of Human Language: Continuities and Discontinuities with Nonhuman Primates* (pp. 329–365). Peter Lang D. https://doi.org/10.3726/b12405

Fagot, J., & Paleressompoulle, D. (2009). Automatic testing of cognitive performance in baboons maintained in social groups. *Behavior Research Methods*, *41*(2), 396–404. https://doi.org/10.3758/BRM.41.2.396

Fagot, J., & Thompson, R. K. R. (2011). Generalized relational matching by Guinea baboons (*Papio papio*) in two-by-two-Item analogy problems. *Psychological Science*, *22*(10), 1304–1309. https://doi.org/10.1177/0956797611422916

Fedurek, P., & Slocombe, K. E. (2011). Primate vocal communication: A useful tool for understanding human speech and language evolution? *Human Biology*, *83*(2), 153–173. https://doi.org/10.3378/027.083.0202

Ferretti, T. R., McRae, K., & Hatherell, A. (2001). Integrating verbs, situation schemas, and thematic role concepts. *Journal of Memory and Language*, *44*(4), 516–547. https://doi.org/10.1006/jmla.2000.2728

Fichtel, C., Dinter, K., & Kappeler, P. M. (2020). The lemur baseline: How lemurs compare to monkeys and apes in the Primate Cognition Test Battery. *PeerJ*, *8*, 1–26. https://doi.org/10.7717/peerj.10025

Fillmore, C. J. (1968). Lexical Entries for Verbs. *Foundations of Language*, *4*(4), 373–393.

Finger, H., Goeke, C., Diekamp, D., Standvoß, K., & König, P. (2017). LabVanced: A unified JavaScript framework for online studies. *International Conference on Computational Social Science*.

Fischer, J., Higham, J. P., Alberts, S. C., Barrett, L., Beehner, J. C., Bergman, T. J., Carter, A. J., Collins, A., Elton, S., Fagot, J., Ferreira da Silva, M. J., Hammerschmidt, K., Henzi, P., Jolly, C. J., Knauf, S., Kopp, G. H., Rogers, J., Roos, C., Ross, C., … Zinner, D. (2019). Insights into the evolution of social systems and species from baboon studies. *eLife*, *8*, e50989. https://doi.org/10.7554/eLife.50989

Fischer, J., Kopp, G. H., Dal Pesco, F., Goffe, A., Hammerschmidt, K., Kalbitzer, U., Klapproth, M., Maciej, P., Ndao, I., Patzelt, A., & Zinner, D. (2017). Charting the neglected West: The social system of Guinea baboons. *American Journal of Physical Anthropology*, *162*(S63), 15–31. https://doi.org/10.1002/ajpa.23144

Fischer, J., & Zinner, D. (2020). Introduction to special issue: Frontiers in baboon research. *Journal of Human Evolution*, *146*, 102822. https://doi.org/10.1016/j.jhevol.2020.102822

Fitch, W. T. (2010). *The Evolution of Language*. Cambridge University Press.

Fitch, W. T. (2019). Animal cognition and the evolution of human language: Why we cannot focus solely on communication. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1789), 20190046. https://doi.org/10.1098/rstb.2019.0046

Frankenhuis, W. E., House, B., Clark Barrett, H., & Johnson, S. P. (2013). Infants' perception of chasing. *Cognition*, *126*(2), 224–233. https://doi.org/10.1016/j.cognition.2012.10.001

Galazka, M., Bakker, M., Gredebäck, G., & Nyström, P. (2016). How social is the chaser? Neural correlates of chasing perception in 9-month-old infants. *Developmental Cognitive Neuroscience*, *19*, 270–278. https://doi.org/10.1016/j.dcn.2016.05.005

Galazka, M., & Nyström, P. (2016). Infants' preference for individual agents within chasing interactions. *Journal of Experimental Child Psychology*, *147*, 53–70. https://doi.org/10.1016/j.jecp.2016.02.010

Gao, T., Baker, C. L., Tang, N., Xu, H., & Tenenbaum, J. B. (2019). The cognitive architecture of perceived animacy: Intention, attention, and memory. *Cognitive Science*, *43*(8), e12775. https://doi.org/10.1111/cogs.12775

Gao, T., Newman, G. E., & Scholl, B. J. (2009). The psychophysics of chasing: A case study in the perception of animacy. *Cognitive Psychology*, *59*(2), 154–179. https://doi.org/10.1016/j.cogpsych.2009.03.001

Gärdenfors, P. (2024). Event structure, force dynamics and verb semantics. *Language Sciences*, *102*, 101610. https://doi.org/10.1016/j.langsci.2023.101610

Gentner, D., & Bowerman, M. (2010). Why some spatial semantic categories are harder to learn than others: The typological prevalence hypothesis. In J. Guo, E. Lieven, N. Budwig, S. Ervin-Tripp, K. Nakamura, & S. Özcaliskan (Eds.), *Crosslinguistic approaches to the psychology of language: Research in the tradition of Dan Isaac Slobin* (pp. 485–500). Psychology Press. https://doi.org/10.4324/9780203837887-51

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naïve theory of rational action. *Trends in Cognitive Sciences*, *7*(7), 287–292. https://doi.org/10.1016/S1364-6613(03)00128-1

Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, *56*(2), 165–193. https://doi.org/10.1016/0010-0277(95)00661-H

Gertner, Y., & Fisher, C. (2012). Predicted errors in children's early sentence comprehension. *Cognition*, *124*(1), 85–94. https://doi.org/10.1016/j.cognition.2012.03.010

Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences*, *23*(5), 389–407. https://doi.org/10.1016/j.tics.2019.02.003

Gibson, E., Piantadosi, S. T., Brink, K., Bergen, L., Lim, E., & Saxe, R. (2013). A Noisy-Channel Account of Crosslinguistic Word-Order Variation. *Psychological Science*, *24*(7), 1079–1088. https://doi.org/10.1177/0956797612463705

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*(1), 3–55.

Gleitman, L., & Papafragou, A. (2005). Relations between language and thought. In K. J. Holyoak & R. G. Morrison (Eds.), *Cambridge handbook of thinking and reasoning* (pp. 633–661). Cambridge University Press.

Goldin-Meadow, S., So, W. C., Özyürek, A., & Mylander, C. (2008). The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences*, *105*(27), 9163–9168. https://doi.org/10.1073/pnas.0710060105

Golinkoff, R. M. (1975). Semantic development in infants: The concepts of agent and recipient. *Merrill-Palmer Quarterly of Behavior and Development*, *21*(3), 181–193.

Goto, K., Lea, S. E., & Dittrich, W. H. (2002). Discrimination of intentional and random motion paths by pigeons. *Animal Cognition*, *5*(3), 119–127. https://doi.org/10.1007/s10071-002-0139-6

Goupil, N., Papeo, L., & Hochmann, J.-R. (2022). Visual perception grounding of social cognition in preverbal infants. *Infancy*, *27*(2), 210–231. https://doi.org/10.1111/infa.12453

Goupil, N., Rayson, H., Serraille, É., Massera, A., Ferrari, P. F., Hochmann, J.-R., & Papeo, L. (2024). Visual preference for socially relevant spatial relations in humans and monkeys. *Psychological Science*, *35*(6), 681–693. https://doi.org/10.1177/09567976241242995

Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In *Universals of language* (pp. 40–70). MIT Press.

Hafri, A., & Firestone, C. (2021). The perception of relations. *Trends in Cognitive Sciences*, *25*(6), 475–492. https://doi.org/10.1016/j.tics.2021.01.006

Hafri, A., Papafragou, A., & Trueswell, J. C. (2013). Getting the gist of events: Recognition of two-participant actions from brief displays. *Journal of Experimental Psychology: General*, *142*(3), 880–905. https://doi.org/10.1037/a0030045

Hafri, A., Trueswell, J. C., & Strickland, B. (2018). Encoding of event roles from visual scenes is rapid, spontaneous, and interacts with higher-level visual processing. *Cognition*, *175*, 36–52. https://doi.org/10.1016/j.cognition.2018.02.011

Hall, M. L., Ahn, Y. D., Mayberry, R. I., & Ferreira, V. S. (2015). Production and comprehension show divergent constituent order preferences: Evidence from elicited pantomime. *Journal of Memory and Language*, *81*, 16–33. https://doi.org/10.1016/j.jml.2014.12.003

Hall, M. L., Mayberry, R. I., & Ferreira, V. S. (2013). Cognitive constraints on constituent order: Evidence from elicited pantomime. *Cognition*, *129*(1), 1–17. https://doi.org/10.1016/j.cognition.2013.05.004

Hard, B. M., Meyer, M., & Baldwin, D. (2019). Attention reorganizes as structure is detected in dynamic action. *Memory & Cognition*, *47*(1), 17–32. https://doi.org/10.3758/s13421-018-0847-z

Hard, B. M., Recchia, G., & Tversky, B. (2011). The shape of action. *Journal of Experimental Psychology: General*, *140*(4), 586–604. https://doi.org/10.1037/a0024310

Hartig, F. (2022). *DHARMa: Residual diagnostics for hierarchical (multi-level / mixed) tegression models* (Version R package version 0.4.6) [Computer software]. https://CRAN.R-project.org/package=DHARMa

Hauser, M. D. (1998). A nonhuman primate's expectations about object motion and destination: The importance of self-propelled movement and animacy. *Developmental Science*, *1*(1), 31–37. https://doi.org/10.1111/1467-7687.00009

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*(5598), 1569–1579. https://doi.org/10.1126/science.298.5598.1569

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology*, *57*(2), 243. https://doi.org/10.2307/1416950

Herrmann, E., Call, J., Hernández-Lloreda, M. V., Hare, B., & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *Science*, *317*(5843), 1360–1366. https://doi.org/10.1126/science.1146282

Herrmann, E., Hare, B., Call, J., & Tomasello, M. (2010). Differences in the cognitive skills of bonobos and chimpanzees. *PLoS ONE*, *5*(8), 2–5. https://doi.org/10.1371/journal.pone.0012438

Hespos, S. J., Grossman, S. R., & Saylor, M. M. (2010). Infants' ability to parse continuous actions: Further evidence. *Neural Networks*, *23*(8), 1026–1032. https://doi.org/10.1016/j.neunet.2010.07.010

Hespos, S. J., Saylor, M. M., & Grossman, S. R. (2009). Infants' ability to parse continuous actions. *Developmental Psychology*, *45*(2), 575–585. https://doi.org/10.1037/a0014145

Hespos, S. J., & Spelke, E. S. (2004). Conceptual precursors to language. *Nature*, *430*(6998), 453–456. https://doi.org/10.1038/nature02634

Hockett, C. F. (1960). The origin of speech. *Scientific American*, *203*(3), 88–97.

