

PSYCHOLOGY

The evolutionary origins of syntax: Event cognition in nonhuman primates

Vanessa A. D. Wilson^{1,2,3*}, Klaus Zuberbühler^{1,3,4}, Balthasar Bickel^{2,3}

Languages tend to encode events from the perspective of agents, placing them first and in simpler forms than patients. This agent bias is mirrored by cognition: Agents are more quickly recognized than patients and generally attract more attention. This leads to the hypothesis that key aspects of language structure are fundamentally rooted in a cognition that decomposes events into agents, actions, and patients, privileging agents. Although this type of event representation is almost certainly universal across languages, it remains unclear whether the underlying cognition is uniquely human or more widespread in animals. Here, we review a range of evidence from primates and other animals, which suggests that agent-based event decomposition is phylogenetically older than humans. We propose a research program to test this hypothesis in great apes and human infants, with the goal to resolve one of the major questions in the evolution of language, the origins of syntax.

INTRODUCTION

Language is largely structured around the description of events (1) and allows speakers of any language to communicate not only “who did what” but also “to whom,” “where,” “when,” or “how.” “Tom chases the dog,” “the monkey stole the sandwich,” and “the ball hit the car” all describe how agents, animate or inanimate (Tom, the monkey, and the ball), interact (chase, steal, and hit) with a patient-like entity (dog, sandwich, and car) in goal-directed ways. One of the challenges in language comprehension is to reconstruct a compositionally organized event encoded by the speaker with a linear speech stream. This task is facilitated by the fact that both speaker and listener share the same basic “who-did-what-to-whom” structure, conveyed in the sentences of their language. Languages differ in how they express this basic structure, but once the rules are known (e.g., word order in English and case markers in Ukrainian), listeners can rapidly and efficiently identify the agent and patient regardless of lexical content. Intriguingly, languages tend to privilege agents in their grammar rules: In most languages, agents are by default placed first, rank highest in syntactic hierarchy, and receive the simplest marking (2–6).

Notably, the ability and propensity to differentiate agents from patients is evident not only in processing linguistic structure but also in how people perceive events nonlinguistically (7–9). For instance, when viewing events, people tend to exhibit a visual bias toward the agent (7–9), a propensity only moderately influenced by the characteristics of specific languages or grammatical choices (10, 11). This bias is evident even in infants as young as 5 months, when viewing interactions of inanimate shapes (12, 13).

Given the curious parallels between how humans perceive events as agent-driven and how languages prioritize agency through grammar, the hypothesis emerges that the evolution of event syntax (i.e., the compositional expression of events in language) has built on this cognitive bias of processing events as component parts (14–16). Here, we refer to this as the agent detection hypothesis. The idea is

that syntax represents the generalization of a compositional schema derived from event cognition. Concretely, hierarchical structure in syntax would mirror a basic schema [a “frame” or “script”; (17)] that links agent-only events (“A lashes out”) to patient-only events (“P hurts”) via a notion of “cause” in a conceptually hierarchical manner: [agent acts] causes [patient changes] (“A strikes P”) (14, 18). This schema, with two subunits embedded inside a larger expression, would then be generalized to other categories (beyond agents and patients) and to richer levels of complexity [perhaps via the emergence of recursion in a faculty of language in the narrow sense (19) or via gradual evolution (20, 21)].

A competing hypothesis reverses the arrow of causation and proposes that the evolution of language’s syntactic structure (via any of the currently proposed scenarios) fundamentally reshaped event cognition in humans so that agent-based event decomposition coevolved with language. Concretely, the core principles of hierarchical composition would have evolved independently of event cognition, but once established, they would have allowed integration of agents and patients into single composite expressions, thereby coevolving new, potentially variable categories (“agent-of,” “patient-of,” “causative,” etc.) in languages. This would, in turn, have fundamentally restructured nonlinguistic event cognition, leading to the agent bias that we find in humans. We refer to this as the event description hypothesis.

The two hypotheses are highly general and allow many modeling choices (e.g., how semantic roles are formally represented, how they relate to conceptual space and morphosyntactic form, how hierarchy is declared or generated, how it relates to processing and learning mechanisms, etc.). These choices are orthogonal, however, to the key question that the hypotheses raise and that we focus on: Does the syntactically composite expression of agents and patients build on prelinguistic modes of cognition, or did this expression establish a novel kind of event cognition in humans?

To address these hypotheses, we review research on event processing in humans and nonhuman species, focusing particularly on primates. We examine the evidence for how animals process events and whether human event cognition, specifically, the way that people apprehend agent-patient interactions in events, is truly unique, thereby allowing us to address the theoretical origins of syntax. We limit our purview to the core syntax of event expression, particularly

Copyright © 2022
The Authors, some
rights reserved;
exclusive licensee
American Association
for the Advancement
of Science. No claim to
original U.S. Government
Works. Distributed
under a Creative
Commons Attribution
NonCommercial
License 4.0 (CC BY-NC).

¹Department of Comparative Cognition, Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland. ²Department of Comparative Language Science, University of Zurich, Zurich, Switzerland. ³Center for the Interdisciplinary Study of Language Evolution (ISLE), University of Zurich, Zurich, Switzerland. ⁴School of Psychology and Neuroscience, University of St Andrews, St. Andrews, Scotland.

*Corresponding author. Email: vanessa.wilson@unine.ch

dyadic interactions that depict agents and patients involved in goal-directed actions (not, for example, instances where two people act simultaneously). This corresponds to what is technically known as “transitive” syntax in linguistics, a type of structure that provides a blueprint for other aspects of event syntax.

We start by outlining findings in human event cognition and the proposed links to syntax. We then examine the evolutionary origins of event cognition through the representation of physical causality, social dynamics, and psychological states in nonhuman primates. After this, we address how agents are identified in different contexts in both humans and other primates. Last, we build on this evidence to propose that certain aspects of event processing, specifically, event decomposition and agency detection, are evolutionarily ancient and predate modern human evolution (see Fig. 1). In doing so, we also highlight what is missing from our understanding and propose the next steps required to fully understand how other primate species perceive the agent/patient dynamic in events. Ultimately, we address theories and debates surrounding the evolution of syntax and provide future directions for this field.

EVENT COGNITION IN HUMANS

“Event cognition” is a broad term that generally encompasses how people apprehend events. One problem is how individuals manage

to structure the continuous flow of information into meaningful chunks, i.e., how they recognize the boundaries of events (1). Another problem is how individuals identify and mentally represent the key features within events, e.g., the fact that an event took place at night, that it was raining, or that a person was handling an umbrella, but humans go beyond perceiving events as assemblies of categorical information. Since Michotte’s (22) pioneering work, we know that humans have something akin to a “causality detector,” as they perceive some interactions between objects as causal and others not (23–27). A person handling an umbrella exerts not only physical causation (changes geometry of umbrella) but also psychological causation (intends to avoid rain). The causal relation is furthermore decomposed into an agent (the person), an action (the manipulation), and a patient (the umbrella). Human event cognition, in other words, encompasses a conglomerate of mental skills, ranging from temporal chunking to categorical assessments to causal judgments. We focus here on the last feature, i.e., the human propensity to decompose natural events into agents, actions, and patients, and how this may provide the grounding for compositionality in event expression.

When people view events, there is a strong tendency to fixate earlier on the agent than the patient (7, 8, 28). This agent bias also manifests itself in higher accuracy for agents when scanning stimuli

