



**University of  
Zurich**<sup>UZH</sup>

**Zurich Open Repository and  
Archive**

University of Zurich  
University Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2024

---

## **A universal preference for animate agents in hominids**

Brocard, Sarah ; Wilson, Vanessa A D ; Berton, Chloé ; Zuberbühler, Klaus ; Bickel, Balthasar

DOI: <https://doi.org/10.1016/j.isci.2024.109996>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-259736>

Journal Article

Published Version



The following work is licensed under a Creative Commons: Attribution 4.0 International (CC BY 4.0) License.

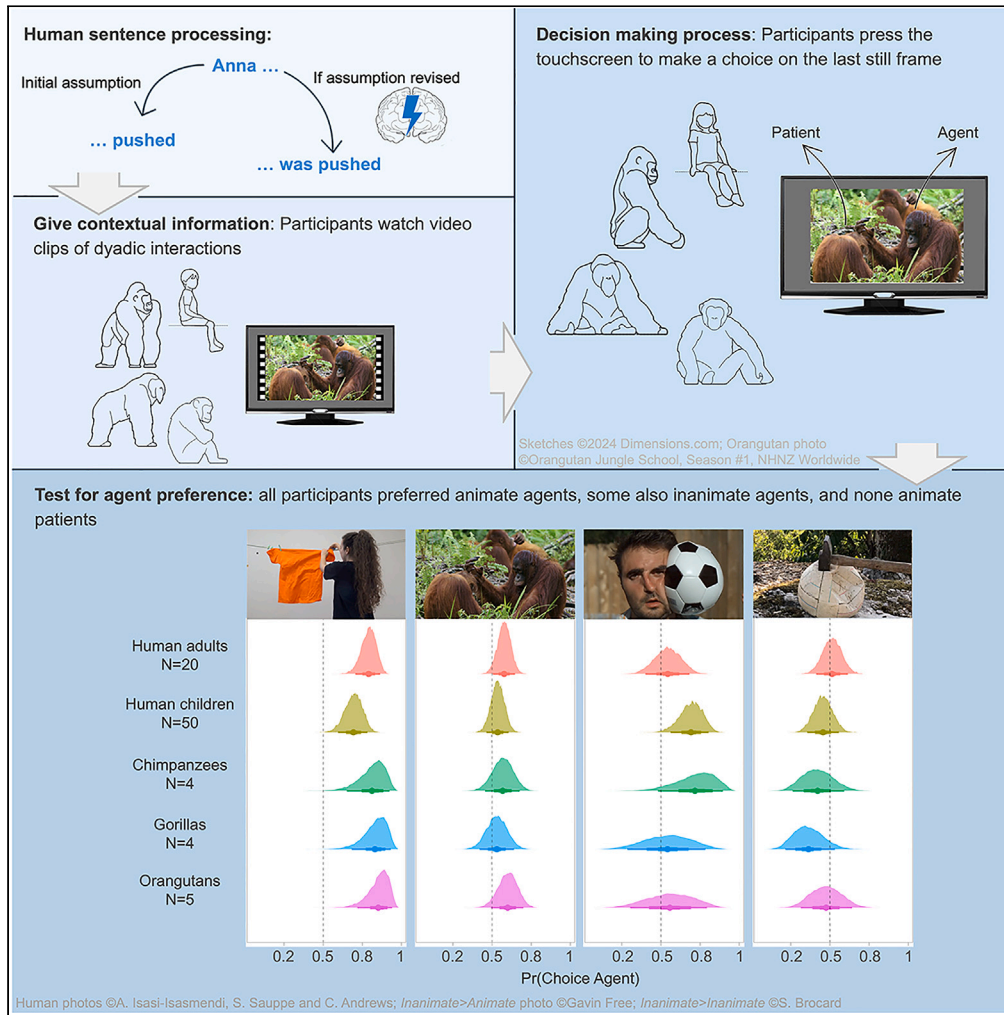
Originally published at:

Brocard, Sarah; Wilson, Vanessa A D; Berton, Chloé; Zuberbühler, Klaus; Bickel, Balthasar (2024). A universal preference for animate agents in hominids. *iScience*, 27(6):109996.