Hofrichter, R., & Rutherford, M. D. (2019). Early attentional capture of animate motion: 4-Year-Olds show a pop-out effect for chasing stimuli. *Perception*, *48*(3), 228–236. https://doi.org/10.1177/0301006619828256

Hurford, J. R. (2007). *Language in the Light of Evolution: Volume 1, The Origins of Meaning*. Oxford University Press.

Isasi-Isasmendi, A., Andrews, C., Flecken, M., Laka, I., Daum, M. M., Meyer, M., Bickel, B., & Sauppe, S. (2023). The agent preference in visual event apprehension. *Open Mind*, *7*, 240–282. https://doi.org/10.1162/opmi_a_00083

Jackendoff, R. (1999). Possible stages in the evolution of the language capacity. *Trends in Cognitive Sciences*, *3*(7), 272–279. https://doi.org/10.1016/S1364-6613(99)01333-9

Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford University Press UK.

Janmaat, K. R. L., Polansky, L., Ban, S. D., & Boesch, C. (2014). Wild chimpanzees plan their breakfast time, type, and location. *Proceedings of the National Academy of Sciences*, *111*(46), 16343–16348. https://doi.org/10.1073/pnas.1407524111

Jensen, G., Alkan, Y., Ferrera, V. P., & Terrace, H. S. (2019). Reward associations do not explain transitive inference performance in monkeys. *Science Advances*. https://doi.org/10.1126/sciadv.aaw2089

Jolly, C. J. (1970). The seed-eaters: A new model of hominid differentiation based on a baboon analogy. *Man*, *5*, 1–26.

Joly, M., Micheletta, J., De Marco, A., Langermans, J. A., Sterck, E. H. M., & Waller, B. M. (2017). Comparing physical and social cognitive skills in macaque species with different degrees of social tolerance. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1862). https://doi.org/10.1098/rspb.2016.2738

Kabata, K. (2013). Goal–source asymmetry and crosslinguistic grammaticalization patterns: A cognitive-typological approach. *Language Sciences*, *36*, 78–89. https://doi.org/10.1016/j.langsci.2012.03.021

Kano, F., & Call, J. (2014). Great apes generate goal-based action predictions: An eye-tracking study. *Psychological Science*, *25*(9), 1691–1698. https://doi.org/10.1177/0956797614536402

Kemmerer, D. (2012). The cross-linguistic prevalence of SOV and SVO word orders reflects the sequential and hierarchical representation of action in Broca's area. *Language and Linguistics Compass*, *6*(1), 50–66. https://doi.org/10.1002/lnc3.322

Kemp, C., Rey, A., Legou, T., Boë, L.-J., Berthommier, F., Becker, Y., & Fagot, J. (2017). Vocal repertoire of captive Guinea baboons (*Papio papio*). In *Origins of Human Language: Continuities and Discontinuities with Nonhuman Primates* (pp. 15–58). Peter Lang D. https://doi.org/10.3726/b12405

King, G. E. (2022). Baboon perspectives on the ecology and behavior of early human ancestors. *Proceedings of the National Academy of Sciences*, *119*(45), e2116182119. https://doi.org/10.1073/pnas.2116182119

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*(31), 10681–10686. https://doi.org/10.1073/pnas.0707835105

Kominsky, J. F., Baker, L., Keil, F. C., & Strickland, B. (2021). Causality and continuity close the gaps in event representations. *Memory & Cognition*, *49*(3), 518–531. https://doi.org/10.3758/s13421-020-01102-9

Kominsky, J. F., & Carey, S. (2024). Infants' representations of Michottean triggering events. *Cognition*, *250*, 105844. https://doi.org/10.1016/j.cognition.2024.105844

Kominsky, J. F., Lucca, K., Thomas, A. J., Frank, M. C., & Hamlin, J. K. (2022). Simplicity and validity in infant research. *Cognitive Development*, *63*, 101213. https://doi.org/10.1016/j.cogdev.2022.101213

Kominsky, J. F., & Scholl, B. J. (2020). Retinotopic adaptation reveals distinct categories of causal perception. *Cognition*, *203*, 104339. https://doi.org/10.1016/j.cognition.2020.104339

Kominsky, J. F., Strickland, B., Wertz, A. E., Elsner, C., Wynn, K., & Keil, F. C. (2017). Categories and constraints in causal perception. *Psychological Science*, *28*(11), 1649–1662. https://doi.org/10.1177/0956797617719930

Kuperberg, G. R. (2021). Tea with milk? A hierarchical generative framework of sequential event comprehension. *Topics in Cognitive Science*, *13*(1), 256–298. https://doi.org/10.1111/tops.12518

Lakusta, L., & Landau, B. (2005). Starting at the end: The importance of goals in spatial language. *Cognition*, *96*(1), 1–33. https://doi.org/10.1016/j.cognition.2004.03.009

Lakusta, L., Wagner, L., O'Hearn, K., & Landau, B. (2007). Conceptual foundations of spatial language: Evidence for a goal bias in infants. *Language Learning and Development*, *3*(3), 179–197. https://doi.org/10.1080/15475440701360168

Lameira, A. R., Hardus, M. E., Kowalsky, B., de Vries, H., Spruijt, B. M., Sterck, E. H. M., Shumaker, R. W., & Wich, S. A. (2013). Orangutan (*Pongo spp.*) whistling and implications for the emergence of an open-ended call repertoire: A replication and extension. *The Journal of the Acoustical Society of America*, *134*(3), 2326–2335. https://doi.org/10.1121/1.4817929

Landau, B., & Jackendoff, R. (1993). "What" and "where" in spatial language and spatial cognition. *Behavioral and Brain Sciences*, *16*(2), 217–238. https://doi.org/10.1017/S0140525X00029733

Langus, A., & Nespor, M. (2010). Cognitive systems struggling for word order. *Cognitive Psychology*, *60*(4), 291–318. https://doi.org/10.1016/j.cogpsych.2010.01.004

le Roux, A., Snyder-Mackler, N., Roberts, E. K., Beehner, J. C., & Bergman, T. J. (2013). Evidence for tactical concealment in a wild primate. *Nature Communications*, *4*(1), 1462. https://doi.org/10.1038/ncomms2468

Lemaire, B. S., Rosa-Salva, O., Fraja, M., Lorenzi, E., & Vallortigara, G. (2022). Spontaneous preference for unpredictability in the temporal contingencies between agents' motion in naïve domestic chicks. *Proceedings of the Royal Society B: Biological Sciences*, *289*, 20221622. https://doi.org/10.1098/rspb.2022.1622

Lemasson, A., Ouattara, K., Petit, E. J., & Zuberbühler, K. (2011). Social learning of vocal structure in a nonhuman primate? *BMC Evolutionary Biology*, *11*(1), 362. https://doi.org/10.1186/1471-2148-11-362

Lenth, R. V. (2022). *emmeans: Estimated marginal means, aka least-squares means* (Version 1.9.0) [Computer software]. https://CRAN.R-project.org/package=emmeans

Leslie, A. M. (1995). A theory of agency. In *Causal cognition: A multidisciplinary debate* (pp. 121–149). Clarendon Press/Oxford University Press.

Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition*, *25*(3), 265–288. https://doi.org/10.1016/S0010-0277(87)80006-9

Levelt, W. J. M. (1993). *Speaking: From Intention to Articulation*. MIT Press.

Levine, D., Buchsbaum, D., Hirsh-Pasek, K., & Golinkoff, R. M. (2019). Finding events in a continuous world: A developmental account. *Developmental Psychobiology*, *61*(3), 376–389. https://doi.org/10.1002/dev.21804

Levinson, S. C., & Wilkins, D. (Eds.). (2006). *Grammars of space: Explorations in cognitive diversity*. Cambridge University Press.

Levinson, S., Meira, S., & Group, T. L. and C. (2003). 'Natural concepts' in the spatial topological domain-adpositional meanings in crosslinguistic perspective: An exercise in semantic typology. *Language*, *79*(3), 485–516.