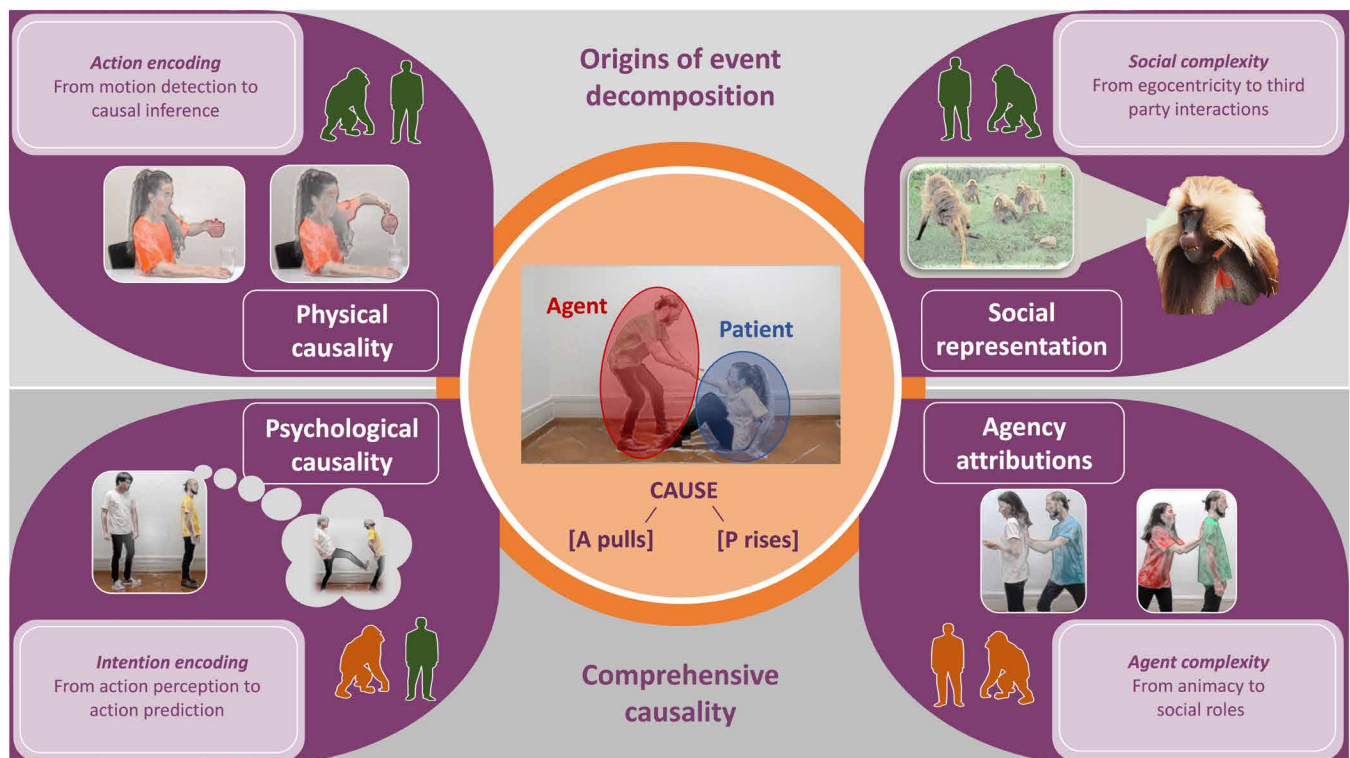


Fig. 1. Cognitive components of event decomposition. The central portion depicts how event decomposition accounts for causal components, something that is reflected in linguistic event representation and its syntax. The top half of the figure represents the hypothetical origins of event cognition, through motion encoding and social representation. The bottom half of the figure represents fine-grained event attributions such as understanding intentionality and which features determine asymmetry in attention between agents and patients. Each of the four components indicates a major cognitive capacity required for attributing semantic roles. For each section, the colored silhouettes represent strength of evidence across nonhuman primates compared with humans. Green indicates that most of the evidence supports the ability to reason about causal relationships, and orange indicates that evidence is mixed or lacking. Within agency attributions, we also coded humans as orange because it remains unclear what features allow people to rapidly extract agency from events. Photo credit: gelada monkeys, Vanessa Wilson, University of Neuchatel; silhouettes, <https://publicdomainvectors.org>; humans, https://osf.io/c5ubv/?view_only=51c5eea60722462e9ec490d94d89bb36.

very rapidly [<100 ms; (29–31)], when remembering events (32), or when detecting small changes between events (33). Indeed, the bias might be part of core cognition (34) in humans, attested early on in infants (13), and inscribed into fundamental mechanisms of event processing (8, 28).

This parallels a preference for agents in language structure, i.e., in grammatical rule systems. For example, in most languages, basic rules of grammar place agents first, so sentences such as “he likes beans” are the rule, while sentences such as “beans he likes” require special contexts, not only in English but also in most languages (5, 35). In addition, agents tend to be treated as the default, normal case, so that most languages use special and complex marking only when a noun is not the agent (such as the passive voice in English, which signals a patient, e.g., “John was hit,” or prepositions, which signal a recipient, e.g., “give it to John”); special marking for agents (known as “ergative case”) occurs, but it tends to be avoided when languages evolve over time (36). Furthermore, in most languages, an expression such as “the chickens should eat” can only be understood as an event where the chickens are the agents and they eat; in only very few languages, the chicken will be understood as the patients, i.e., that they are eaten. Patient-favoring rules of this kind exist, but they are exceedingly rare (4). Lastly, languages recurrently code agents as the starting point of an event, aligned by default with the point of view of the speaker (or someone that the speaker socializes and empathizes with), making, for example, sentences such as “he was told off by me” much less natural than “I told him off” (37, 38).

Social contexts and beliefs about the agent also play a critical role in the use of case marking. For example, in Western Samoan speakers, the use of ergative case depends on socially regulated assumptions about agency attribution (39). In one discussion, the high chief, who takes a mediatory role in political discussion, used agent nouns with a special agency marker (“ergative”) nine times, but in all of these instances, it was to refer to God. In comparison, two senior orators who were involved in the same debate used the ergative case with higher frequency and in more variable contexts, often using it to emphasize accusations against third parties. This example demonstrates that defining agency in language may depend on social variables, such as status within the community. In an analysis of German and Polish corpora, certain groups of target agents showed an increased likelihood to be followed by verbs, with men and younger people having a higher likelihood of being placed before a verb than women and older people, respectively (40), suggesting a social bias in how people attribute agency to these groups. These studies indicate that agency attribution may be inextricably entwined with language, culture, and cognition. Indeed, Chinese and American newspaper reports of murder cases have been shown to differ in their attributions for the murders, with Chinese reports focusing on situational factors (such as societal and relationship problems), while American reports focused on personal problems (personality and attitudes of the murderer) (41).

The parallelism between the cognitive and the linguistic agent bias, with all its social ramifications, suggests that they coevolved, perhaps analogously to the evolution of a precise number system in thought and language (42). Indeed, one could argue that the cognitive and the linguistic system are the same, barring evidence of differences in terms of specificity (concrete cognition versus abstract language) and related effects (43). However, there is mounting evidence that cognitive event decomposition and a preference for agents

in that decomposition might both be primary and precede language, as we argue in the following paragraphs.

First, evidence from deaf children suggests that language is not necessary to discriminate agents from patients: Deaf children who were never exposed to a conventional sign system produced spontaneous gestures for semantic roles, separate from event gestures (44). Consistent with this, language loss in aphasia does not compromise the ability to detect agents and patients in picture plausibility tasks although it weakens performance in sentence-picture matching tasks (45).

Second, while nonlinguistic event apprehension can be influenced by specific properties of one’s language (46, 47), the bias to apprehend agents faster and more accurately seems fairly robust against these differences. Crucially, the bias extends to languages such as Hindi or Basque that do not take agents to be the default and instead use ergative markers to signal them. This has been found in eye tracking when subjects are asked to look at the picture of an event in order to describe it later (48). The specific affordances of Hindi and Basque can modulate the bias, but they cannot overturn it.

Third, during language comprehension, the brain transiently and preferably assigns an agent role to the first unmarked noun in a sentence regardless of its actual role, e.g., a noun such as “the cow” or “the car” is assigned an agent role even when the sentence ends in a passive like “the cow/car was sold.” This preference is evidenced by error biases in understanding (49) and by electrophysiological deflections (event-related potentials) when the role turns out to be a patient at the end (50–52). Intriguingly, the preference and its electrophysiological effect persist for unmarked nouns in Hindi, which tend to be patients because agents receive ergative marking (36). In other words, even when nouns denote patients by default, the processing system transiently assumes that they are agents, triggering an electrophysiological signal when the overall sentence confirms that they are patients after all.

Fourth, neurobiological evidence indicates that there are neural overlaps in the processing of sentences and images that depict events (53–55). This would suggest that a language network emerged out of a neural basis dedicated to event apprehension, perhaps explaining why agents and patients are processed by distinct neural populations during sentence comprehension (56). However, conflicting results present an as yet unresolved picture as to the extent of this overlap (45, 57) and as to how and when representations are processed, beyond “where.”

These findings are particularly remarkable because, during language production, speakers do not always obey an agent-first principle: We simply say first what matters most, using noncanonical word orders, passives, and all sorts of other strategies that our grammar systems offer (58–60). Despite these habits in our productions and the freedom we have in it, the agent bias persists in event perception and sentence comprehension, as well as in the evolution and distribution of grammars worldwide. In short, the evidence reviewed here suggests that the bias has evolutionary roots that precede language.

EVENT COGNITION IN NONHUMAN PRIMATES Evolutionary origins of event cognition: The role of physical causality

To understand the evolutionary origins of human event cognition, the foremost question to address is whether nonhuman animals

process events in the same way as humans, i.e., whether they compositionally distinguish agents and patients and whether they grant agents privileged attention. Alternatively, animals including great apes may perceive events very differently, either as unstructured holistic occurrences or structured along other features, such as social dynamics (dominance, friendship, and kinship) or mere perceptual salience.

To consider this question in detail, the best place to start is perhaps with how animals process causality, a key aspect of event cognition. As noted earlier, humans have a strong predisposition to perceive contiguous events as causal. However, the processing of physical relationships in event scenarios is unlikely to be unique to humans. As we review below, a number of studies indicate the ability of primates (and, to some extent, other species) to reason about physical causality. While causal cognition may itself be a relatively derived trait, processing of low-level features of causal mechanisms likely emerged much earlier in evolution. Specifically, agency cues are often automatically encoded in biological motion dynamics (61). Distinguishing between different types of motion, particularly between predators, prey, and conspecifics, is most likely highly adaptive for survival (62), probably emerging over 500 million years ago among early bilaterals—cephalized species with unidirectional motion, active predation, and a necessity for mate seeking. These early neural systems would provide a basic framework that, through gradual refinement (63), would build causal detection mechanisms and the ability to attribute goal-directed behavior among third parties.

We can already make several distinctions along this spectrum of causal detection. One distinction is causal orientation. Newly hatched domestic chicks (*Gallus gallus*) exposed to computer-generated animated sequences depicting causal and noncausal interactions (64) exhibited imprinting preference (measured through approach behavior) for a self-propelled object playing a causative-agent role; they exhibited no preference when there was no causal interaction or when the causality was disrupted by a delay. These findings suggest an innate orientation to a self-propelled, causal agent in chickens, proposing that causal orientation may be an intermediary step in the emergence of causal reasoning. Similarly, both domestic dogs and humans differentiate between causal and independent animations, with a longer looking time to the noncausal over causal animation (65). The authors propose that the causal pattern is recognizable and thereby produced a more rapid habituation than the noncausal movement. However, the limitation of looking time paradigms is that they reveal little about the underlying motivation for visual bias (66); thus, interpretations can only be made with caution.

A second distinction is the ability to detect differences in an action, in terms of not only speed or direction but also self-control and efficiency. Hauser (67) used a looking time paradigm to examine cotton-top tamarins' (*Saguinus oedipus*) expectations about object location when the objects varied in self-propelled motion and animacy. He reports that, when viewed in a novel location, monkeys exhibited longer looking time to objects that were inanimate or lacked self-propelled motion compared to animate or self-propelled objects. This suggests that the monkeys formed expectations about the objects' abilities to move unaided, and the longer looking time in the "unexpected" conditions reflects this violation of expectation. Using a similar paradigm, long-tailed and southern pig-tailed macaques (*Macaca fascicularis* and *Macaca nemestrina*) were familiarized to a reaching action over a barrier: Removing the barrier resulted in looking longer to the inefficient action (as if the barrier was still present) than the efficient action (a linear reach to an object) (68),

suggesting that monkeys interpret actions in functional, goal-directed ways. Notably, this effect vanished when subjects (this time, southern pig-tailed and rhesus macaques, *Macaca mulatta*) observed unfamiliar actions, suggesting that expectations about goal-directed actions are limited to the monkeys' own motor repertoires or to previously observed actions. Two studies in free-ranging rhesus macaques present conflicting support for these interpretations. Monkeys exhibit avoidance behavior to a throwing action despite being unable to throw themselves (69). Moreover, the monkeys were attentive to specific variables, such as the speed and direction of the throw, as well as whether an object was thrown or not. This suggests that they can form expectations about the potential goal based on the specific action, an interpretation that goes beyond simple associative learning. In contrast, the same authors, in the same study population, found that in a feeding context, the monkeys were only sensitive to actions within their own behavioral repertoire (70). When presented with a choice of two coconuts, they were more likely to approach the one that was contacted by the experimenter with an action within their own repertoire but approached at chance when the action was not within their repertoire. It is notable that the monkeys did not differentiate between familiar actions within these conditions. While these findings remain inconclusive about whether macaques are able to process familiar actions outside their own behavioral repertoire, one interpretation is that their responses to these actions may be context specific.