DOI: <https://doi.org/10.1016/j.isci.2024.109996>

Article

# A universal preference for animate agents in hominids



Sarah Brocard,  
Vanessa A.D.  
Wilson, Chloé  
Berton, Klaus  
Zuberbühler,  
Balthasar Bickel

sarah.brocard1@gmail.com

**Highlights**

All species tested showed a robust preference for animate agents acting on objects

The preference sometimes generalized to inanimate agents but never to animate patients

The preference is specific to decision task and does not generalize to event perception

The preference parallels the agent preference guiding sentence processing in humans

Brocard et al., iScience 27, 109996  
June 21, 2024 © 2024 The Author(s). Published by Elsevier Inc.  
<https://doi.org/10.1016/j.isci.2024.109996>



## Article

## A universal preference for animate agents in hominids

Sarah Brocard,<sup>1,6,\*</sup> Vanessa A.D. Wilson,<sup>1,2,3</sup> Chloé Berton,<sup>1</sup> Klaus Zuberbühler,<sup>1,3,4,5</sup> and Balthasar Bickel<sup>2,3,5</sup>

## SUMMARY

**When conversing, humans instantaneously predict meaning from fragmentary and ambiguous mspeech, long before utterance completion. They do this by integrating priors (initial assumptions about the world) with contextual evidence to rapidly decide on the most likely meaning. One powerful prior is attentional preference for agents, which biases sentence processing but universally so only if agents are animate. Here, we investigate the evolutionary origins of this preference, by allowing chimpanzees, gorillas, orangutans, human children, and adults to freely choose between agents and patients in still images, following video clips depicting their dyadic interaction. All participants preferred animate (and occasionally inanimate) agents, although the effect was attenuated if patients were also animate. The findings suggest that a preference for animate agents evolved before language and is not reducible to simple perceptual biases. To conclude, both humans and great apes prefer animate agents in decision tasks, echoing a universal prior in human language processing.**

## INTRODUCTION

Understanding spoken language is a mostly effortless and instantaneous process, despite the fact that speech tends to be rapid and information-rich, but often also fragmented and uncertain. Listeners make on-the-fly decisions about meanings as each word is received long before, and regardless of, utterance completion.<sup>1–3</sup> These real-time decisions critically rely on integrating current linguistic (e.g., lexical) and non-linguistic (e.g., contextual) sensory information<sup>4</sup> with prior expectations about intended meanings,<sup>5,6</sup> i.e., beliefs about plausible outcomes before evidence is observed.<sup>7,8</sup>

One powerful prior in humans is the disposition to preferentially assume that any person or animate being first mentioned in a sentence will play an agent (rather than a patient) role in an upcoming interaction — in other words, that they will instigate rather than undergo an event. Specifically, sentence processing is guided across languages by a prior expectation that role-ambiguous noun phrases default to an agent interpretation. When this expectation fails upon utterance completion, it triggers electrophysiologically measurable self-correction.<sup>9–13</sup> Language models that include this agent preference outperform simpler models in predicting event-related potentials (chiefly, the N400 effect as an indicator of role prediction failure) across structurally highly diverse languages.<sup>14</sup> This prior is universally robust, particularly for animate agents. The preference can be overridden for inanimate agents (as in “the waves smashed the canoes” or “this water makes me sick”) but, as far as is currently known, only in the very few languages that place agents last in sentences.<sup>15</sup> The prior furthermore leads to simpler, “unmarked” encoding of animate agents across languages,<sup>13,16–18</sup> although this often results in increased neural activity during sentence planning due to uncertainty.<sup>19,20</sup> The agent preference in language processing is further echoed by an attentional preference for agents when looking at still images,<sup>21–25</sup> even when they are presented only as briefly as the blink of an eye, demanding a high-speed decision.<sup>26,27</sup> This suggests that event roles can even be rapidly extracted from still images when participants engage in appropriate tasks.