Lind, J., Enquist, M., & Ghirlanda, S. (2015). Animal memory: A review of delayed matching-to-sample data. *Behavioural Processes*, *117*, 52–58. https://doi.org/10.1016/j.beproc.2014.11.019

Luo, Y., Kaufman, L., & Baillargeon, R. (2009). Young infants' reasoning about physical events involving inert and self-propelled objects. *Cognitive Psychology*, *58*(4), 441–486. https://doi.org/10.1016/j.cogpsych.2008.11.001

Lyn, H., Russell, J. L., Leavens, D. A., Bard, K. A., Boysen, S. T., Schaeffer, J. A., & Hopkins, W. D. (2014). Apes communicate about absent and displaced objects: Methodology matters. *Animal Cognition*, *17*(1), 85–94. https://doi.org/10.1007/s10071-013-0640-0

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*(3), 177–197. https://doi.org/10.1111/j.1439-0310.1993.tb00988.x

Maciej, P., Ndao, I., Hammerschmidt, K., & Fischer, J. (2013). Vocal communication in a complex multi-level society: Constrained acoustic structure and flexible call usage in Guinea baboons. *Frontiers in Zoology*, *10*(1), 58. https://doi.org/10.1186/1742-9994-10-58

MacWhinney, B. (1977). Starting points. *Language*, *53*(1), 152–168. https://doi.org/10.2307/413059

Magliano, J. P., Radvansky, G. A., Forsythe, J. C., & Copeland, D. E. (2014). Event segmentation during first-person continuous events. *Journal of Cognitive Psychology*, *26*(6), 649–661. https://doi.org/10.1080/20445911.2014.930042

Majid, A., Bowerman, M., Kita, S., Haun, D. B. M., & Levinson, S. C. (2004). Can language restructure cognition? The case for space. *Trends in Cognitive Sciences*, *8*(3), 108–114. https://doi.org/10.1016/j.tics.2004.01.003

Maldonado, M., & Culbertson, J. (2022). Person of interest: Experimental investigations into the learnability of person systems. *Linguistic Inquiry*, *53*(2), 295–336. https://doi.org/10.1162/ling_a_00406

Malt, B. C., Sloman, S. A., & Gennari, S. P. (2003). Speaking versus thinking about objects and actions. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in Mind: Advances in the Study of Language and Thought*. The MIT Press. https://direct.mit.edu/books/edited-volume/1917/chapter/52673/Speaking-versus-Thinking-about-Objects-and-Actions

Mandler, J. M. (1992). How to build a baby: II. Conceptual primitives. *Psychological Review*, *99*(4), 587–604. https://doi.org/10.1037/0033-295x.99.4.587

Martin, A., Adger, D., Abels, K., Kanampiu, P., & Culbertson, J. (2024). A universal cognitive bias in word order: Evidence from speakers whose language goes against it. *Psychological Science*, *35*(3), 304–311. https://doi.org/10.1177/09567976231222836

Mascalzoni, E., Regolin, L., & Vallortigara, G. (2010). Innate sensitivity for self-propelled causal agency in newly hatched chicks. *Proceedings of the National Academy of Sciences*, *107*(9), 4483–4485. https://doi.org/10.1073/pnas.0908792107

Mascalzoni, E., Regolin, L., Vallortigara, G., & Simion, F. (2013). The cradle of causal reasoning: Newborns' preference for physical causality. *Developmental Science*, *16*(3), 327–335. https://doi.org/10.1111/desc.12018

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

Matsuno, T., & Tomonaga, M. (2017). Causal capture effects in chimpanzees (*Pan troglodytes*). *Cognition*, *158*, 153–164. https://doi.org/10.1016/j.cognition.2016.10.023

Matsuzawa, T. (1985). Use of numbers by a chimpanzee. *Nature*, *315*(6014), 57–59. https://doi.org/10.1038/315057a0

McDonough, L., Choi, S., & Mandler, J. M. (2003). Understanding spatial relations: Flexible infants, lexical adults. *Cognitive Psychology*, *46*(3), 229–259. https://doi.org/10.1016/S0010-0285(02)00514-5

Meewis, F., Barezzi, I., Joël, F., Claidière, N., & Dautriche, I. (2024). *Preprint: A comparative study of causal perception in Guinea baboons (Papio papio) and human adults*. OSF. https://doi.org/10.31219/osf.io/mejwp

Meir, I., Aronoff, M., Börstell, C., Hwang, S.-O., Ilkbasaran, D., Kastner, I., Lepic, R., Lifshitz Ben-Basat, A., Padden, C., & Sandler, W. (2017). The effect of being human and the basis of grammatical word order: Insights from novel communication systems and young sign languages. *Cognition*, *158*, 189–207. https://doi.org/10.1016/j.cognition.2016.10.011

Meyerhoff, H. S., Schwan, S., & Huff, M. (2014a). Interobject spacing explains the attentional bias toward interacting objects. *Psychonomic Bulletin & Review*, *21*(2), 412–417. https://doi.org/10.3758/s13423-013-0496-y

Meyerhoff, H. S., Schwan, S., & Huff, M. (2014b). Perceptual animacy: Visual search for chasing objects among distractors. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(2), 702–717. https://doi.org/10.1037/a0034846

Michotte, A. (1963). *The Perception of Causality*. Routledge. https://doi.org/10.4324/9781315519050

Moors, P., Wagemans, J., & de-Wit, L. (2017). Causal events enter awareness faster than non-causal events. *PeerJ*, *5*, e2932. https://doi.org/10.7717/peerj.2932

Muentener, P., & Bonawitz, E. (2017). The development of causal reasoning. In M. R. Waldmann (Ed.), *The Oxford Handbook of Causal Reasoning* (p. 0). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199399550.013.40

Muentener, P., & Carey, S. (2010). Infants' causal representations of state change events. *Cognitive Psychology*, *61*(2), 63–86. https://doi.org/10.1016/j.cogpsych.2010.02.001

Mulcahy, N. J., & Call, J. (2006). Apes save tools for future use. *Science*, *312*(5776), 1038–1040. https://doi.org/10.1126/science.1125456

Murai, C., Tanaka, M., & Sakagami, M. (2011). Physical intuitions about support relations in monkeys (*Macaca fuscata*) and apes (*Pan troglodytes*). *Journal of Comparative Psychology*, *125*(2), 216–226. https://doi.org/10.1037/a0022099

Myowa-Yamakoshi, M., Scola, C., & Hirata, S. (2012). Humans and chimpanzees attend differently to goal-directed actions. *Nature Communications*, *3*(1), 693. https://doi.org/10.1038/ncomms1695

Napoli, D. J., & Sutton-Spence, R. (2014). Order of the major constituents in sign languages: Implications for all language. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00376

Needham, A., & Baillargeon, R. (1993). Intuitions about support in 4.5-month-old infants. *Cognition*, *47*(2), 121–148. https://doi.org/10.1016/0010-0277(93)90002-D

New, J., Cosmides, L., & Tooby, J. (2007). Category-specific attention for animals reflects ancestral priorities, not expertise. *Proceedings of the National Academy of Sciences*, *104*(42), 16598–16603. https://doi.org/10.1073/pnas.0703913104

New, J., Schultz, R. T., Wolf, J., Niehaus, J. L., Klin, A., German, T. C., & Scholl, B. J. (2010). The scope of social attention deficits in autism: Prioritized orienting to people and animals in static natural scenes. *Neuropsychologia*, *48*(1), 51–59. https://doi.org/10.1016/j.neuropsychologia.2009.08.008

Newtson, D., & Engquist, G. (1976). The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology*, *12*(5), 436–450. https://doi.org/10.1016/0022-1031(76)90076-7

Oakes, L. M. (1994). Development of infants' use of continuity cues in their perception of causality. *Developmental Psychology*, *30*, 869–879. https://doi.org/10.1037/0012-1649.30.6.869

Oakes, L. M. (2010). Using habituation of looking time to assess mental processes in infancy. *Journal of Cognition and Development*, *11*(3), 255–268. https://doi.org/10.1080/15248371003699977

Oakes, L. M. (2017). Sample size, statistical power, and false conclusions in infant looking-time research. *Infancy*, *22*(4), 436–469. https://doi.org/10.1111/infa.12186

Oakes, L. M., & Cohen, L. B. (1990). Infant perception of a causal event. *Cognitive Development*, *5*(2), 193–207. https://doi.org/10.1016/0885-2014(90)90026-P

O'Connell, S., & Dunbar, R. I. M. (2005). The perception of causality in chimpanzees (*Pan spp.*). *Animal Cognition*, *8*(1), 60–66. https://doi.org/10.1007/s10071-004-0231-1

Ohl, S., & Rolfs, M. (2023). *Visual routines for detecting causal interactions are tuned to motion direction* [Preprint]. Neuroscience. https://doi.org/10.1101/2023.08.22.554237

Overduin-de Vries, A. M., Bakker, F. A. A., Spruijt, B. M., & Sterck, E. H. M. (2016). Male long-tailed macaques (*Macaca fascicularis*) understand the target of facial threat. *American Journal of Primatology*, *78*(7), 720–730. https://doi.org/10.1002/ajp.22536

Packer, M. J., & Moreno-Dulcey, F. A. (2022). Theory of puppets?: A critique of the use of puppets as stimulus materials in psychological research with young children. *Cognitive Development*, *61*, 101146. https://doi.org/10.1016/j.cogdev.2021.101146

Papafragou, A., & Grigoroglou, M. (2019). The role of conceptualization during language production: Evidence from event encoding. *Language, Cognition and Neuroscience*, *34*(9), 1117–1128. https://doi.org/10.1080/23273798.2019.1589540

Papeo, L. (2020). Twos in human visual perception. *Cortex*, *132*, 473–478. https://doi.org/10.1016/j.cortex.2020.06.005

Papeo, L., & Abassi, E. (2019). Seeing social events: The visual specialization for dyadic human–human interactions. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(7), 877–888. https://doi.org/10.1037/xhp0000646

Papeo, L., Stein, T., & Soto-Faraco, S. (2017). The two-body inversion effect. *Psychological Science*, *28*(3), 369–379. https://doi.org/10.1177/0956797616685769

Papeo, L., Vettori, S., Serraille, E., Odin, C., Rostami, F., & Hochmann, J.-R. (2024). Abstract thematic roles in infants' representation of social events. *Current Biology*. https://doi.org/10.1016/j.cub.2024.07.081

Pashler, H., & Wagenmakers, E. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, *7*(6), 528–530. https://doi.org/10.1177/1745691612465253

Penn, D. C., & Povinelli, D. J. (2007). Causal cognition in human and nonhuman animals: A comparative, critical review. *Annual Review of Psychology*, *58*(1), 97–118. https://doi.org/10.1146/annurev.psych.58.110405.085555

Phillips, W., & Santos, L. R. (2007). Evidence for kind representations in the absence of language: Experiments with rhesus monkeys (*Macaca mulatta*). *Cognition*, *102*(3), 455–463. https://doi.org/10.1016/j.cognition.2006.01.009

Premack, D. (1976). Language and intelligence in ape and man: How much is the gap between human and animal intelligence narrowed by recent demonstrations of language in chimpanzees? *American Scientist*, *64*(6), 674–683.