The ability to track subjects (observers) or objects (things) is also a necessary step in following an event sequence. A relevant consideration then is how primates track objects through cohesion sequences. In a forced choice task, Western gorillas (*Gorilla gorilla*) and orangutans (*Pongo pygmaeus*) were consistently able to choose a larger over a smaller cracker, even when the larger cracker had been broken into smaller pieces (71). With the exception of the larger cracker being ground into crumbs, the apes were successfully able to track the object through the cohesion sequence. In rhesus macaques, the results for object tracking are not so clear. Subjects visually differentiated between an event where two food items were lifted together and an event where one food item was lifted away from another (72). They did not, however, differentiate between whole and halved food items. This latter result reflects findings in 11-month-old infants, who could differentiate cracker quantity but not once a cracker had been split into two halves (73). That is, cohesion violation seemed to disrupt their ability to represent the objects.

What about other physical properties? Chimpanzees' responses to physical support violations were assessed using videos of physically possible and impossible events (74). In the possible event, chimpanzees watched a banana being pushed across a platform, whereas in the impossible event, the banana was slid onto a Perspex base and lit in a way as to be invisible, thereby creating the pretense that the banana was suspended in the air. Subjects looked longer at the physically impossible event, suggesting that the banana's stability with an apparent lack of support violated their expectation about the causal relationship between objects and surfaces. This "surprise" response was maintained when only 15% of the banana was supported by a surface, in comparison to when 70% of the banana was supported. This response however only held true for vertical supports and not for horizontal ones, i.e. when an object received support from the side rather than from below. Similarly, both chimpanzees and Japanese macaques (*Macaca fuscata*) looked longer to the impossible over the possible support event but again did not differentiate

between possible and impossible supports when the surface was vertical (75). These studies reflect findings for 4-month-old human infants, who looked longer at an impossible event (a box placed on a transparent platform, apparently suspended in midair) over a possible event (box placed on a visible platform) (76). Further evidence from chimpanzees suggests that they can distinguish probable from improbable events (e.g., a banana being moved by a hand versus moving on its own). Using a habituation/dishabituation paradigm, O'Connell and Dunbar (77) found that subjects looked longer when an improbable clip was played following a habituated probable clip but not when the probable clip was played following the improbable clip. The authors propose that this indicates that the chimpanzees found the improbable scene causally more incongruous when viewed second.

An additional consideration for assessing causal knowledge is that of temporal contiguity (78). Forming causal connections between separate time points requires one to keep track of not only what happened but also when. Connecting events that are both spatially and temporally separated is argued to be necessary for "strong" causal knowledge (79). Examining memory may provide a window into understanding how primates perceive events. Rhesus macaques, for example, are able to differentiate time points within short video sequences that are then represented as still frames (80). Videos depicted various animal species undertaking unidirectional motion. When presented with still frames of the video on a touchscreen, monkeys had to choose the frame from the earlier time point, which they did above chance, indicating that they can form temporal order judgements about event details. In a real-world context, when presented with a choice of two food items after either a short (5 min) or long (1 hour) delay, apes (bonobos, *Pan paniscus*, chimpanzees, and orangutans) were able to distinguish between different food types depending on the time that had elapsed since they were placed in those positions (81). After a 5-min interval, they preferred to choose frozen juice over a grape, but after an hour, they were more likely to choose the grape, indicating that to make a choice, they integrated knowledge about the temporal state of each food item. However, the results were driven by only 4 of the 12 individuals tested, suggesting that this response does not necessarily generalize among the species tested. Alongside evidence that primates (and a few other species) attend to physical properties of causal relations, these studies suggest that they also exhibit event recall and form expectations about future event outcomes based on prior experience.

From basic motion detection systems, causal detection mechanisms and the ability to reason about causal relationships have emerged, incorporating components of causal orientation, sensitivity to action control and efficiency, object tracking, and reasoning about physical probability and temporal contiguity. While it is not yet clear when causal inference began to appear in the evolutionary timeline, the evidence presented here indicates that this goes at least as far back as nonhuman primates (including apes, Old World and New World monkeys, and with limited evidence beyond primates) and probably further (Fig. 1, top left). While these findings suggest that there may be differences in causal physical reasoning between humans and other species, these interpretations should be taken with caution without direct interspecific comparisons or a broader taxa assessment.

Social origins of event cognition: From recognition to representation

A second consideration in understanding the evolution of semantic role attribution is the origin of agency detection in the social

domain, i.e., situations where the observer is not an active participant but a passive observer of interactions between agents and patients. In group-living species, keeping track of who did what to whom is important for future decision-making about how to interact with other group members, such as who to groom, share food with, or support in a fight (82). Group-living species almost continuously interact with conspecifics, who are simultaneously both taxing competitors for shared resources and valuable cooperators toward shared goals. A group-living individual's fitness will therefore largely depend on how it navigates between these two opposing forces by keeping records of how other group members perform during social interactions. While in the above section, it is clear that primates reason to some extent about physical causal relationships, we propose that this causal representation was driven in large part by social forces; that is, the ability not only to distinguish but also to keep track of agent-patient events evolved with complex social living (Fig. 1, top right).

Concordantly, research in nonhuman primates sheds light on the role of social dynamics in keeping track of agent-patient roles. Playback studies with baboons (*Papio cynocephalus ursinus*) found that simulated vocal interactions between group members that violated the dominance hierarchy (i.e., a subordinate threatening a dominant) elicited longer looking times than interactions that were consistent with the hierarchy (83–85). Similar results were found for zoo-housed chimpanzees, where call sequences portrayed agonistic interactions between differently ranking group members (86). When viewing visual scenarios of conspecifics depicting congruent social information (dominant group mate threatening a subordinate) or incongruent information (subordinate group mate threatening a dominant), male long-tailed macaques (but not females) looked longer toward the incongruent than the congruent scenes (87). Conversely, monkeys did not differentiate between congruent and incongruent scenarios depicting bared teeth rather than open-mouth expressions. The authors suggest that this difference is due to different expressions holding different values for third parties. Together, these studies imply that species that form well-defined dominance hierarchies not only have knowledge about the dominance relationships but also form specific expectations about the behavior of other group members toward each other. These and similar findings have been put forward to make arguments about the evolution of communicative complexity (84). A need to keep track of social interactions should therefore favor a cognitive apparatus capable of distinguishing agency and patience.

Expectations about conspecifics' social roles appear to extend beyond relatively fixed hierarchies to other social contexts. In sanctuary-housed bonobos, vocal responses of victims to unexpected, unprovoked conflicts were generally more severe than responses to expected conflicts (such as after responses to food theft), suggesting that unprovoked, unjustified aggression violated subjects' expectations about social behavior (88). Keeping track of these interactions, that is, who violates social norms, could be beneficial to making future decisions about who to interact with. In these contexts, it would be advantageous to monitor the perpetrators (agents) of unjustified attacks. Similarly, captive chimpanzees (*Pan troglodytes*) looked longer at video scenes depicting infanticide than other forms of high aggression or hunting behavior, regardless of novelty and other potentially intervening variables, suggesting that great apes categorize and have expectations about different agent-patient dynamics (89). Indeed, when presented with images of conspecifics that clearly depict

agent-patient interactions (grooming and aggression), rhesus macaques attributed more attention to patients than agents, indicating that they can visually distinguish the two (90). Since the same pattern was not found for nonconspicuous images and was limited in social context, it is unclear to what extent attention patterns were driven by the social variables depicted, but this study certainly lends support to the notion of prelinguistic semantic role attribution.

Agency may also be important for predicting social outcomes. A crested macaque (*Macaca nigra*) was trained to match videos of social scenarios with images of different social outcomes, and results suggested that she selected outcomes that were best predicted by the facial expressions of the scenarios (91). It remains unclear however whether the subject truly understood the temporal association implied between the scenario and the social outcomes.

Social contexts also seem to influence memory for events. In a study by Howard *et al.* (92), subjects (chimpanzees and gorillas) watched a video of a tower being built by either a hand (social condition) or a claw (nonsocial condition). They were then shown images of two towers: the one they saw being built and a novel one. Subjects looked longer, as predicted, toward the new versus the old tower in the social, but not the nonsocial, condition. These findings are reflected in both prelingual infants (93) and brown capuchin monkeys (94) and fit with findings in adult humans for a natural bias toward social agency (95). The role of social stimuli in memory also extends, unsurprisingly, beyond primates; for example, male meadow voles (*Microtus pennsylvanicus*) alter their search behavior of an apparatus depending on the previous presence of a sexually receptive female 24 hours earlier (96). While the findings presented here emphasize the importance of social stimuli in event processing, in actuality, things may be more complex than that. Howard (92) proposes that action encoding benefits from a relatable (i.e. social) agent and a goal-directed action, which fits with previously discussed findings (64, 97, 98). Consistent with this interpretation, when infants were first familiarized with the goal-directed action of the claw in a real-life setting, they exhibited better memory for the tower in the claw condition compared with when the claw was unfamiliar (93).

In sum, these studies provide preliminary evidence for causally organized event representation in nonhuman primates, which goes beyond simple agency detection to social roles and expectations of how conspecifics should behave. Building on this, in the following section, we explore how nonhuman primates respond to internal states, a crucial consideration in the evolution of social cognition.