The origin of the agent preference in language processing remains unclear. One hypothesis that derives from the above literature is that it is a by-product of linguistic communication, mirroring a universal trend across languages toward expressing agents before patients<sup>28–33</sup> as the prime topics of conversational interest. Only very few languages place agents last in sentences, such as *Āiwoo* spoken on the Solomon Islands.<sup>15</sup> The overwhelming majority of languages exhibit a preference for placing the agent first, and this tendency extends to subjects requested to communicate about an event with gestures instead of words.<sup>29,31</sup> A possible driver of this ordering preference is that agents are more commonly maintained as the topic of interest in conversations<sup>34,35</sup> and are therefore easier to retrieve during speech planning than patients, leading to earlier production.<sup>36–38</sup> However, recent work suggests that these order preferences might be primarily driven by a generally stronger interest in human and animate referents than in inanimate referents, independent of their roles.<sup>39,40</sup>

<sup>1</sup>Department of Comparative Cognition, Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland

<sup>2</sup>Department of Comparative Language Science, University of Zurich, Zurich, Switzerland

<sup>3</sup>Center for the Interdisciplinary Study of Language Evolution (ISLE), University of Zurich, Zurich, Switzerland

<sup>4</sup>School of Psychology & Neuroscience, University of St Andrews, St Andrews, Scotland (UK)

<sup>5</sup>These authors contributed equally

<sup>6</sup>Lead contact

\*Correspondence: [sarah.brocard1@gmail.com](mailto:sarah.brocard1@gmail.com)

<https://doi.org/10.1016/j.isci.2024.109996>



An alternative hypothesis is that the agent preference is born out of more basic non-verbal core cognition,<sup>41</sup> based on the assumption that early recognition of agents is evolutionarily important for understanding and predicting events, such as social interactions between competing group members or predicting predator behavior. This hypothesis is consistent with the fact that the human brain has a tendency to assign causality to almost any interaction, independent of language,<sup>42</sup> and that even pre-verbal children assign agent and patient roles to entities of events.<sup>41</sup> The preference of agents might therefore have evolved long before the emergence of language,<sup>43</sup> possibly as a mechanism to predict the outcome of events in a more general attempt to reduce uncertainty.<sup>7</sup> However, such an agent preference might be more nuanced than initially presumed.<sup>43</sup> It may not generalize to any agent (i.e., animate or inanimate), as human preference for agent can vary according to the context, and appears fully resilient only when agents are animate.<sup>15</sup>

Here, we investigated the universal human preference for agents with a comparative study on our closest living relatives—four western chimpanzees (*Pan troglodytes verus*), four western lowland gorillas (*Gorilla gorilla gorilla*), and five Sumatran orangutans (*Pongo abelii*)—and 20 human adults and 50 children of different ages. If the agent preference was pre-linguistic, we expected it to be shared across apes; if it is a by-product of usage patterns in language, we expected it to be specific to humans, possibly subject to developmental patterns.

Assessing agent preference non-linguistically is a challenging task since it predominantly emerges during sentence processing or picture viewing. Here, we developed an ape-friendly, non-linguistic approach that recreates the decision processes when contextual information is integrated with prior expectations. First, participants received contextual information by watching short video clips of natural events to mimic the continuous flow of incoming information in language. We used natural scenes with different hominid species to promote socio-ecological validity of the stimuli. For this reason, we did not standardize the videos in favor of reflecting the diversity of events encountered in the natural world. Then, with customized touchscreen devices,<sup>44</sup> we tested the decision-making process to access any eventual preference.

We used simple, unrestricted setups that allowed similar testing conditions for all participants, and the protocol was engaging to enhance participation from both great apes and children from 2.5 to 12 years old. Prior to the test, great apes underwent a three-stage training process that involved watching video clips featuring a single unknown conspecific actor (i.e., without event roles), while human volunteers received instructions but remained naive to the goal of the experiment (more details in the [STAR methods](#)). Participants then watched videos of two entities naturally interacting in clear causal *agent*>*patient* relations (“>” stands for “acts on”). At the end of the video, the last still frame remained on screen and participants had to make a choice by touching either the agent or the patient, in the absence of any additional information and without any subsequent reward ([Figure S1](#)).

Events were divided into four conditions: (1) *Animate*>*Inanimate*: an animate agent (chimpanzee, gorilla, orangutan, human or non-primate) acting on an inanimate patient (e.g., a cow playing with a ball, an orangutan manipulating Lego blocks), (2) *Animate*>*Animate*: an animate agent acting on a conspecific in different types of interactions, i.e., agonistic (e.g., aggressions), play, affiliative (e.g., grooming), or cooperative (e.g., help), (3) *Inanimate*>*Animate*: an inanimate agent (i.e., object) acting on a human patient (e.g., a ball smashing into the face of a person), and (4) *Inanimate*>*Inanimate*: an inanimate agent (different objects, such as balls, cars, and trees) acting on another object in different types of interaction (e.g., colliding, falling on; see [STAR methods](#)).