Prinzmetal, W., McCool, C., & Park, S. (2005). Attention: Reaction time and accuracy reveal different mechanisms. *Journal of Experimental Psychology: General*, *134*(1), 73–92. https://doi.org/10.1037/0096-3445.134.1.73

Quilty-Dunn, J., Porot, N., & Mandelbaum, E. (2023). The best game in town: The reemergence of the language-of-thought hypothesis across the cognitive sciences. *Behavioral and Brain Sciences*, *46*, e261. https://doi.org/10.1017/S0140525X22002849

Quinn, P. C., Cummins, M., Kase, J., Martin, E., & Weissman, S. (1996). Development of categorical representations for above and below spatial relations in 3- to 7-month-old infants. *Developmental Psychology*, *32*(5), 942–950.

Quinn, P. C., Norris, C. M., Pasko, R. N., Schmader, T. M., & Mash, C. (1999). Formation of a categorical representation for the spatial relation between by 6- to 7-month-old infants. *Visual Cognition*, *6*(5), 569–585. https://doi.org/10.1080/135062899394948

R Core Team. (2021). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/.

Radvansky, G. A., Tamplin, A. K., Armendarez, J., & Thompson, A. N. (2014). Different Kinds of Causality in Event Cognition. *Discourse Processes*, *51*(7), 601–618. https://doi.org/10.1080/0163853X.2014.903366

Radvansky, G. A., & Zacks, J. M. (2011). Event perception. *WIREs Cognitive Science*, *2*(6), 608–620. https://doi.org/10.1002/wcs.133

Rakison, D. H., & Krogh, L. (2012). Does causal action facilitate causal perception in infants younger than 6 months of age? *Developmental Science*, *15*(1), 43–53. https://doi.org/10.1111/j.1467-7687.2011.01096.x

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, *8*(5), 368–373. https://doi.org/10.1111/j.1467-9280.1997.tb00427.x

Reynolds, J. R., Zacks, J. M., & Braver, T. S. (2007). A computational model of event segmentation from perceptual prediction. *Cognitive Science*, *31*(4), 613–643. https://doi.org/10.1080/15326900701399913

Rips, L. J. (2011). Causation from perception. *Perspectives on Psychological Science*, *6*(1), 77–97. https://doi.org/10.1177/1745691610393525

Rissman, L., & Goldin-Meadow, S. (2017). The development of causal structure without a language model. *Language Learning and Development*, *13*(3), 286–299. https://doi.org/10.1080/15475441.2016.1254633

Rissman, L., & Majid, A. (2019). Thematic roles: Core knowledge or linguistic construct? *Psychonomic Bulletin & Review*, *26*(6), 1850–1869. https://doi.org/10.3758/s13423-019-01634-5

Rochat, M. J., Serra, E., Fadiga, L., & Gallese, V. (2008). The evolution of social cognition: Goal familiarity shapes monkeys' action understanding. *Current Biology*, *18*(3), 227–232. https://doi.org/10.1016/j.cub.2007.12.021

Rochat, P., Morgan, R., & Carpenter, M. (1997). Young infants' sensitivity to movement information specifying social causality. *Cognitive Development*, *12*, 537–561.

Rochat, P., Striano, T., & Morgan, R. (2004). Who is doing what to whom? Young infants' developing sense of social causality in animated displays. *Perception*, *33*(3), 355–369. https://doi.org/10.1068/p3389

Rogers, J., Raveendran, M., Harris, R. A., Mailund, T., Leppälä, K., Athanasiadis, G., Schierup, M. H., Cheng, J., Munch, K., Walker, J. A., Konkel, M. K., Jordan, V., Steely, C. J., Beckstrom, T. O., Bergey, C., Burrell, A., Schrempf, D., Noll, A., Kothe, M., … Baboon Genome Analysis Consortium. (2019). The comparative genomics and complex population history of *Papio* baboons. *Science Advances*, *5*(1), eaau6947. https://doi.org/10.1126/sciadv.aau6947

Rolfs, M., Dambacher, M., & Cavanagh, P. (2013). Visual adaptation of the perception of causality. *Current Biology*, *23*(3), 250–254. https://doi.org/10.1016/j.cub.2012.12.017

Rosa-Salva, O., Grassi, M., Lorenzi, E., Regolin, L., & Vallortigara, G. (2016). Spontaneous preference for visual cues of animacy in naïve domestic chicks: The case of speed changes. *Cognition*, *157*, 49–60. https://doi.org/10.1016/j.cognition.2016.08.014

Sandler, W., Meir, I., Padden, C., & Aronoff, M. (2005). The emergence of grammar: Systematic structure in a new language. *Proceedings of the National Academy of Sciences*, *102*(7), 2661–2665. https://doi.org/10.1073/pnas.0405448102

Santos, L. R. (2004). 'Core Knowledges': A dissociation between spatiotemporal knowledge and contact-mechanics in a non-human primate? *Developmental Science*, *7*(2), 167–174. https://doi.org/10.1111/j.1467-7687.2004.00335.x

Santos, L. R., Seelig, D., & Hauser, M. D. (2006). Cotton-top tamarins' (*Saguinus oedipus*) expectations about occluded objects: A dissociation between looking and reaching tasks. *Infancy*, *9*(2), 147–171. https://doi.org/10.1207/s15327078in0902_4

Saxe, R., & Carey, S. (2006). The perception of causality in infancy. *Acta Psychologica*, *123*(1–2), 144–165. https://doi.org/10.1016/j.actpsy.2006.05.005

Saxe, R., Tenenbaum, J. B., & Carey, S. (2005). Secret agents: Inferences about hidden causes by 10- and 12-month-old infants. *Psychological Science*, *16*(12), 995–1001. https://doi.org/10.1111/j.1467-9280.2005.01649.x

Saylor, M. M., Baldwin, D. A., Baird, J. A., & LaBounty, J. (2007). Infants' on-line segmentation of dynamic human action. *Journal of Cognition and Development*, *8*(1), 113–128. https://doi.org/10.1080/15248370709336996

Schafroth, J. L., Basile, B. M., Martin, A., & Murray, E. A. (2021). No evidence that monkeys attribute mental states to animated shapes in the Heider–Simmel videos. *Scientific Reports*, *11*(1), Article 1. https://doi.org/10.1038/s41598-021-82702-6

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*(10), e76674. https://doi.org/10.1371/journal.pone.0076674

Schlenker, P., Bonnet, M., Lamberton, J., Lamberton, J., Chemla, E., Santoro, M., & Geraci, C. (2024). Iconic syntax: Sign language classifier predicates and gesture sequences. *Linguistics and Philosophy*, *47*(1), 77–147. https://doi.org/10.1007/s10988-023-09388-z

Schlenker, P., Chemla, E., Schel, A. M., Fuller, J., Gautier, J.-P., Kuhn, J., Veselinović, D., Arnold, K., Cäsar, C., Keenan, S., Lemasson, A., Ouattara, K., Ryder, R., & Zuberbühler, K. (2016). Formal monkey linguistics. *Theoretical Linguistics*, *42*(1–2), 1–90. https://doi.org/10.1515/tl-2016-0001

Schlenker, P., Chemla, E., & Zuberbühler, K. (2016). What do monkey calls mean? *Trends in Cognitive Sciences*, *20*(12), 894–904. https://doi.org/10.1016/j.tics.2016.10.004

Schlenker, P., Coye, C., Leroux, M., & Chemla, E. (2023). The ABC-D of animal linguistics: Are syntax and compositionality for real? *Biological Reviews*, *98*(4), 1142–1159. https://doi.org/10.1111/brv.12944

Schloegl, C., & Fischer, J. (2017). *Causal reasoning in non-human animals* (M. R. Waldmann, Ed.; Vol. 1). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199399550.013.36

Schlottmann, A., & Shanks, D. R. (1992). Evidence for a distinction between judged and perceived causality. *The Quarterly Journal of Experimental Psychology*, *44*(2), 321–342. https://doi.org/10.1080/02724989243000055

Schmitt, V., Pankau, B., & Fischer, J. (2012). Old world monkeys compare to apes in the primate cognition test battery. *PLoS ONE*, *7*(4). https://doi.org/10.1371/journal.pone.0032024

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, *4*(8), 299–309. https://doi.org/10.1016/S1364-6613(00)01506-0

Schouwstra, M., & de Swart, H. (2014). The semantic origins of word order. *Cognition*, *131*(3), 431–436. https://doi.org/10.1016/j.cognition.2014.03.004

Seed, A., Seddon, E., Greene, B., & Call, J. (2012). Chimpanzee 'folk physics': Bringing failures into focus. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1603), 2743–2752. https://doi.org/10.1098/rstb.2012.0222

Seyfarth, R. M., & Cheney, D. L. (2008). Primate social knowledge and the origins of language. *Mind & Society*, *7*(1), 129–142. https://doi.org/10.1007/s11299-007-0038-2

Seyfarth, R. M., & Cheney, D. L. (2014). The evolution of language from social cognition. *Current Opinion in Neurobiology*, *28*, 5–9. https://doi.org/10.1016/j.conb.2014.04.003

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980). Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication. *Science*, *210*(4471), 801–803. https://doi.org/10.1126/science.7433999