Psychological causality

The problem with studying perception of goal-directed actions is that goals are not necessarily visible and can represent internal psychological states that can be difficult to discern (99), such as intentionality. Intentionality requires careful definition since it must be inferred from behavior: In a social context, it may incorporate a combination of means-end dissociation (i.e., the same signals in different contexts and different signals in the same contexts), the use of signal repetition and flexible signal combinations, to achieve a particular social goal, as well as attending to the attentional state of a conspecific and adapting responses accordingly (100). More simply, intentionality may be considered as a behavior that is goal-directed in nature (101), a definition that can be adopted in the following examples.

Young (5 to 7 year old) chimpanzees had to distinguish between actors who performed either accidental (spilling a cup of juice) or

intentional (pouring the juice onto the floor) actions but exhibited no clear preferences for either actor (102). When forced to choose between two food containers, chimpanzees and orangutans showed preferences for the box that received an intentional action, i.e., when a marker was placed on top of the box, instead of an accidental action, when the marker fell onto it (103). While these findings could not be explained by additional behavioral or gaze cues, they did show only a borderline significant effect and only for initial trials. Moreover, only one individual selected the intentional action above chance levels. In contrast, 3-year-old children readily distinguish intentional compared with accidental actions (103). A similar study with cotton-top tamarins, rhesus macaques, and chimpanzees revealed that they were more likely to inspect the container that received a goal-directed, grasping action rather than a non-goal-directed hand flop action, although for cotton-top tamarins, the reported *P* value indicates a nonsignificant effect (104). Critics of this study argue that inconsistencies in the training and testing procedures between species make interspecific comparisons difficult to interpret (105). Notably, attempts to replicate these findings in Tonkean macaques (*Macaca tonkeana*) and brown capuchins (*Sapajus apella*) were not successful (105). Overall, evidence from these studies that nonhuman primates differentiate between goal-directed and non-goal-directed actions in an experimental context is weak.

A different approach to assessing intentionality understanding examined apes' [chimpanzees, bonobos, gorillas, and orangutans (*Pongo abelii*)] responses to an experimenter when an action (opening a box with food) was consistent between conditions, but the prior context varied (106). In the experimental condition, the apes were accustomed to the experimenter opening a box to give them food from inside. In the control condition, the box was opened, but the food was given instead from a pocket of the experimenter's overalls, consequently rendering the box opening a meaningless action. After repeated exposure to either condition, the box was baited by a second experimenter in view of the apes and then the first experimenter returned to perform the same opening action. Chimpanzees and bonobos in the control condition (that did not receive the grape from the box) were quicker to leave than those that had previously received the grape from the box, suggesting that they anticipated the goal of the experimenter based on the context. In a follow-up study, the authors examined how chimpanzees would anticipate the goal of the experimenter based on a variable context. In this setup, the chimpanzees were accustomed to being fed food from one of two buckets. They were also occasionally fed food from the second bucket and thus learned to anticipate that when the experimenter stood up, she was about to approach the second bucket. In the experiment, before moving to the second feed location, the experimenter was either distracted by another researcher or interrupted feeding to conduct an unrelated action (dropping a clipboard). The experimenter's actions were otherwise identical in each condition, that is, she approached the second bucket regardless; accordingly, the chimpanzees would have to infer from the context alone whether the experimenter's goal was still to provide food from the second bucket. The results revealed that the chimpanzees adapted their behavior depending on the context: They were quicker to leave the testing area in the control condition, when the experimenter dropped the clipboard, compared with the other interruptions. These findings suggest that the chimpanzees were able to anticipate the goal of the experimenter in a context-specific manner and act accordingly.

An alternative method examines response to "willingness" of actors to provide food, compared with actors that are simply unable to

provide food. Differentiation between these conditions implies the ability to understand environmental constraints on agency. Chimpanzees were tested across a series of conditions reflecting either unwilling or unable behavior (107). They exhibited a higher rate of behavioral responses, such as poking fingers through the feeding holes, pushing on the apparatus, or knocking on the plexiglass divider to the experimenter, in the unwilling versus unable conditions but, notably, only when the experimenter remained present. This suggests that their behaviors were directed toward the experimenters rather than being a result of their own attempts to obtain the food, thereby supporting the notion that, in the unwilling condition, they could determine the lack of physical constraints on the agent. Similarly, Tonkean macaques spent more time begging for food and threatening the experimenter in the unwilling condition compared with unable or distracted conditions (108), and brown capuchins were quicker to leave the testing area in the unwilling compared with the unable condition where the experimenter dropped the grape, a difference that was not accounted for by differences in the motion of the action (109). Notably, the authors also examined the monkeys' responses to animate agents (human hands) compared with inanimate agents (a spoon stick) with the experimenters concealed behind a curtain. While the monkeys differentiated between the unwilling (hand withdrawn behind curtain) and unable (a second hand snatches the food) conditions for human hands, they did not make the same differentiation for the spoon (spoon withdrawn behind curtain versus a second stick snatches the food) (109). Consistent with cotton-top tamarins (67), this suggests that the monkeys have some concept of the role of animacy in agent actions. This experiment has also been extended to horses, revealing that similar to primates, they too differentiate between unwilling versus unable behavior in human actors (110). The difference between these findings and those of the accidental versus intentional studies may come down to the ability to account for why an action takes place.

The development of eye tracking studies has provided additional insight into how nonhuman primates process goal-directed actions. Kano and Hirata (111) used eye tracking to examine how apes (bonobos and chimpanzees) responded to videos when shown again on sequential days. They found that on the second view, subjects produced anticipatory fixations to regions where actions were about to take place. For example, they looked to a target door (as opposed to a distractor door) where a Kong character would emerge and to a target tool that an actor grabbed, even after it was spatially dislocated in the second viewing. These findings thus indicate that apes can recall events and form expectations about an actor's goal based on previous information.

Together, these findings present mixed evidence for primates' ability to understand intentionality (and limited evidence from nonprimates; Fig. 1, bottom left). Notably, when considering whether apes differentiate causally relevant and irrelevant actions, studies so far found that apes (chimpanzees, orangutans, and bonobos) do not emulate causally irrelevant actions (112–114). This suggests that, while they easily account for physical causality in their actions (i.e., which actions are necessary to achieve the required outcome), they are more limited in accounting for psychological causality (why an action, however irrelevant, is carried out in the first place). In addition to the above-reviewed findings on physical and temporal attributes, this would suggest that nonhuman primates may exhibit similar responses to humans under physical

causal scenarios but divergent responses when responding to intentional actions.

AGENCY ATTRIBUTIONS IN HUMANS AND NONHUMAN PRIMATES

The ability to ascribe agency to an action sequence is an important aspect of event cognition and is, as earlier discussed, a key part of constructing language around events. The evidence presented so far suggests that primates may indeed process events in a compositional manner, i.e., differentiating between agents and patients. However, given the human bias toward agency, to what extent could we expect to see this same bias in other primates? Evidence from monkeys indicates that they do differentiate between agents based on certain characteristics, such as animacy (67, 109). Taking this one step further, several studies have examined nonhuman primates' expectations about goal-directed actions using varying agent characteristics.

To examine apes' anticipation of goal-directed actions, bonobos, chimpanzees, and orangutans viewed video repetitions of a human hand reaching for one of two objects in a familiarization task (115). In a test sequence, they then saw the hand half reaching for the target object, except the target and distractor objects had now switched location. Using eye tracking, apes showed significantly higher anticipatory looks to the target object over the distractor, suggesting that they were successfully able to infer the goal of the agent, that is, assign agency to the human hand. Interestingly, when the same action sequence was repeated but with a mechanical claw instead of a human hand, anticipatory looks did not differ significantly between the two objects. This reflects earlier findings in human infants (116–118), with this difference ascribed to a lack of familiarity with the action of the claw or a lack of clues as to its goal (115). Although, more recent findings dispute the role of familiarity in goal anticipation (119).

Goal attribution may depend on the presence of conspecific-like features. Common marmosets (*Callithrix jacchus*) were familiarized to an object-grasping sequence from either a live demonstration from a human experimenter or a video of a conspecific (120). They were then shown either a congruent condition, demonstrator grasps the same object in a different location, or an incongruent condition, demonstrator grasps a different object in the same location. The expectation was that, if monkeys learned to anticipate the goal of the agent, then they should look longer in the incongruent condition since the goal is different. The authors found this to be the case for both human and conspecific demonstrators and when they replaced the animal agents with a quadrupedal robot. However, when the robot was covered with a box, they did not find the same effect. The authors suggest that the monkeys may rely on certain morphological cues to attribute agency to others, which could explain why an unfamiliar robot might be characterized as an agent, but a box is not.

A secondary consideration in assessing response to agency is considering not only how subjects respond to an action sequence based on agent features but how subjects respond to the agent itself. A comparative study with chimpanzees, human adults, and human infants presented subjects with videos of an agent (human or chimpanzee) performing an action (121). In a sequence showing a person pouring juice into a cup, chimpanzees, similar to adult humans, showed anticipation of the goal, that is, they looked at the cup before the juice was poured into it. Neither 12- nor 8-month-old infants showed a significant pattern of anticipatory looking. Notably, in comparison to humans, chimpanzees showed overall less attention

to the agent's face and more attention to the patient (the cup) in the action sequence. A similar difference in attention to the face was found when a video of a chimpanzee performing an action was viewed, but this species difference was only significant in the video onset, before contact with the goal was made. Perhaps the most interesting finding of this paper is that human adults attended overall relatively longer to the agent's face when a goal-directed action was performed as compared to when a non-goal-directed action was performed. In contrast, chimpanzees did not show any such differentiation.