## RESULTS

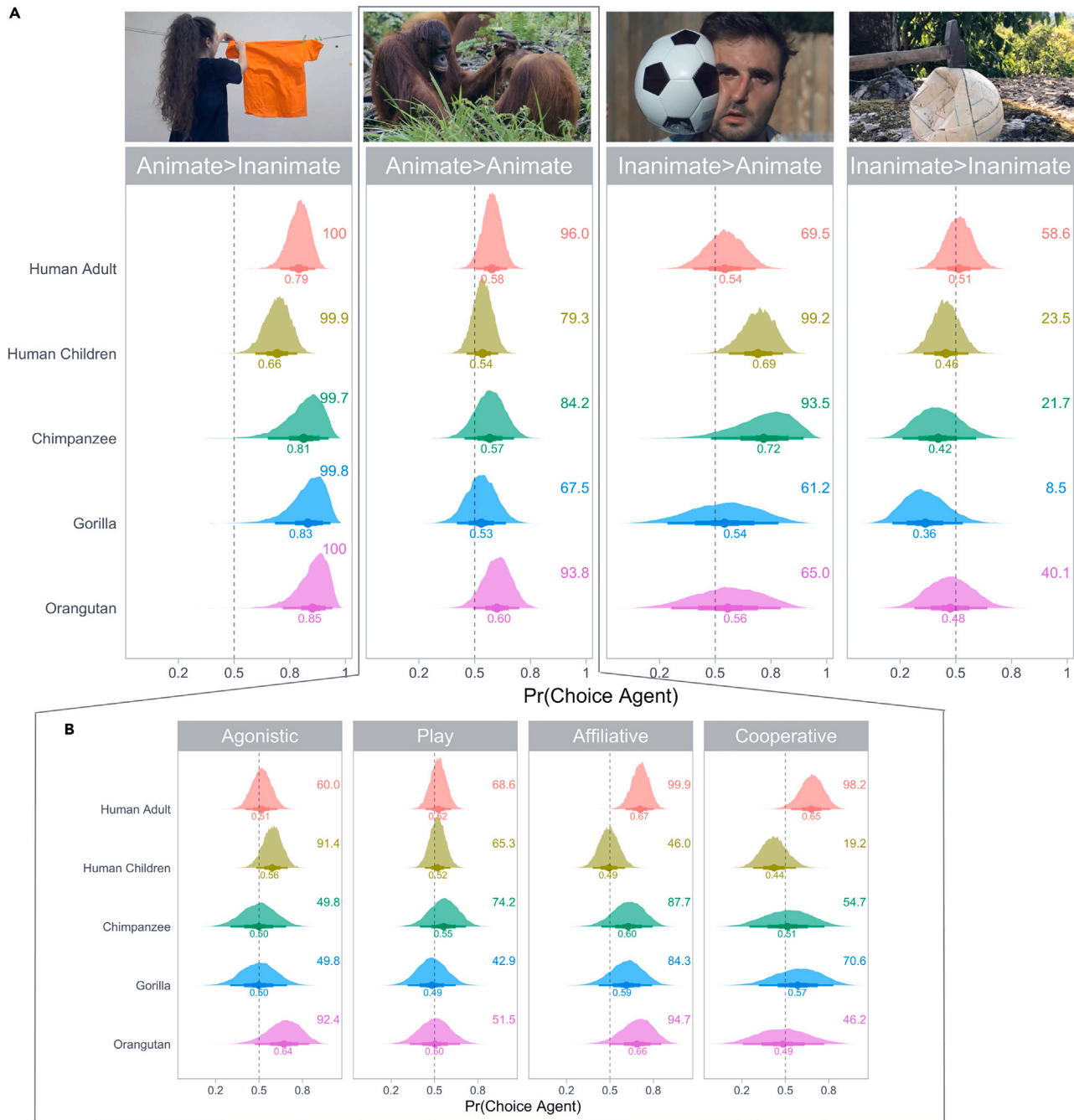
We first tested whether the event role distinction (agent vs. patient) was predictive of decisions, above and beyond relevant covariates (side of choice, conditions, species of the participant, size of the actors, and type of stimuli; see [Table S1](#)). We found that a Bayesian log-linear model with event role distinction had much better predictive performance than a model without these role distinctions, under leave-one-out cross-validation ( $\text{elpd}_{\text{Full model}} = -703.3$ ,  $\text{SE} = 12.6$  and  $\text{elpd}_{\text{Model without event roles}} = -4040.2$ ,  $\text{SE} = 290.7$ ; see [STAR methods](#) for detailed explanations).