Shiffman, D. (2012). Chapter 6: Autonomous agents. In *The nature of code*. https://natureofcode.com/autonomous-agents/

Simion, F., Di Giorgio, E., Leo, I., & Bardi, L. (2011). The processing of social stimuli in early infancy: From faces to biological motion perception. In O. Braddick, J. Atkinson, & G. M. Innocenti (Eds.), *Progress in Brain Research* (Vol. 189, pp. 173–193). Elsevier. https://doi.org/10.1016/B978-0-444-53884-0.00024-5

Slobin, D. I. (1987). Thinking for speaking. *Annual Meeting of the Berkeley Linguistics Society*, 435–445. https://doi.org/10.3765/bls.v13i0.1826

Slocombe, K. E., Kaller, T., Call, J., & Zuberbühler, K. (2010). Chimpanzees extract social information from agonistic screams. *PLoS ONE*, *5*(7), e11473. https://doi.org/10.1371/journal.pone.0011473

Smith, P. L., & Ratcliff, R. (2009). An integrated theory of attention and decision making in visual signal detection. *Psychological Review*, *116*(2), 283–317. https://doi.org/10.1037/a0015156

Sommerville, J. A., Woodward, A. L., & Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. *Cognition*, *96*(1), B1–B11. https://doi.org/10.1016/j.cognition.2004.07.004

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review*, *99*, 605–632. https://doi.org/10.1037/0033-295X.99.4.605

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*(1), 89–96. https://doi.org/10.1111/j.1467-7687.2007.00569.x

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and Cognition* (2nd edition). Wiley-Blackwell.

Spinozzi, G., Lubrano, G., & Truppa, V. (2004). Categorization of above and below spatial relations by tufted capuchin monkeys (*Cebus apella*). *Journal of Comparative Psychology*, *118*(4), 403–412. https://doi.org/10.1037/0735-7036.118.4.403

Springer, M. S., Meredith, R. W., Gatesy, J., Emerling, C. A., Park, J., Rabosky, D. L., Stadler, T., Steiner, C., Ryder, O. A., Janečka, J. E., Fisher, C. A., & Murphy, W. J. (2012). Macroevolutionary dynamics and historical biogeography of primate diversification inferred from a species supermatrix. *PLOS ONE*, *7*(11), e49521. https://doi.org/10.1371/journal.pone.0049521

Stahl, A. E., Romberg, A. R., Roseberry, S., Golinkoff, R. M., & Hirsh-Pasek, K. (2014). Infants segment continuous events using transitional probabilities. *Child Development*, *85*(5), 1821–1826. https://doi.org/10.1111/cdev.12247

Strickland, B. (2017). Language reflects "core" cognition: A new theory about the origin of cross-linguistic regularities. *Cognitive Science*, *41*(1), 70–101. https://doi.org/10.1111/cogs.12332

Strickland, B., & Chemla, E. (2018). Cross-linguistic regularities and learner biases reflect "core" mechanics. *PLOS ONE*, *13*(1), e0184132. https://doi.org/10.1371/journal.pone.0184132

Strickland, B., Geraci, C., Chemla, E., Schlenker, P., Kelepir, M., & Pfau, R. (2015). Event representations constrain the structure of language: Sign language as a window into universally accessible linguistic biases. *Proceedings of the National Academy of Sciences*, *112*(19), 5968–5973. https://doi.org/10.1073/pnas.1423080112

Strickland, B., & Keil, F. (2011). Event completion: Event based inferences distort memory in a matter of seconds. *Cognition*, *121*(3), 409–415. https://doi.org/10.1016/j.cognition.2011.04.007

Swallow, K. M., Kemp, J. T., & Candan Simsek, A. (2018). The role of perspective in event segmentation. *Cognition*, *177*, 249–262. https://doi.org/10.1016/j.cognition.2018.04.019

Swedell, L., & Plummer, T. (2012). A papionin multilevel society as a model for hominin social evolution. *International Journal of Primatology*, *33*(5), 1165–1193. https://doi.org/10.1007/s10764-012-9600-9

Tafreshi, D., Thompson, J. J., & Racine, T. P. (2014). An analysis of the conceptual foundations of the infant preferential looking paradigm. *Human Development*, *57*(4), 222–240. https://doi.org/10.1159/000363487

Tardif, T., Fletcher, P., Liang, W., Zhang, Z., Kaciroti, N., & Marchman, V. A. (2008). Baby's first 10 words. *Developmental Psychology*, *44*(4), 929–938. https://doi.org/10.1037/0012-1649.44.4.929

Thompson, R. K. R., Oden, D. L., & Boysen, S. T. (1997). Language-naive chimpanzees (*Pan troglodytes*) judge relations between relations in a conceptual matching-to-sample task. *Journal of Experimental Psychology: Animal Behavior Processes*, *23*(1), 31–43. https://doi.org/10.1037/0097-7403.23.1.31

Tincoff, R., & Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science*, *10*(2), 172–175. https://doi.org/10.1111/1467-9280.00127

Tiran, T., Meewis, F., Fourtassi, A., & Dautriche, I. (2023). Typology of topological relations using machine translation. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *45*. https://doi.org/10.31234/osf.io/95nrd

Townsend, S. W., & Manser, M. B. (2013). Functionally referential communication in mammals: The past, present and the future. *Ethology*, *119*(1), 1–11. https://doi.org/10.1111/eth.12015

Uller, C. (2004). Disposition to recognize goals in infant chimpanzees. *Animal Cognition*, *7*(3). https://doi.org/10.1007/s10071-003-0204-9

Ünal, E., Ji, Y., & Papafragou, A. (2021). From event representation to linguistic meaning. *Topics in Cognitive Science*, *13*(1), 224–242. https://doi.org/10.1111/tops.12475

Ünal, E., & Papafragou, A. (2016). Interactions between language and mental representations. *Language Learning*, *66*(3), 554–580. https://doi.org/10.1111/lang.12188

Ünal, E., Richards, C., Trueswell, J. C., & Papafragou, A. (2021). Representing agents, patients, goals and instruments in causative events: A cross-linguistic investigation of early language and cognition. *Developmental Science*, *24*(6). https://doi.org/10.1111/desc.13116

Ünal, E., Wilson, F., Trueswell, J., & Papafragou, A. (2024). Asymmetries in encoding event roles: Evidence from language and cognition. *Cognition*, *250*, 105868. https://doi.org/10.1016/j.cognition.2024.105868

van Rooijen, R., Ploeger, A., & Kret, M. E. (2017). The dot-probe task to measure emotional attention: A suitable measure in comparative studies? *Psychonomic Bulletin & Review*, *24*(6), 1686–1717. https://doi.org/10.3758/s13423-016-1224-1

Vauclair, J., Fagot, J., & Hopkins, W. D. (1993). Rotation of mental images in baboons when the visual input is directed to the left cerebral hemisphere. *Psychological Science*, *4*(2), 99–103. https://doi.org/10.1111/j.1467-9280.1993.tb00468.x

Visch, V. T., & Tan, E. S. (2009). Categorizing moving objects into film genres: The effect of animacy attribution, emotional response, and the deviation from non-fiction. *Cognition*, *110*(2), 265–272. https://doi.org/10.1016/j.cognition.2008.10.018

Völter, C. J., & Call, J. (2017). Causal and inferential reasoning in animals. In *APA handbook of comparative psychology: Perception, learning, and cognition, Vol. 2* (pp. 643–671). American Psychological Association. https://doi.org/10.1037/0000012-029

Völter, C. J., & Huber, L. (2021). Dogs' looking times and pupil dilation response reveal expectations about contact causality. *Biology Letters*, *17*(12), 20210465. https://doi.org/10.1098/rsbl.2021.0465

Von Fersen, L., & Lea, S. E. G. (1990). Category discrimination by pigeons using five polymorphous features. *Journal of the Experimental Analysis of Behavior*, *54*(2), 69–84. https://doi.org/10.1901/jeab.1990.54-69

Wagner, L., & Lakusta, L. (2009). Using language to navigate the infant mind. *Perspectives on Psychological Science*, *4*(2), 177–184. https://doi.org/10.1111/j.1745-6924.2009.01117.x

Webb, A., Knott, A., & MacAskill, M. R. (2010). Eye movements during transitive action observation have sequential structure. *Acta Psychologica*, *133*(1), 51–56. https://doi.org/10.1016/j.actpsy.2009.09.001

Wegdell, F., Hammerschmidt, K., & Fischer, J. (2019). Conserved alarm calls but rapid auditory learning in monkey responses to novel flying objects. *Nature Ecology & Evolution*, *3*(7), 1039–1042. https://doi.org/10.1038/s41559-019-0903-5

Weiss, D. J., & Newport, E. L. (2006). Mechanisms underlying language acquisition: Benefits from a comparative approach. *Infancy*, *9*(2), 241–257. https://doi.org/10.1207/s15327078in0902_8

Wilson, F., Papafragou, A., Bunger, A., & Trueswell, J. (2011). Rapid extraction of event participants in caused motion events. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 7.