In a follow-up study, the authors examined how humans and chimpanzees differ in their attention patterns in comparative conditions where the goal is either achieved or not achieved (for example, pouring juice onto the table instead of into a cup) (122). They found that, in the condition where the goal was not achieved, i.e., the outcome was incongruent with the subjects' expectations of the action outcome, adult humans, but neither chimpanzees, young children (3.5 years), nor infants (12 months), exhibited an increase in gaze toward the agent's face following the action. This was not the case when the action was congruent. This suggests that, when adult humans encounter an unexpected action outcome, they look to the agent of the action for additional social cues. The authors suggest that this could provide evidence of species differences in implicit versus explicit understanding of goal-directed actions. That is, humans may attempt to understand why an action did not meet their expectations, which involves higher-level cognitive functions such as advanced perspective taking, abilities that may not be cognitively accessible for nonhuman apes. This interpretation fits with the findings of earlier studies that indicate that both apes and monkeys struggle to explicitly differentiate goal-directed and non-goal-directed actions (102–105).

The studies outlined so far in this section indicate the ability of nonhuman primates to detect social agents. But how is the agent identified in the first place? At the simplest level, agency may be encoded from motion (61, 123). Movement is a primary source of information in humans to differentiate between animate and inanimate agents on the basis of pathway type (124). A visual bias for biological over nonbiological motion patterns is also evident in newborn infants (125) and visually naive chicks (126), suggesting that this may be an ancient and universal aspect of agency detection among animals. There are however species differences in how motion is processed: Mapping motion processing to cortical regions has revealed, for example, that humans have a larger area seemingly dedicated to the processing of motion, compared with monkeys [*Macaca* species; (123)].

In infants, a number of studies indicate that directed motion plays an important role in agent identification. Two-day-old infants looked longer toward actions that directed away from the body of an actor, when an object was present as compared to absent, suggesting an innate attentional bias toward goal-directed actions, which could help early on to determine agency (97). Similarly, 15- and 20-month-old infants exhibit better recall (measured through action imitation) for goal-directed actions (i.e., preceding a goal) than non-goal-directed actions (following the achievement of a goal) (127). Eight-month-old infants also appear to differentiate between continuous and discontinuous movement of launch displays (23). In addition, in 3- and 5-year-old children, younger, but not older, children looked longer when the agent and patient switched places in a dishabituation trial (an event sequence depicting the interaction of two balls) but only when the action sequence was in the correct

order (12). The authors interpreted this as evidence that the younger children could attribute intent in the correct, but not reverse, sequence order, whereas the older children could attribute intent in both sequence orders, which was reflected by their lack of differentiation between the two role-reversal sequences. Infants even differentiate moving animations as "agents" and "patients" based on motion direction (13), consistent with Michotte's (22) findings that people easily attribute causality even to simple action sequences. Indeed, evidence suggests that infants can identify an agent if it exhibits internal control over its actions (128). Seven-month-old children form different expectations about movement by animate and inanimate agents (129): In these scenarios, people changed their motion, either after coming into contact with another object (or person) or after stopping short of that contact. Infants looked longer in the no-contact condition when objects were depicted, but not people, suggesting that they discriminated animate from inanimate agents on the basis of changes in motion patterns.

It is clear, however, that motion alone is not enough to identify an agent since previous reports find a bias toward animacy (95), self-propulsion (67), directionality (97), and causality (64). What is not clear is to what extent a combination of these or additional variables play a role in agency detection. The nature of this detection may be very fine-tuned, as demonstrated using point-light displays (130). Point-light displays consist of several light dots placed around a body, which create the impression of that body, even when only the light dots are visible. People can accurately detect motion direction from a moving point-light display, and their accuracy tends to decrease below chance when the points are spatially scrambled. What is notable is that if all the body points except the feet are scrambled, people can still accurately detect motion direction, suggesting that specific features, such as feet, may be used as cues in identifying agents. An interesting investigation would therefore be to examine how subjects identify agents when certain features, such as the intensity or direction of an action, are manipulated.

Agency determination may also depend on nonvisible factors, such as a belief about who the agent is, in the absence of visual cues. In humans, manipulated beliefs are evidenced to result in altered behavior patterns (98). Participants were required to respond on a Simon task, where different colored targets are presented on either the left or right of the screen, and should be responded to with a left or right key press (e.g., green target, press left key, and red target, press right key). When there is incongruence between the side presented and the key required, reaction time slows. In this task, Tsai and colleagues (98) split the task so that the participant had to only respond to one color and a partner had to respond to the other. The subject believed that they were either working with a human partner or with a computer (in fact, in both instances, it was just a computer). While the condition of believed partner had no overall effect on response, participants were significantly slower to respond on the incongruent trials when they believed their partner to be a human, as opposed to a computer. The authors suggest that this points to social tuning of action planning, that is, perceiving intentionality of another's actions mediates one's own actions. Whether expected intentionality may be invoked here is debatable since the study focused on the believed interaction with a social or nonsocial agent but did not examine intentionality directly.

As we saw in the earlier section on event cognition, social and cultural variables influence language use in regard to agent description (39–41). Social variables are likely also a key aspect of agency

detection, such as sex, age, and social status. For example, in species for whom rank plays an important role, any differentiation between an agent and a patient may be defined by their rank differences, such that attentional bias is attributed to a higher-ranking individual [e.g., (131)], regardless of their role as agent or patient in the event. Equally, one may find a stronger bias for own-species or own-sex agents compared with outgroups. Accordingly, these are variables that should be accounted for in any comparative assessment of the agency detection hypothesis.

In sum, the above studies suggest that the ability to identify an agent is likely not dependent on one variable but draws on multiple sources of information, such as animacy, motion, action type, and agent-specific features, including physical features, social variables, and internal beliefs (Fig. 1, bottom right). While it is highly possible that, similar to humans, other species will exhibit a visual bias to the agent (compared with the patient), this is likely dependent on the specific features that agents and patients exhibit. Social information, in particular, seems to be key in identifying, responding to, and contextualizing agents. However, it is clear that there is much still to understand about the role of social status in agency attribution. It would be especially important for future research to disentangle social, linguistic, and cultural overlaps, which may expose internal biases about traditional agent/patient roles. Indeed, since causality and its associated features seem to be at the basis of semantic role attribution, a first step to understanding agency attribution is to examine the role of low-level features on implicit responses by manipulating causal information. Given the possible important role of social status (39, 132) and variables such as age and gender (40) in how people attribute agency, it could also be beneficial to examine cross-cultural differences in agency attribution depending on status asymmetries of, or social beliefs about, agents and patients.

CONCLUSIONS AND FUTURE DIRECTIONS

The above work encompasses discussion of primates' abilities to process physical causality, temporal information, social knowledge, intent, and features of agency. Given the parallels found between humans and nonhuman primates in these domains, the evidence leans in favor of the agency detection hypothesis, that is, that agent-based event decomposition is not unique to humans. We therefore conclude with the argument that the ability to distinguish semantic roles was a prerequisite for compositionality in human event expressions and emerged at least early on in the primate lineage or likely earlier. We propose that a basic neural framework to automatically detect low-level features found in causal relationships already occurred early on in the animal kingdom among species with bilateral symmetry. This provided the basis for later emergence of causal detection, orientation, and reasoning, which was likely driven by social factors: Given the strong role that sociality seems to play in agent detection and attending to goal-directed actions, it seems likely that social selection pressures and the need to keep track of social events drove the ability to represent agents and patients distinctly but jointly in event scenarios. Given some apparent differences between humans and apes in responding to agents (122), it is possible that humans have experienced divergent selection pressures for attention to events, alongside other cognitive developments, that ultimately allowed for language to evolve.

To build a clearer picture of the evolution of event cognition, we need to clarify the following points: (i) To what extent does event

processing in nonhuman primates reflect event processing in humans? Are there interspecific parallels, or differences, in attentional differences to the agent versus the patient? (ii) How ancient is the agent/patient distinction, and did it develop further under social pressures? To answer this question, further research from a broader phylogenetic perspective would be required. (iii) What is the relative importance of physical and psychological causation in event cognition? Are these prerequisites to differentiating semantic roles within events? (iv) If other species exhibit event processing parallels with humans, then why do they communicate about events holistically, as opposed to syntactically? (v) To understand how event cognition developed in humans, one should also ask whether semantic role attribution precedes the development of language in ontogeny. (vi) Last, to fully understand the developmental processes of event cognition in humans, it would be important to assess how children with atypical social development, such as those on the autistic spectrum, process events.

Given these points, there are a number of steps that can be taken to expand on the current state of knowledge surrounding event cognition. The first step, following a traditional comparative cognition approach, is to directly test the agency detection hypothesis by examining event processing in our closest relatives, the great apes, and to draw direct comparisons with humans. Comparative research is most productive if different species can be tested with ecologically relevant problems presented to them in a comparable way (133–135). Recent advances in eye tracking (111, 121, 136, 137) and touch-screen technology (66, 138, 139) render such an approach both feasible and simple to implement in a noninvasive setting. For example, eye tracking studies suggest that apes may be able to perform high-level theory of mind assessments, such as understanding false belief, at an implicit level (136, 140, 141), and these implicit responses may also extend to Old World monkeys (142). These findings, which have advanced our understanding of apes' perceptions of others' actions through the use of eye tracking, indicate that exploring fine-grained species' differences in temporal event sequences is a promising avenue for understanding the evolution of event cognition. The use of touchscreens could expound on eye tracking by examining explicit preferences (66) for agents or patients. One problem with many cognitive studies in primates is that they examine responses to human actors rather than to conspecifics, which could confound results (143). Thus, future research in event cognition should be sure to examine responses to species-specific stimuli. It should also avoid relying too heavily on a pan-centric approach (144). Comparative work in apes could therefore be complemented with studies that examine links between event processing and simple syntactic communication, such as in birds [e.g., (145)], as well as drawing comparisons between social and asocial species from different taxa. Last, choosing a paradigm that is easily implementable across different species (such as eye tracking) provides a fully comparative approach to examining directly the similarities and differences that can be seen when viewing events in individuals of different species, ages, and abilities.

REFERENCES AND NOTES

1. J. M. Zacks, Event perception and memory. *Rev. Psychol.* **71**, 165–191 (2020).
2. J. H. Greenberg, Some universals of grammar with particular reference to the order of meaningful elements, in *Universals of Language*, J. H. Greenberg, Ed. (MIT Press, 1963), pp. 73–113.
3. C. J. Fillmore, The case for case, in *Universals in Linguistic Theory*, E. Bach, R. T. Harms, Eds. (Holt, Rinehart & Winston, 1968).
4. B. Bickel, Grammatical relations typology, in *The Oxford Handbook of Language Typology*, J. J. Song, Ed. (Oxford Univ. Press, 2011), pp. 399–444.