We then estimated agent choice directly in a multi-level Bayesian logistic model, retaining only covariates that were needed to maintain predictive performance under leave-one-out cross-validation ( $\text{elpd}_{\text{Full model}} = -4399.5$ ,  $\text{SE} = 32.2$  and  $\text{elpd}_{\text{Simplified model}} = -4399.4$ ,  $\text{SE} = 31.5$ ): condition (e.g., *Animate*>*Animate*, *Animate*>*Inanimate*), species of the participant, side of choice, side of agent, status of event (i.e., action still on-going or completed), self-propelledness of agent (i.e., agent moves without propulsion or agency), centrality of agent (i.e., closeness of agent to the center of the screen relative to the patient), and differences in movement between agent and patient (i.e., difference in total amount of time agent and patient moved during the event) ([Figure S2](#); see [Table S1](#) for details on the covariates).

Our model revealed a strong effect of the condition (see [Tables S2–S4](#) for posterior estimates), after controlling for the retained covariates. We declared a preference in agent choice as “robust” if at least 90% of the posterior estimates were above chance, i.e., a 0.5 probability of agent choice.<sup>45</sup>

In the *Animate*>*Inanimate* condition, we found that all species had a robust preference for the agent (all  $\text{Pr}(\text{choice}>0.5) > 0.99$ ; [Figure 1A](#), first panel). In humans, we found an age effect insofar as agent preference increased with age, most steeply for young adults ([Figure S3](#); [Table S5](#)). For the apes, estimates were uncertain because of the small number of subjects.

In the *Animate*>*Animate* condition, arguably much more complex, the agent preference was robust only in adult humans and orangutans ([Figure 1A](#), second panel,  $\text{Pr}(\text{choice}>0.5) > 0.93$ ) with no age effect in humans ([Figure S3](#); [Table S5](#)). Closer inspection of the specific content suggested that the effects were slightly different in human adults and orangutans ([Figure 1B](#)). While both showed a preference in affiliative settings ( $\text{Pr}(\text{choice}>0.5) > 0.94$ ), humans additionally showed it in cooperative settings ( $\text{Pr}(\text{choice}>0.5) > 0.98$ ) while orangutans showed it in agonistic settings ( $\text{Pr}(\text{choice}>0.5) > 0.92$ ). Apart from these, we found an agent preference only in human children in the agonistic setting ( $\text{Pr}(\text{choice}>0.5) > 0.91$ ), similar to the orangutans. In all other settings, the *Animate*>*Animate* condition elicited a balanced preference to agents and patients.



**Figure 1. Posterior probability of the agent choice for each species across**

(A) Across conditions and (B) across type of interaction. The dashed line corresponds to a random choice (0.5). The points represent the mean of the estimates and the three levels of thickness of the point intervals represent the 30%, 60%, and 90% credible intervals. Values presented next to each distribution are the percentages of posterior estimates above a 0.5 probability of agent choice. Twenty human adults, 50 children, four chimpanzees, four gorillas, and five orangutans were tested for each condition, but one adult was removed from the *Animate>Animate* condition (“>” stands for “acting on”), as she was identified as an outlier based on her age. Twenty-five clips were presented in the *Animate>Inanimate* condition, 74 in the *Animate>Animate* condition (19 clips depicted an agonistic interaction, 27, 20, and 7 clips represented play, affiliative, and cooperative interactions, respectively), 15 in the *Inanimate>Animate* condition, and 41 in the *Inanimate>Inanimate* condition. See also Figure S3 and Tables S2–S5. Photo credits: (*Animate>Inanimate*) A. Isasi-Isasmendi, S. Sauppe, and C. Andrews; (*Animate>Animate*) Orangutan Jungle School, Season #1, NHNZ Worldwide; (*Inanimate>Animate*) GAVIN FREE and (*Inanimate>Inanimate*) S. Brocard.