Wilson, V. A. D., Sauppe, S., Brocard, S., Ringen, E., Daum, M. M., Wermelinger, S., Gu, N., Andrews, C., Isasi-Isasmendi, A., Bickel, B., & Zuberbühler, K. (2023). *Primate origins of human event cognition* (p. 2023.11.23.568086). bioRxiv. https://doi.org/10.1101/2023.11.23.568086

Wilson, V. A. D., Zuberbühler, K., & Bickel, B. (2022). The evolutionary origins of syntax: Event cognition in nonhuman primates. *Science Advances*, *8*(25), eabn8464. https://doi.org/10.1126/sciadv.abn8464

Wood, J. N., Glynn, D. D., & Hauser, M. D. (2008). Rhesus monkeys' understanding of actions and goals. *Social Neuroscience*, *3*(1), 60–68. https://doi.org/10.1080/17470910701563442

Yates, T. S., Sherman, B. E., & Yousif, S. R. (2023). More than a moment: What does it mean to call something an 'event'? *Psychonomic Bulletin & Review*, *30*(6), 2067–2082. https://doi.org/10.3758/s13423-023-02311-4

Yates, T. S., Skalaban, L. J., Ellis, C. T., Bracher, A. J., Baldassano, C., & Turk-Browne, N. B. (2022). Neural event segmentation of continuous experience in human infants. *Proceedings of the National Academy of Sciences*, *119*(43), e2200257119. https://doi.org/10.1073/pnas.2200257119

Yin, J., & Csibra, G. (2015). Concept-based word learning in human infants. *Psychological Science*, *26*(8), 1316–1324. https://doi.org/10.1177/0956797615588753

Young, M. E., Beckmann, J. S., & Wasserman, E. A. (2006). Pigeons's discrimination of Michotte's launching effect. *Journal of the Experimental Analysis of Behavior*, *86*(2), 223–237. https://doi.org/10.1901/jeab.2006.60-05

Zacks, J. M. (2020). Event perception and memory. *Annual Review of Psychology*, *71*(1), 165–191. https://doi.org/10.1146/annurev-psych-010419-051101

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind/brain perspective. *Psychological Bulletin*, *133*(2), 273–293. https://doi.org/10.1037/0033-2909.133.2.273

Zacks, J. M., & Swallow, K. M. (2007). Event Segmentation. *Current Directions in Psychological Science*, *16*(2), 80–84. https://doi.org/10.1111/j.1467-8721.2007.00480.x

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, *127*(1), 3–21.

Zacks, J. M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General*, *130*(1), 29–58. https://doi.org/10.1037/0096-3445.130.1.29

Zinner, D., Groeneveld, L. F., Keller, C., & Roos, C. (2009). Mitochondrial phylogeography of baboons (*Papio spp.*) – Indication for introgressive hybridization? *BMC Evolutionary Biology*, *9*(1), 83. https://doi.org/10.1186/1471-2148-9-83

Zuberbühler, K. (2019). Evolutionary roads to syntax. *Animal Behaviour*, *151*, 259–265. https://doi.org/10.1016/j.anbehav.2019.03.006

Zuberbühler, K. (2020). Syntax and compositionality in animal communication. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1789), 20190062. https://doi.org/10.1098/rstb.2019.0062

Zuberbühler, K. (2022). Event parsing and the origins of grammar. *WIREs Cognitive Science*, *13*(3). https://doi.org/10.1002/wcs.1587

Zuberbühler, K., & Bickel, B. (2022). Transition to language: From agent perception to event representation. *WIREs Cognitive Science*, *13*(6). https://doi.org/10.1002/wcs.1594

Zwitserlood, P., Bölte, J., Hofmann, R., Meier, C. C., & Dobel, C. (2018). Seeing for speaking: Semantic and lexical information provided by briefly presented, naturalistic action scenes. *PLOS ONE*, *13*(4), e0194762. https://doi.org/10.1371/journal.pone.0194762

ANNEXES

A. Cogsci 2023 proceedings paper

Typology of topological relations using machine translation

Tanguy Tiran[1] (tanguy.tiran@etu.univ-amu.fr)
Floor Meewis[1] (floor.meewis@univ-amu.fr)
Abdellah Fourtassi[2] (abdellah.fourtassi@univ-amu.fr)
Isabelle Dautriche[1] (isabelle.dautriche@cnrs.fr)

[1]Aix Marseille Université, CNRS, LPC, Marseille, France
[2]Aix Marseille Université, Université de Toulon, CNRS, LIS, Marseille, France

Abstract

Languages describe spatial relations in different manners. It is however hypothesized that highly frequent ways of categorizing spatial relations across languages correspond to the natural ways humans conceptualize them. In this study, we explore the use of machine translation to gather data in semantic typology to address whether different languages show similarities in how they carve up space. We collected spatial descriptions in English, translated them using machine translation, and subsequently extracted spatial terms automatically. Our results suggest that most spatial descriptions are accurately translated. Despite limitations in our extraction of spatial terms, we obtain meaningful patterns of spatial relation categorization across languages. We discuss translation limits for semantic typology and possible future directions.

Keywords: Semantic typology; machine translation; spatial relations; semantic universals

Introduction

Languages vary in the way they encode thoughts, leading to differences in how – and even which – concepts are expressed. Such differences can be found in the domain of topological relationships, i.e., non-perspectival spatial relations which include relations such as support, containment, and proximity (Levinson & Wilkins, 2006). For instance, whereas English conflates horizontal and vertical support under a single term ("*on*"), Dutch employs a finer-grained distinction and uses different terms ("*op*" and "*aan*", respectively) for both of these contact relationships (Bowerman & Choi, 2001).

How do these semantic variations in spatial language relate to cognition? One possibility is that highly frequent ways of categorizing spatial relations across languages correspond to the natural ways humans conceptualize them; a proposition put forward as the typological prevalence hypothesis (Gentner & Bowerman, 2009). To test this hypothesis, Levinson and Meira (2003) elicited spatial terms from native speakers of 9 languages who were asked to describe 71 pictures covering a wide range of topological relationships (Topological Relations Picture Series or TRPS, Bowerman & Pederson, 1992). They found that, while the 9 languages under study did not share the same basic topological categories, they did tend to organize their semantics around conceptual attractors. Those attractors included notions of attachment, containment, superposition, as well as a category conflating "near" and "under" relations. These results, while incompatible with an absolute universal position, hint nevertheless at the existence of a distributional universal in the domain of topological relationships. Such a distributional universal could be linked to the existence of naturally existing conceptual categories, in line with the typological prevalence hypothesis.

Linguistic universals, or the absence of thereof, are often justified by cross-linguistic analyses. Levinson and Meira (2003) for instance, chose to study 9 languages that belong to distinct language families in order to control for any similarity that could have been accounted for by historical proximity. However, while it is desirable for typologists to investigate genetically diverse languages - as languages exert influence over each other, including as many languages as possible may also be important - as some differences may exist even between closely related languages (as exemplified earlier with the differences between English and Dutch). Yet, while several studies have investigated topological relations (e.g., Beekhuizen & Stevenson, 2015; Bowerman & Choi, 2001; Levinson & Meira, 2003), hardly any has done so with a large set of languages. An important reason behind this shortcoming is the substantial time and monetary expenditure that is required to conduct this type of research.

We aim at bypassing those limits and target a wide range of languages by exploring the use of a novel method in semantic typology: automatic translation. Despite improvements in recent years, machine translation has, to our knowledge, only been used once to study spatial terms before, and on single words only (Strickland & Chemla, 2018).

We aim here at using automatic translation to translate sentences containing spatial descriptions in many languages. Critically, neural-based models use context to translate an utterance and appear to be able to pick up the necessary information from the rest of the sentence to accurately translate a spatial description. For instance, consider the two English sentences "the picture is on the wall" and "the picture is on the table", which use the same term ("*on*") to express two different relations. A machine translation model such as Google Translate distinguishes these two relations when translating both sentences to a language that expresses them differently (producing in Dutch, for example, "De foto hangt *aan* de muur" and "De foto staat *op* tafel").

Our goals are (1) to assess whether automatic translation can be a valid tool for semantic typology and if so, (2) to ultimately provide a typology of topological relations for a large set of languages. Such typology could answer whether (distributional) universal conceptual biases govern the way we carve up the semantic space.

To achieve these goals, we plan to (1) elicit spatial descriptions in English with the help of the TRPS, (2) translate the elicited sentences in other languages and extract the spatial terms within those translations, and (3) evaluate in a sample of languages the quality of both translations and extractions. Lastly (4), we will use clustering analyses to see whether different TRPS pictures are described by similar terms across the different languages.

Acquirement of the translated data

Crowdsourcing elicitation data

We conducted an elicitation task in which English participants were asked to describe pictures using spatial terms. The reasons why we gathered spatial elicitations ourselves instead of relying on available data were: (1) that we needed sufficient elicitations for each picture to capture the wide range of ways of describing spatial relations, and (2) that we needed those elicitations to be in a standardized format to be able to translate them efficiently.

The experiment was hosted online. Beekhuizen and Stevenson (2015) have compared data obtained online with data obtained in-person and have shown that crowdsourcing is an appropriate method to obtain data in semantic typology. In total, 509 English speakers agreed to take part in the experiment, which was hosted on the Amazon Mechanical Turk crowdsourcing platform.

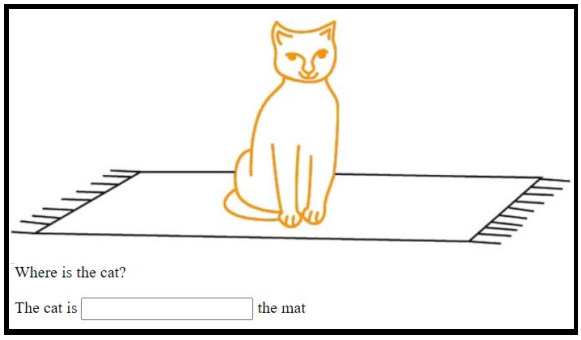


Figure 1: Example of a picture that participants had to describe in the elicitation task.

The original TRPS was used as stimuli for elicitation. The 71 pictures were divided into seven subsets, each containing 10 pictures (except for one subset which contained 11 pictures). Participants saw all pictures of a subset in random order, and were asked for each picture to answer the question "where is the [yellow object/figure]" using the first words that came to their mind (example provided in Fig. 1). Participants typed in their answers, filling in sentences of the kind "The *figure* is [BLANK] the *ground*". This standardized format allowed us to translate sentences efficiently.