5. M. S. Dryer, Order of subject, object, and verb, in *The World Atlas of Language Structures Online*, M. S. Dryer, M. Haspelmath, Eds. (Max-Planck-Institut für Evolutionäre Anthropologie, 2013).
6. L. Rissman, A. Majid, Thematic roles: Core knowledge or linguistic construct? *Psychon. Bull. Rev.* **26**, 1850–1869 (2019).
7. A. Webb, A. Knott, M. R. Macaskill, Eye movements during transitive action observation have sequential structure. *Acta Psychol. (Amst)* **133**, 51–56 (2010).
8. F. Wilson, A. Papafragou, A. Bunger, J. Trueswell, Rapid extraction of event participants in caused motion events, in *Proceedings of the Annual Meeting of the Cognitive Science* (eScholarship, 2011), pp. 1206–1211.
9. S. Sauppe, E. Norcliffe, A. E. Konopka, R. D. van Valin, S. C. Levinson, Dependencies first: Eye tracking evidence from sentence production in Tagalog, in *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*, M. Knauff, M. Pauen, N. Sebanz, I. Wachsmuth, Eds. (eScholarship, 2013), pp. 1265–1270.
10. A. Papafragou, J. Hulbert, J. Trueswell, Does language guide event perception? Evidence from eye movements. *Cognition* **108**, 155–184 (2008).
11. S. Sauppe, M. Flecken, Speaking for seeing: Sentence structure guides visual event apprehension. *Cognition* **206**, 104516 (2021).
12. V. Dasser, I. Ulbaek, D. Premack, The perception of intention. *Science* **243**, 365–367 (1989).
13. M. Galazka, P. Nyström, Infants' preference for individual agents within chasing interactions. *J. Exp. Child Psychol.* **147**, 53–70 (2016).
14. D. Kemmerer, The cross-linguistic prevalence of SOV and SVO word orders reflects the sequential and hierarchical representation of action in Broca's area. *Lang. Linguist. Compass* **6**, 50–66 (2012).
15. C. Sinha, P. Gärdenfors, Time, space, and events in language and cognition: A comparative view. *Ann. N. Y. Acad. Sci.* **1326**, 72–81 (2014).
16. K. Zuberbühler, Evolutionary roads to syntax. *Anim. Behav.* **151**, 259–265 (2019).
17. C. J. Fillmore, Frame semantics, in *Linguistics in the Morning Calm* (Hanshin Publishing Co, 1982), pp. 111–37.
18. D. R. Dowty, *Word Meaning and Montague Grammar* (Reidel, 1979).
19. M. D. Hauser, N. Chomsky, W. T. Fitch, The faculty of language: What is it, who has it, and how did it evolve? *Science* **298**, 1569–1579 (2002).
20. S. Kirby, Culture and biology in the origins of linguistic structure. *Psychon. Bull. Rev.* **24**, 118–137 (2017).
21. M. Christiansen, N. Chater, Language as shaped by the brain. *Behav. Brain Sci.* **31**, 489–509 (2008).
22. A. Michotte, *The Perception of Causality* (Basic Books, 1963).
23. A. M. Leslie, The perception of causality in infants. *Perception* **11**, 173–186 (1982).
24. B. J. Scholl, P. D. Tremoulet, Perceptual causality and animacy. *Trends Cogn. Sci.* **4**, 299–309 (2000).
25. P. Moors, J. Wagemans, L. De-Wit, Causal events enter awareness faster than non-causal events. *PeerJ.* **5**, e2932 (2017).
26. E. C. Tecwyn, C. Bechliyanidis, D. Lagnado, C. Hoerl, S. Lorimer, E. Blakey, T. McCormack, M. J. Buehner, Causality influences children's and adults' experience of temporal order. *Dev. Psychol.* **56**, 739–755 (2020).
27. A. Hafri, C. Firestone, The perception of relations. *Trends Cogn. Sci.* **25**, 475–492 (2021).
28. J. Gerwien, M. Flecken, First things first? Top-down influences on event apprehension, in *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*, A. Papafragou, D. Grodner, D. Mirman, J. C. Trueswell, Eds. (Max Planck Institute for Psycholinguistics, 2016), pp. 2633–2638.
29. C. Dobel, H. Gumnior, J. Bölte, P. Zwitserlood, Describing scenes hardly seen. *Acta Psychol. (Amst)* **125**, 129–143 (2007).
30. A. Hafri, A. Papafragou, J. C. Trueswell, Getting the gist of events: Recognition of two-participant actions from brief displays. *J. Exp. Psychol.* **142**, 880–905 (2013).
31. A. Hafri, J. C. Trueswell, B. Strickland, Encoding of event roles from visual scenes is rapid, spontaneous, and interacts with higher-level visual processing. *Cognition* **175**, 36–52 (2018).
32. A. Jessop, F. Chang, Thematic role information is maintained in the visual object-tracking system. *Q. J. Exp. Psychol.* **73**, 146–163 (2020).
33. E. Únal, C. Richards, J. C. Trueswell, A. Papafragou, Representing agents, patients, goals and instruments in causative events: A cross-linguistic investigation of early language and cognition. *Dev. Sci.* **24**, e13116 (2021).
34. E. S. Spelke, K. D. Kinzler, Core knowledge. *Dev. Sci.* **1**, 89–96 (2007).
35. D. J. Napoli, R. Sutton-spence, Order of the major constituents in sign languages: Implications for all language. *Front. Psychol.* **5**, 376 (2014).
36. B. Bickel, A. Witzlack-Makarevich, K. K. Choudhary, M. Schlesewsky, I. Bornkessel-Schlesewsky, The neurophysiology of language processing shapes the evolution of grammar: Evidence from case marking. *PLOS ONE* **10**, e0132819 (2015).
37. S. DeLancey, An interpretation of split ergativity and related patterns. *Language* **57**, 626–657 (1981).
38. S. Kuno, *Functional Syntax: Anaphora, Discourse and Empathy* (University of Chicago Press, 1987).
39. A. Duranti, *From Grammar to Politics: Linguistic Anthropology in a Western Samoan Village* (University of California Press, 1994).
40. M. Formanowicz, J. Roessel, C. Saitner, A. Maass, Verbs as linguistic markers of agency: The social side of grammar. *Eur. J. Soc. Psychol.* **47**, 566–579 (2017).
41. M. W. Morris, K. Peng, Culture and cause: American and Chinese attributions for social and physical events. *J. Pers. Soc. Psychol.* **67**, 949–971 (1994).
42. P. Wolff, K. J. Holmes, Linguistic relativity. *Wiley Interdiscip. Rev. Cogn. Sci.* **2**, 253–265 (2011).
43. S. C. Levinson, From outer to inner space: Linguistic categories and non-linguistic thinking, in *Language and Conceptualization*, J. Nuyts, E. Pederson, Eds. (Cambridge Univ. Press, 1997), pp. 13–45.
44. S. Goldin-meadow, C. Mylander, Spontaneous sign systems created by deaf children in two cultures. *Nature* **391**, 279–281 (1998).
45. A. A. Ivanova, Z. Mineroff, V. Zimmerer, N. Kanwisher, R. Varley, E. Fedorenko, The language network is recruited but not required for nonverbal event semantics. *Neurobiol. Lang.* **2**, 176–201 (2021).
46. H. Hwang, E. Kaiser, The role of the verb in grammatical function assignment in English and Korean. *J. Exp. Psychol.* **40**, 1363–1376 (2014).
47. M. Flecken, J. Gerwien, M. Carroll, C. von Stutterheim, Analyzing gaze allocation during language planning: A cross-linguistic study on dynamic events. *Lang. Cogn.* **7**, 138–166 (2015).
48. S. Sauppe, C. K. Kamal, N. Giroud, D. E. Blasi, E. Norcliffe, S. Bhattamishra, M. Gulati, A. Egrtzeig, I. Bornkessel-Schlesewsky, M. Meyer, B. Bickel, Neural signatures of syntactic variation in speech planning. *PLOS Biol.* **19**, e3001038 (2021).
49. F. Ferreira, G. Rehrig, Linearisation during language production: Evidence from scene meaning and saliency maps. *Lang. Cogn. Neurosci.* **34**, 1129–1139 (2019).
50. F. S. Haupt, M. Schlesewsky, D. Roehm, A. D. Friederici, I. Bornkessel-schlesewsky, The status of subject-object reanalyses in the language comprehension architecture. *J. Memory Lang.* **59**, 54–96 (2008).
51. S. B. Demiral, M. Schlesewsky, I. Bornkessel-schlesewsky, On the universality of language comprehension strategies: Evidence from Turkish. *Cognition* **106**, 484–500 (2008).
52. L. Wang, M. Schlesewsky, B. Bickel, I. Bornkessel-Schleswsky, Exploring the nature of the 'subject'-preference: Evidence from the online comprehension of simple sentences in Mandarin Chinese. *Lang. Cogn. Process.* **24**, 1180–1226 (2009).
53. A. L. Jouen, T. M. Ellmore, C. J. Madden, C. Pallier, P. F. Dominey, J. Ventre-Dominey, Beyond the word and image: Characteristics of a common meaning system for language and vision revealed by functional and structural imaging. *Neuroimage* **106**, 72–85 (2015).
54. Z. Hu, H. Yang, Y. Yang, S. Nishida, C. Madden-Lombardi, J. Ventre-Dominey, P. F. Dominey, K. Ogawa, Common neural system for sentence and picture comprehension across languages: A Chinese–Japanese bilingual study. *Front. Hum. Neurosci.* **13**, 380 (2019).
55. J. Feldman, S. Narayanan, Embodied meaning in a neural theory of language. *Brain Lang.* **89**, 385–392 (2004).
56. S. M. Frankland, J. D. Greene, An architecture for encoding sentence meaning in left mid-superior temporal cortex. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 11732–11737 (2015).
57. G. Thierry, C. J. Price, Dissociating verbal and nonverbal conceptual processing in the human brain. *J. Cogn. Neurosci.* **18**, 1018–1028 (2006).
58. S. Goldin-Meadow, Thought before language: Do we think ergative? in *Language in Mind: Advances in the Study of Language and Thought*, D. Gentner, S. Goldin-Meadow, Eds. (MIT Press, 2003), pp. 493–522.
59. M. C. Macdonald, How language production shapes language form and comprehension. *Front. Psychol.* **4**, 226 (2013).
60. I. Meir, M. Aronoff, C. Börstell, S. Hwang, D. Ilkbasaran, I. Kastner, R. Lepic, A. L. Ben-Basat, C. Padden, W. Sandler, 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 (2017).
61. U. Frith, C. Frith, The social brain: Allowing humans to boldly go where no other species has been. *Philos. Trans. R. Soc. B Biol. Sci.* **365**, 165–176 (2010).
62. M. G. Paulin, J. Cahill-lane, Events in early nervous system evolution. *Topics Cogn. Sci.* **13**, 25–44 (2021).
63. P. Cisek, Resynthesizing behavior through phylogenetic refinement. *Atten. Percept. Psychophys.* **81**, 2265–2287 (2019).
64. E. Mascalzoni, L. Regolin, G. Vallortigara, Innate sensitivity for self-propelled causal agency in newly hatched chicks. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 4483–4485 (2010).
65. J. Abdai, B. Ferdinandy, Perception of animacy in dogs and humans. *Biol. Lett.* **13**, 20170156 (2017).
66. V. A. D. Wilson, C. Kade, J. Fischer, Testing the relationship between looking time and choice preference in long-tailed macaques. *Anim. Behav. Cogn.* **8**, 351–375 (2021).
67. M. D. Hauser, A nonhuman primate's expectations about object motion and destination: The importance of self-propelled movement and animacy. *Dev. Psychol.* **1**, 31–37 (1998).