Participants' data were rejected when participants wrote the same answer for every picture (N=17), when it was clear they did not do the task (e.g., some forms only contained numbers or random words) (N=15), when they tried participating more than once for the same subset of pictures (N=3) or when participants declared speaking more than one language at the end of the experiment (N = 155), as the experiment was intended for monolingual English speakers only. This left a total of 377 participants (66.5% of the initial pool of participants), yielding on average 55 elicitations for each picture (min=49, max=60).

Data was then cleaned by fixing clear typos (e.g., "hangnig on") and removing elicitations that did not fit grammatically in the sentences (e.g., "The cup is top the table"). Even though we constrained the answers heavily and asked participants to only complete sentences with spatial terms, many participants failed to do the task properly. This is comparable to previous studies which suggest that Mechanical Turk has experienced a decrease in participants responses' quality in recent years (e.g., Chmielewski & Kucker, 2020).

Finally, in order to cross-check participants' answers, two native English speakers assessed whether the elicited terms were semantically appropriate for each picture. The consultants judged at least one spatial expression as correct for each picture (min=1, max=8, mean=3.1). Note that the two consultants agreed on 82% of the proposed spatial terms (Cohen $\kappa = 0.64$) suggesting that there is some variability on how to describe the spatial relations depicted in the TRPS.

Translation of the elicited spatial descriptions

In order to understand how spatial relations are described across languages, the elicited sentences were translated in all languages available on Google Translate (GT) using the GT API in Python 3.9.5, in October 2022. GT is a state-of-the-art translation system which uses recent neural-based techniques (Wu et al., 2016). It was chosen for its accessibility and its extensiveness: the system offers support for more than a hundred languages belonging to 16 different language families.

Extraction of spatial terms within spatial descriptions

Because of the large number of languages in our sample, we sought to extract spatial terms from the translated sentences automatically.

We extracted spatial terms from the translated sentences using AWESoME, a recent alignment model which achieves state-of-the-art performance (Dou & Neubig, 2021). It uses multilingual BERT representations to align sentences of different languages and is available in more than 100 languages.

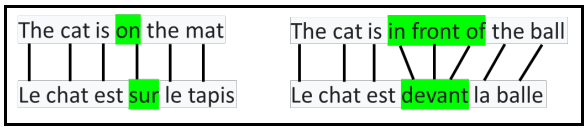


Figure 2: Illustration of alignment and extraction of spatial terms for two sentences.

AWESoME takes as input tokenized sentences (here, sentences split into words) in a source language (in this case, English) and a target language and matches their words (as in Fig. 2). Sentences were tokenized using mosestokenizer (Koehn et al., 2007) through the sacremoses library in python. The final list of investigated languages is the intersection between languages available on GT and languages available on AWESoME (minus Japanese, which was dropped because of tokenization issues). It includes 73 languages covering 14 different language families.

The spatial terms of the translated sentences were retrieved by extracting the words that aligned with the English spatial terms of the original sentences.

AWESoME's alignment performance ostensibly varied across languages. To infer whether it correctly aligned sentences, we used an alignment score for spatial terms for each language (henceforth referred to as the alignment score). We defined this score as the percentage of times AWESoME identified matching words for the English spatial terms in the translated sentences. Most languages revealed high alignment scores (mean≈86%) with alignments for spatial terms found in every sentence (most of these languages belong to the Indo-European family). However, alignment scores were considerably lower in some other languages (min≈57%). Most languages with low alignment performance turned out to be agglutinative, i.e., languages where spatial relations can be expressed through affixes appended to verbs or nouns. While the alignment scores help us hypothesize for which languages alignment could fail, it does not disclose whether the right spatial terms were extracted. Evaluating the extraction of spatial terms is consequently a necessity.

Evaluation of the translated data

We evaluated the sentence translations and the accuracy of the spatial term alignment in a sample of eight languages. Those languages were selected according to their genetic diversity and/or their morphological complexity, as well as native speaker availability.

Native speakers (one for each language) were shown the translated sentences and extracted spatial terms in their languages in a random order. They had to judge (1) if the (automatically translated) sentences were correct descriptions of their matching pictures (an evaluation of our translation tool, GT) and (2) if the spatial terms that were extracted using alignment were indeed correct spatial terms (an evaluation of our alignment tool, AWESoME).

Table 1 presents the evaluation results. In our sample, the evaluators judged most sentences as correct descriptions of their matching pictures in their native languages (except for Hindi) (mean percentage of sentences judged correct = 78%, sd = 14%). This indicates that GT properly translated most sentences in seven of our eight sampled languages.

We obtained more mitigated results on spatial terms extractions (mean of scores = 51%, sd = 23). Spatial terms were on average correctly retrieved in Dutch, Arabic, and to some extent, French. On the other hand, we obtained low scores in our sample's remaining languages.

We also found that patterns of errors varied across languages. In French and Hindi, most errors occurred when extracted terms did not contain all relevant words (e.g., "l'intérieur" instead of "à l'intérieur"/inside). Other languages exhibited the opposite pattern, with most errors corresponding to extractions containing more than only the relevant terms ("tableON" types of error for example). This was especially the case for agglutinative languages (e.g., Turkish or Hungarian). We consider other potential ways to extract spatial terms from translations in the discussion.

Table 1: Results of the translation and alignment evaluations in eight languages. Left: Percentage of translated sentences (by GT) judged correct by native speakers. Right: Percentage of extracted spatial terms (by AWESoME) identified as spatial terms by native speakers.

Language	Sentences (%)	Spatial terms (%)
Dutch	84	81
French	85	66
Hindi	48	28
Arabic	79	76
Hebrew	98	59
Turkish	78	42
Hungarian	78	42
Chinese	78	15

Clustering analysis

We investigated the extent to which different pictures are described by the same terms in each language. The method we follow here can be summarized in three steps.

Step 1: matrices representing the linguistic dissimilarity between pictures' description were computed for each language. Similarity between pictures *i* and *j* was computed as the number of spatial terms shared between them divided by the number of unique terms used for either picture.

$$D_{i,j} = \frac{|st_i \cap st_j|}{|st_i \cup st_j|}$$

Where st_k corresponds to the spatial terms used to describe picture *k*.

This measure resulted in 71 by 71 dissimilarity (or distance) matrices for each language where the value in a cell $D_{i,j}$ corresponded to the extent to which picture *i* and *j* differed in their linguistic treatment, with scores close to 1

meaning they were described using different spatial terms, and scores close to 0 meaning they were described using the same spatial terms.

Step 2: All matrices were weighted to mitigate the effect of genetic relatedness. The matrices of each language were multiplied by the inverse of the number of languages belonging to their family. They were subsequently added together and the final matrix was divided by the number of families in our set, so as to get values ranging between 0 and 1.

Step 3: We ultimately performed two statistical analyses on the final matrix. Hierarchical clustering (using the factoextra library (Kassambara & Mundt, 2017) in R) was applied to find clusters of pictures that were treated similarly across languages. Another statistical tool, Multidimensional Scaling (or MDS), was used to help visualize the linguistic treatment of pictures across languages. MDS takes as input the 71 by 71 matrices and reduces them to 71 two-dimensional points that can then be plotted in a cartesian plane. Because matrices represent distances between pictures, points that will be close on this plot will correspond to pictures that tend to be treated the same across languages, while points that are further apart will correspond to pictures seldomly described using the same terms.

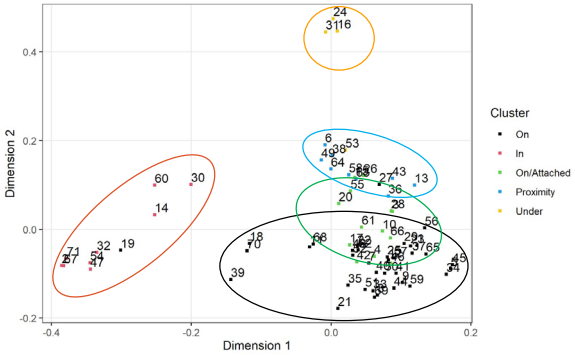


Figure 3: Pictures' similarity between how spatial relations are described across different languages plotted in a two-dimensional plane. Each point corresponds to a picture of the TRPS. Points that are close together indicate that they are described with the same spatial terms in different languages. Visual inspection revealed five clusters of pictures that can be described by similar spatial terms.

From Figure 3, we can see that pictures are not randomly distributed in the 2D space but tend to cluster together. We identified 5[1] clusters corresponding to containment ("In"), "Under", Proximity (e.g., "dog next to a doghouse", "lamp above a table"), Support, and a final cluster of relations – which we label "On/Attached" – where figures are attached to the ground and cannot be easily displaced (e.g., "face on a stamp", "strap attached to a bag"). Despite the poor quality of spatial terms extraction, it thus seems that our method is able to pick up cross-linguistic regularities in the way languages express topological relationships.

Discussion

In this project, we explored the use of machine translation to acquire data in semantic typology. We had the following two objectives: (1) assess whether automatic translation could be used to answer questions in semantic typology, and if so (2) take advantage of machine translation's assets to provide a typology of topological relations for a large set of languages.

Is automatic translation appropriate for semantic typology?

The results on translations are promising in all but one language (Hindi), with four out of five (79%) sentences being correctly translated on average, indicating they are valid descriptions of the TRPS pictures. Yet this leaves around 20% of translated sentences judged as incorrect. It should however be taken into account that only one speaker evaluated the translations for each language. It is thus possible that the consultants judged as incorrect some sentences that other speakers of those languages would have found correct. Recall that the two English consultants who helped construct the final English dataset only agreed on 82% of cases suggesting that inter-individual variability exists and is non-negligeable when expressing spatial relations. Overall, these results suggest that, while not being perfect, Google Translate, our translation tool, successfully managed to use context embedded in the source English sentences to pick appropriate spatial expressions in the target languages. Although we only evaluated translations in eight languages, these languages belonged to different typological families, thus we expect these conclusions to apply to the rest of languages available on GT.

Whilst the scores we obtain on translation are promising, results on the extraction of spatial terms are much more mixed: 51% of spatial terms are correctly extracted over the eight languages. The extraction performance varied greatly between languages and was poor in half of our evaluation sample's languages. Two possibilities may explain this result. First, mBert representations, which are used by AWESoME, are trained unevenly on each language (Wu & Dredze, 2020). AWESoME is thus maximally performant when used between pairs of high-resource languages. Secondly, one limit of our extraction technique is that we are unable to retrieve spatial affixes, which are especially common in agglutinative languages.

We have explored another way to extract spatial terms automatically and bypass the issues mentioned above. This other method took advantage of the constrained format of participants' responses in the crowdsourcing task. We created an additional dataset of English sentences by completing all sentences (e.g., "The cat is … the mat") with various spatial terms. We then translated those additional

[1] An objective method to find optimal number of clusters, the Silhouette method (Rousseeuw, 1987) suggested the data comprised six clusters. As one of the six clusters could not be interpretated, we display here results with five clusters.

sentences. Since for each picture, only the spatial terms varied across sentences, words that were less frequent in the translations were likely to be spatial terms. Importantly, this other extraction method (1) did not rely on a neural model and thus could perform similarly on low-resource and high-resource languages, and (2) allowed us to extract spatial affixes by looking at the frequencies of strings of letters instead of full words. Although this method did help retrieve spatial suffixes in agglutinative languages, it did not yield better results than alignment overall.

Despite its problem, it is noteworthy that our method yields interpretable clusters of topological relations across languages. This suggests that the analysis of categorization patterns could be robust to extraction errors when investigating a large number of languages. The errors may especially be randomly distributed across languages and pictures, and could be canceled out when investigating many languages. In this case, errors brought by our method could thus be compensated by the quantitative scale machine translation allows us to reach.

It is also worth mentioning that our method, while building on elicitations, has ties with another source of data that has been used in typology: parallel corpora, that is, texts which have been translated (usually not automatically) in different languages (e.g., Cysouw & Wälchli, 2007). Parallel corpora have been used to investigate semantic spaces across languages, especially in abstract domains (such as indefinite pronouns, as in Beekhuizen, Watson & Stevenson, 2017) where words (e.g., "someone" or "anything") are difficult to elicit.

In this work, by automatically translating English elicitation data, we have effectively created a parallel corpus of topological relationships. Our study thus relates to parallel corpora usage in semantic typology and bridges this strand of work to traditional elicitation studies. Additionally, machine translation presents potential benefits over these two methods. Firstly, it allows to collect data much faster than elicitation and across many languages. This data could also include abstract words that are difficult to study using elicitations only, and future research could explore in this direction. Secondly, this method is not restricted to scarcely available massively parallel texts (i.e., texts that are already translated in many languages). Moreover, automatically translated texts could be designed to be much more tailored to specific research topics, facilitating appropriate data collection.

However, a potential shortcoming of using automatic translation could be the presence of mistranslations (an already existing risk with parallel corpora studies). Ensuring the translated data is not akin to "*translationese*" and reflect how native speakers would have described the different pictures is thus a necessity, which we discuss in the following section.

Comparison with elicitation data

The translated data can only be used for semantic typology purposes if the translations adequately reflect how native speakers would have described the different spatial relations. Even though, on average, the translations correctly matched their accompanying picture, they might not necessarily correspond to how native speakers may spontaneously describe these pictures.

To test this, we asked native speaker consultants of six different languages (one for each language) to describe the TRPS pictures and compared these elicited descriptions to our translations. The consultants had to type a full sentence describing the picture they were shown and identify themselves the spatial words present in their descriptions. For each language, we computed three measures (recall, precision, and the F-score) comparing the elicited spatial terms with terms extracted manually (by the first author) from the translated sentences[1]. Recall is defined as the proportion of elicited spatial terms that were present in our translated data, while precision corresponds to the proportion of spatial terms in our translated data that were also elicited. F-score is the harmonic mean between these two metrics. An F-score of 1 would mean that the translated data perfectly matches the elicited spatial terms. The results of the comparison between those terms and the elicited terms are shown in Table 2.

Table 2: Comparison between the elicited spatial terms and manually extracted terms from the translations.

Language	Recall	Precision	F-score
Dutch	0.67	0.52	0.58
French	0.70	0.38	0.49
Russian	0.83	0.44	0.58
Turkish	0.57	0.36	0.44
Hungarian	0.71	0.41	0.52
Chinese	0.76	0.43	0.55
Mean	0.71	0.42	0.53

Our results in a sample of six languages suggest the two types of data moderately match (F-score = 0.53). The high recall rate indicates that the translations managed to effectively capture spatial terms that are used by native speakers. Notably, recall was much higher than precision across the six languages. This result can be explained by the fact that the translated data contained on average more sentences by language (M=219) than the elicited data (M=80). It is critical to note that there are some limitations to

[1] We also compared the elicited spatial terms with the automatically extracted spatial terms. Yet, as we described previously, the extraction method worked poorly on some languages, and this unsurprisingly leads to a low average F-score (M=0.30). While this strengthen our conclusions that spatial term extraction is challenging, this comparison cannot be used to evaluate whether our translated sentences match native speakers' spontaneous descriptions of the pictures as the spatial term, while not being extracted correctly, may still be present in the sentence.

the interpretation of these scores. Importantly, the elicitation data was gathered from only one speaker for each language and given the inter-individual variability for describing spatial relations, this might not reflect the full range of topological descriptions used for that given language.

In sum, this suggests that translation data, albeit not perfect, does reflect how native speakers use spatial terms.

Conclusions

In our study, we explored whether machine translation could be a valuable tool to derive a semantic typology of spatial relations.

Our findings indicate that current translation tools can fairly accurately translate elicitations of spatial relations from English to other languages. However, extracting spatial terms automatically from sentences remains a challenge. We found that alignment varies considerably across languages and performs especially badly on less-documented languages (a recurrent challenge for NLP tools) and languages that use spatial affixes. Further research could look into methods of improving extraction of words of interest especially for these kinds of languages.

Regardless of these concerns, our clustering revealed non-random patterns. While these non-random patterns are encouraging for the use of machine translation in semantic typology, they might however appear despite:

- A bias towards the source language (here, English). Translation could indeed be less accurate for languages which categorize spatial relations differently from English. Therefore, it remains unclear if translating from a single source language favors categorization patterns that are similar to that of the source language.
- Occasional translation errors. If randomly distributed across languages and pictures, these errors could nevertheless be canceled out when investigating many languages all together.

A future step we should take before interpretating clusters of pictures in depth would be to investigate these two possibilities.

Based on our study, we conclude that there are still some technological hurdles to take before machine translation can be effectively implemented to create a semantic typology of spatial relations. However, because our exploratory analyses showed some meaningful similarities across languages, we are positive that an improved method could prove a fruitful tool to answer whether universal biases exist in how we conceptually represent space.

Acknowledgments

This work was funded by the French National Agency for Research (ANR-20-CE28-0005).

We would like to thank everyone involved in making this study possible. We especially thank the consultants for their help in evaluating the sentences in English and in our sampled languages, as well as reviewers for their insightful comments.

References

Beekhuizen, B., & Stevenson, S. (2015). Crowdsourcing elicitation data for semantic typologies. In *CogSci*.

Beekhuizen, B., Watson, J., & Stevenson, S. (2017). Semantic Typology and Parallel Corpora: Something about Indefinite Pronouns. In *CogSci*.

Bowerman, M., & Choi, S. (2001). Shaping meanings for language: universal and language-specific in the acquisition of semantic categories. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development*. Cambridge: CUP.

Bowerman, M., & Pederson, E. (1992). Topological relations picture series. In *Space stimuli kit 1.2* (p. 51). Max Planck Institute for Psycholinguistics.

Chmielewski, M., & Kucker, S. C. (2020). An MTurk crisis? Shifts in data quality and the impact on study results. *Social Psychological and Personality Science*, 11(4), 464-473.

Cysouw, M., & Wälchli, B. (2007). Parallel texts: Using translational equivalents in linguistic typology. *Language typology and universals*, 60(2), 95-99.

Dou, Z. Y., & Neubig, G. (2021). Word alignment by fine-tuning embeddings on parallel corpora. *arXiv preprint arXiv:2101.08231*.

Gentner, D., & Bowerman, M. (2009). Why some spatial semantic categories are harder to learn than others. The Typological Prevalence Hypothesis. In J. Guo et al. (Ed.), *Crosslinguistic approaches to the psychology of language. Research in the tradition of Dan Isaac Slobin*. New York: Psychology Press.

Kassambara, A., & Mundt, F. (2017). Package 'factoextra'. Extract and visualize the results of multivariate data analyses, 76.

Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., ... & Herbst, E. (2007, June). Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th annual meeting of the association for computational linguistics companion volume proceedings of the demo and poster sessions* (pp. 177-180).

Levinson, S. C., Meira, S., & The Language and Cognition Group. (2003). 'Natural concepts' in the spatial topological domain – Adpositional meanings in crosslinguistic perspective: An exercise in semantic typology. *Language*, 79(3), 485–516.

Levinson, S. C., & Wilkins, D. P. (2006). The background to the study of the language of space. In *Grammars of space: Explorations in cognitive diversity*. Cambridge University Press.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20, 53-65.

Strickland, B., & Chemla, E. (2018). Cross-linguistic regularities and learner biases reflect "core" mechanics. *Plos one*, 13(1), e0184132.

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., ... & Dean, J. (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*.