68. M. J. Rochat, E. Serra, L. Fadiga, V. Gallese, The evolution of social cognition: Goal familiarity shapes monkeys' action understanding. *Curr. Biol.* **18**, 227–232 (2008).
69. J. N. Wood, D. D. Glynn, M. D. Hauser, The uniquely human capacity to throw evolved from a non-throwing primate: An evolutionary dissociation between action and perception. *Biol. Lett.* **3**, 360–365 (2007).
70. J. N. Wood, D. D. Glynn, M. D. Hauser, Rhesus monkeys' understanding of actions and goals. *Soc. Neurosci.* **3**, 60–68 (2008).
71. T. Cacchione, J. Call, Do gorillas (*Gorilla gorilla*) and orangutans (*Pongo pygmaeus*) fail to represent objects in the context of cohesion violations? *Cognition* **116**, 193–203 (2010).
72. Y. Munakata, L. R. Santos, E. S. Spelke, M. D. Hauser, R. C. O. Reilly, Visual representation in the wild: How rhesus monkeys parse objects. *J. Cogn. Neurosci.* **13**, 44–58 (2001).
73. E. W. Cheries, S. R. Mitroff, K. Wynn, B. J. Scholl, Cohesion as a constraint on object persistence in infancy. *Dev. Sci.* **11**, 427–432 (2008).
74. T. Cacchione, H. Krist, Recognizing impossible object relations: Intuitions about support in chimpanzees (*Pan troglodytes*). *J. Comp. Psychol.* **118**, 140–148 (2004).
75. C. Murai, M. Tanaka, M. Sakagami, Physical intuitions about support relations in monkeys (*Macaca fuscata*) and apes (*Pan troglodytes*). *J. Comp. Psychol.* **125**, 216–226 (2011).
76. A. Needham, R. Baillargeon, Intuitions about support in 4.5-month-old infants. *Cognition* **47**, 121–148 (1993).
77. S. O'Connell, R. I. M. Dunbar, The perception of causality in chimpanzees (*Pan spp.*). *Anim. Cogn.* **8**, 60–66 (2005).
78. A. Dickinson, D. Shanks Instrumental action and causal representation, in *Causal Cognition: A Multidisciplinary Debate*, D. Sperber, D. Premack, A. J. Premack, Eds. (Oxford Univ. Press, 1996), pp. 5–25.
79. H. Kummer, Causal knowledge in animals, in *Causal Cognition: A Multidisciplinary Debate*, D. Sperber, D. Premack, A. J. Premack, Eds. (Oxford Univ. Press, 1996), pp. 26–39.
80. L. Wang, S. Zuo, Y. Cai, B. Zhang, H. Wang, Y. Zhou, S. C. Kwok, Fallacious reversal of event-order during recall reveals memory reconstruction in rhesus monkeys. *Behav. Brain Res.* **394**, 112830 (2020).
81. G. Martin-Ordas, D. Haun, F. Colmenares, J. Call, Keeping track of time: Evidence for episodic-like memory in great apes. *Anim. Cogn.* **13**, 331–340 (2010).
82. M. Donald, The slow process: A hypothetical cognitive adaptation for distributed cognitive networks. *J. Physiol.* **101**, 214–222 (2007).
83. D. L. Cheney, R. M. Seyfarth, J. B. Silk, The responses of female baboons (*Papio cynocephalus ursinus*) to anomalous social interactions: Evidence for causal reasoning? *J. Comp. Psychol.* **109**, 134–141 (1995).
84. R. M. Seyfarth, D. L. Cheney, T. J. Bergman, Primate social cognition and the origins of language. *Trends Cogn. Sci.* **9**, 264–266 (2005).
85. T. J. Bergman, J. C. Beehner, D. L. Cheney, R. M. Seyfarth, Hierarchical classification by rank and kinship in baboons. *Science* **302**, 1234–1236 (2003).
86. K. E. Slocombe, T. Kaller, J. Call, K. Zuberbuehler, Chimpanzees extract social information from agonistic screams. *PLOS ONE* **5**, e11473 (2010).
87. A. M. Overduin-de Vries, F. A. A. Bakker, B. M. Spruijt, E. H. M. Sterck, Male long-tailed macaques (*Macaca fascicularis*) understand the target of facial threat. *Am. J. Primatol.* **78**, 720–730 (2016).
88. Z. Clay, L. Ravaux, F. de Waal, K. Zuberbuehler, Bonobos (*Pan paniscus*) vocally protest against violations of social expectations. *J. Comp. Psychol.* **130**, 44–54 (2016).
89. C. R. Von Rohr, C. P. van Schaik, A. Kislasing, J. M. Burkart, Chimpanzees' bystander reactions to infanticide an evolutionary precursor of social norms? *Hum Nat.* **26**, 143–160 (2015).
90. R. McFarland, H. Roebuck, Y. Yan, B. Majolo, W. Li, K. Guo, Social interactions through the eyes of macaques and humans. *PLOS ONE* **8**, e56437 (2013).
91. B. M. Waller, J. Whitehouse, J. Micheletta, Macaques can predict social outcomes from facial expressions. *Anim. Cogn.* **19**, 1031–1036 (2016).
92. L. H. Howard, K. E. Wagner, A. L. Woodward, S. R. Ross, L. M. Hopper, Social models enhance apes' memory for novel events. *Sci. Rep.* **7**, 40926 (2017).
93. L. H. Howard, A. L. Woodward, Human actions support infant memory. *J. Cogn. Dev.* **20**, 772–789 (2019).
94. L. H. Howard, C. Festa, E. Lonsdorf, M. College, Through their eyes: The influence of social models on attention and memory in capuchin monkeys (*Sapajus apella*). *J. Comp. Psychol.* **132**, 210–219 (2018).
95. J. New, L. Cosmides, J. Tooby, Category-specific attention for animals reflects ancestral priorities, not expertise. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 16598–16603 (2007).
96. M. H. Ferkin, A. Combs, J. DelBarco-Trillo, A. A. Pierce, S. Franklin, Meadow voles, *Microtus pennsylvanicus*, have the capacity to recall the "what", "where", and "when" of a single past event. *Anim. Cogn.* **11**, 147–159 (2008).
97. L. Craighero, I. Leo, C. Umiltà, F. Simion, Newborns' preference for goal-directed actions. *Cognition* **120**, 26–32 (2011).
98. C. Tsai, W. Kuo, D. L. Hung, O. J. L. Tzeng, Action co-representation is tuned to other humans. *J. Cogn. Neurosci.* **20**, 2015–2024 (2008).
99. J. Smortchkova, Seeing goal-directedness: A case for social perception. *Br. J. Philos. Sci.* **71**, 855–879 (2020).
100. M. Tomasello, J. Call, J. Warren, G. T. Frost, M. Carpenter, K. Nagell, The ontogeny of chimpanzee gestural signals. *Evol. Commun.* **1**, 223–259 (1997).
101. D. C. Dennett, Intentional systems. *J. Philos.* **68**, 87–106 (1971).
102. D. J. Povinelli, H. K. Perilloux, J. E. Reaux, D. T. Bierschwale, Young and juvenile chimpanzees' (*Pan troglodytes*) reactions to intentional versus accidental and inadvertent actions. *Behav. Processes* **42**, 205–218 (1998).
103. J. Call, M. Tomasello, Distinguishing intentional from accidental actions in orangutans (*Pongo pygmaeus*), chimpanzees (*Pan troglodytes*) and human children (*Homo sapiens*). *J. Comp. Psychol.* **112**, 192–206 (1998).
104. J. N. Wood, D. D. Glynn, B. C. Phillips, M. D. Hauser, The perception of rational, goal-directed action in nonhuman primates. *Science* **317**, 1402–1405 (2007).
105. M. Costes-Thiré, M. Levé, P. Uhlrich, C. Pasquaretta, A. de Marco, B. Thierry, Evidence that monkeys (*Macaca tonkeana* and *Sapajus apella*) read moves, but no evidence that they read goals. *J. Comp. Psychol.* **129**, 304–310 (2015).
106. D. Buttelmann, S. Schütte, M. Carpenter, J. Call, M. Tomasello, Great apes infer others' goals based on context. *Anim. Cogn.* **15**, 1037–1053 (2012).
107. J. Call, B. Hare, M. Carpenter, M. Tomasello, Unwilling versus "unable": Chimpanzees' understanding of human intentional action. *Dev. Sci.* **7**, 488–498 (2004).
108. C. Canteloup, H. Meunier, Unwilling versus "unable": Tonkean macaques' understanding of human goal-directed actions. *PeerJ.* **5**, e3227 (2017).
109. W. Phillips, J. L. Barnes, N. Mahajan, M. Yamaguchi, L. R. Santos, Unwilling versus "unable": Capuchin monkeys' (*Cebus apella*) understanding of human intentional action. *Dev. Sci.* **12**, 938–945 (2009).
110. M. Trösch, E. Bertin, L. Calandrea, R. Nowak, L. Lansade, Unwilling or willing but unable: Can horses interpret human actions as goal directed? *Anim. Cogn.* **23**, 1035–1040 (2020).
111. F. Kano, S. Hirata, Great apes make anticipatory looks based on long-term memory of single events. *Curr. Biol.* **25**, 2513–2517 (2015).
112. V. Horner, A. Whiten, Causal knowledge and imitation/emulation switching in chimpanzees (*Pan troglodytes*) and children (*Homo sapiens*). *Anim. Cogn.* **8**, 164–181 (2005).
113. M. Nielsen, E. W. E. Susianto, Failure to find over-imitation in captive orangutans (*Pongo pygmaeus*): Implications for our understanding of cross-generation information transfer, in *Developmental Psychology*, J. Håkansson, Eds. (Nova Science Publishers, 2010), pp. 153–67.
114. Z. Clay, C. Tennie, Is overimitation a uniquely human phenomenon? Insights from human children as compared to bonobos. *Child Dev.* **89**, 1535–1544 (2018).
115. F. Kano, J. Call, Great apes generate goal-based action predictions: An eye-tracking study. *Psychol. Sci.* **25**, 1691–1698 (2014).
116. A. L. Woodward, Infants selectively encode the goal object of an actor's reach. *Cognition* **69**, 1–34 (1998).
117. M. M. Daum, G. Gredebäck, The development of grasping comprehension in infancy: Covert shifts of attention caused by referential actions. *Exp. Brain Res.* **208**, 297–307 (2011).
118. E. N. Cannon, A. L. Woodward, Infants generate goal-based action predictions. *Dev. Sci.* **15**, 292–298 (2012).
119. K. Ganglmayer, M. Attig, M. M. Daum, M. Paulus, Infants' perception of goal-directed actions: A multi-lab replication reveals that infants anticipate paths and not goals. *Infant Behav. Dev.* **57**, 101340 (2019).
120. A. Kupferberg, S. Glasauer, J. M. Burkart, Do robots have goals? How agent cues influence action understanding in non-human primates. *Behav. Brain Res.* **246**, 47–54 (2013).
121. M. Myowa-Yamakoshi, C. Scola, S. Hirata, Humans and chimpanzees attend differently to goal-directed actions. *Nat. Commun.* **3**, 693 (2012).
122. M. Myowa-Yamakoshi, C. Yoshida, S. Hirata, Humans but not chimpanzees vary face-scanning patterns depending on contexts during action observation. *PLOS ONE* **10**, 1–15 (2015).
123. Z. Kourtzi, B. Kreklerberg, R. J. A. van Wezel, Linking form and motion in the primate brain. *Trends Cogn. Sci.* **12**, 230–236 (2008).
124. R. Gelman, F. Durgin, L. Kaufman, Distinguishing between animates and inanimates: Not by motion alone, in *Causal Cognition: A Multidisciplinary Debate*, D. Sperber, D. Premack, A. J. Premack, Eds. (Clarendon Press, 1996).
125. F. Simion, L. Regolin, H. Bulf, A predisposition for biological motion in the newborn baby. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 809–813 (2008).
126. G. Vallortigara, L. Regolin, F. Marconato, Visually inexperienced chicks exhibit spontaneous preference for biological motion patterns. *PLoS Biol.* **3**, e208 (2005).
127. L. L. Travis, Goal-based organization of event memory in toddlers, in *Developmental Spans in Event Comprehension and Representation: Bridging Fictional and Actual Events*, P. van den Broek, P. J. Bauer, T. Bourg, Eds. (Lawrence Erlbaum Associates Inc., 1997), pp. 111–138.
128. R. Baillargeon, R. M. Scott, L. Bian, Psychological reasoning in infancy. *Annu. Rev. Psychol.* **67**, 159–186 (2016).

129. A. Woodward, A. Phillips, E. S. Spelke, Infants' expectations about the motions of inanimate vs. animate objects, in *Proceedings of the Cognitive Science Society* (Erlbaum; 1993).
130. N. F. Troje, C. Westhoff, The inversion effect in biological motion perception: Evidence for a 'life detector?'. *Curr. Biol.* **16**, 821–824 (2006).
131. L. S. Lewis, F. Kano, J. M. G. Stevens, J. G. Dubois, J. Call, C. Krupenye, Bonobos and chimpanzees preferentially attend to familiar members of the dominant sex. *Anim. Behav.* **177**, 193–206 (2021).
132. C. N. DeWall, J. K. Maner, High status men (but not women) capture the eye of the beholder. *Evolutionary Psychology.* **6**, 328–341 (2008).
133. E. Herrmann, J. Call, M. V. Hernandez-Lloreda, B. Hare, M. Tomasello, Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *Science* **317**, 1360–1366 (2007).
134. E. L. Maclean, B. Hare, C. L. Nunn, E. Addessi, F. Amici, R. C. Anderson, F. Aureli, J. M. Baker, A. E. Bania, A. M. Barnard, N. J. Boogert, E. M. Brannon, E. E. Bray, J. Bray, L. J. N. Brent, J. M. Burkart, J. Call, J. F. Cantlon, L. G. Cheke, N. S. Clayton, M. M. Delgado, L. J. Di Vincenti, K. Fujita, E. Herrmann, C. Hiramatsu, L. F. Jacobs, K. E. Jordan, J. R. Laude, K. L. Leimgruber, E. J. E. Messer, A. C. de A. Moura, L. Ostojčić, A. Picard, M. L. Platt, J. M. Plotnik, F. Range, S. M. Reader, R. B. Reddy, A. A. Sandel, L. R. Santos, K. Schumann, A. M. Seed, K. B. Sewall, R. C. Shaw, K. E. Slocombe, Y. Su, A. Takimoto, J. Tan, R. Tao, C. P. van Schaik, Z. Virányi, E. Visalberghi, J. C. Wade, A. Watanabe, J. Widness, J. K. Young, T. R. Zentall, Y. Zhao, The evolution of self-control. *Proc. Natl. Acad. Sci. U.S.A.*, E2140–E2148 (2014).
135. M. Primates, D. M. Altschul, M. J. Beran, M. Bohn, J. Call, S. J. Duguid, C. L. Egelkamp, C. Fichtel, J. Fischer, M. Flessert, D. Hanus, D. B. M. Haun, L. M. Haux, R. A. Hernandez-Aguilar, E. Herrmann, L. M. Hopper, M. Joly, F. Kano, S. Keupp, A. P. Melis, A. M. Rodrigo, S. R. Ross, A. Sánchez-Amaro, Y. Sato, V. Schmitt, M. K. Schweinfurth, A. M. Seed, D. Taylor, C. J. Völter, E. Warren, J. Watzek, Establishing an infrastructure for collaboration in primate cognition research. *PLOS ONE* **14**, e0223675 (2019).
136. C. Krupenye, F. Kano, S. Hirata, J. Call, M. Tomasello, Great apes anticipate that other individuals will act according to false beliefs. *Science* **354**, 110–114 (2016).
137. L. M. Hopper, R. A. Gulli, L. H. Howard, F. Kano, C. Krupenye, A. M. Ryan, A. Paukner, The application of noninvasive, restraint-free eye-tracking methods for use with nonhuman primates. *Behav. Res. Methods*, 1003–1030 (2021).
138. L. Parr, W. D. Hopkins, F. B. de Waal, The perception of facial expressions by chimpanzees, *Pan troglodytes*. *Evol. Commun.* **2**, 1–23 (1998).
139. C. L. Egelkamp, S. R. Ross, A review of zoo-based cognitive research using touchscreen interfaces. *Zoo Biol.*, 220–235 (2019).
140. C. Krupenye, F. Kano, S. Hirata, J. Call, C. Krupenye, A test of the submentalizing hypothesis: Apes' performance in a false belief task inanimate control. *Commun. Integr. Biol.* **10**, 1–5 (2017).
141. F. Kano, C. Krupenye, S. Hirata, J. Call, Eye tracking uncovered great apes' ability to anticipate that other individuals will act according to false beliefs. *Commun. Integr. Biol.* **10**, 1–7 (2017).
142. T. Hayashi, R. Akikawa, K. Kawasaki, J. Egawa, T. Minamimoto, K. Kobayashi, S. Kato, Y. Hori, Y. Nagai, A. Iijima, T. Someya, I. Hasegawa, Macaques exhibit implicit gaze bias anticipating others' false-belief-driven actions via medial prefrontal cortex. *Cell Rep.* **30**, 4433–4444.e5 (2020).
143. H. L. Marsh, M. Legerstee, Awareness of goal-oriented behavior during infancy and early childhood, in human- and non-human primates. *Infant Behav. Dev.* **48**, 30–37 (2017).
144. A. R. Lameira, J. Call, Understanding language evolution: Beyond pan-centrism. *Bioessays* **42**, e1900102 (2020).
145. S. Engesser, A. R. Ridley, S. W. Townsend, Meaningful call combinations and compositional processing in the southern pied babbler. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 5976–5981 (2016).

Acknowledgments

Funding: This research received funding from a Swiss National Science Foundation Project grant 310030_185324 (K.Z.) and the NCCR Evolving Language, Swiss National Science Foundation Agreement #51NF40_180888 (B.B. and K.Z.). **Ethics statement:** No ethical approval was required for this paper. **Author contributions:** Conceptualization: K.Z., V.A.D.W., and B.B. Literature search: V.A.D.W. Supervision: K.Z. and B.B. Writing (original draft): V.A.D.W. Writing (review and editing): V.A.D.W., K.Z., and B.B. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper.

Submitted 23 December 2021

Accepted 5 May 2022

Published 22 June 2022

10.1126/sciadv.abn8464