Since the first two conditions did not allow to disentangle the effect of animacy from agency and the agent preference was universally robust only in the *Animate>Inanimate* condition, we furthermore tested whether this effect might be driven by animacy differences rather than agency. To do so, we tested another set of videos, this time depicting events involving inanimate agents. In the *Inanimate>Animate* condition, human children and chimpanzees continued to prefer (inanimate) agents over animate patients ( $\text{Pr}(\text{choice}>0.5) > 0.93$ ), while for the other species the agent preference weakened considerably compared to the *Animate>Inanimate* condition (Figure 1A, third panel; Figure S3; Table S5). Critically, no participant favored the animate patient, suggesting that the agent preference in the *Animate>Inanimate* condition was driven by a preference for agents and not by a preference for animate entities independent of role.

For inanimate agents acting on inanimate patients (*Inanimate>Inanimate* condition), finally, all species lost their agent preference and, if anything, shifted toward a patient preference (Figure 1A, fourth panel). This reversal was most pronounced in gorillas (posterior mean of patient choice = 0.64,  $\text{Pr}(\text{patient choice}>0.5) = 0.91$ ).

## DISCUSSION

We replicated with video stimuli the effect of a powerful prior preference<sup>9–13,21–25</sup> that in humans is known to guide sentence processing and still picture apprehension: a preference for interpreting an incoming stimulus as representing an agent. In language, this preference has been shown to be universally robust for animate agents, while it can disappear for inanimate agents when the statistical distributions of a language disfavor the agent, notably when they follow a patient-verb-agent word order.<sup>15</sup>

Analogous to this, we found robust preferences for animate agents in all four tested species of hominids. The preference was reduced in the *Animate>Animate* condition, i.e., when the animate agents acted on animate instead of inanimate patients. One possible explanation for this is that video stimuli were perceived as much more complex when they involved two animate entities. If one is animate and the other inanimate, agency is much more straightforward to estimate since animates are more common and more natural agents than inanimates.<sup>46</sup> Moreover, in events with two animates, there are many additional aspects of interest to viewers, such as age/sex class (male vs. female, adult vs. juvenile), kinship relations, or facial expressions, and these might confound the choice between agent and patient. There is also little doubt that both human and non-human primates are very sensitive to the social factors in this, particularly along the competitive-cooperative divide.<sup>47,48</sup> Consistent with this, languages tend to evolve extra case marking when patients are animate,<sup>49</sup> reflecting the additional complexity of the situation,<sup>17</sup> although it is debated how universal this trend is.<sup>13</sup>

Another possible explanation of the results lies in the nature of the events themselves. Across species, the agent preference was reduced in events that were intrinsically more symmetrical and reversible, such as reciprocal events in play (e.g., “they chased each other”) or affiliative behavior (e.g., grooming in turns). When participants arrived at the freeze frame, they might already be well aware of the agent/patient roles, but were more interested in the patient, expecting reversal or reciprocity in these types of events. The agent preference was also reduced in agonistic events, arguably because there the patient was of heightened interest (e.g., “he attacked her”) and participants tracked the social consequences (will she fight back? suffer?). Remarkably, however, the agent preference did persist in agonistic contexts in children and orangutans, and it uniquely persisted in humans for cooperative events. This variation calls for targeted follow-up experiments on the role of empathy and cooperation in event cognition across species.

Surprisingly, and contrary to human adults, the agent preference also persisted in the *Inanimate>Animate* condition for human children and chimpanzees. It is notable that chimpanzees, as a species, are much more manipulative and tool-obsessed than other great apes (especially for extractive-foraging and complex tools<sup>50</sup>). In the wild, chimpanzees use tools to impact on the animate world (e.g., crushing of parasites with leaves,<sup>51</sup> hunting with tools,<sup>52,53</sup> and attacking leopards with sticks<sup>54,55</sup>), which may explain their heightened interest for how objects impact on animate patients. Arguably, the same is true for children, who are in an ontogenetic phase where they still need to learn much about how to navigate in a world of artifacts supplied by their culture. Overall, these findings in the *Inanimate>Animate* condition support the idea that great apes, including humans, do not simply respond to animacy but appear to process something about the causes and consequences relating to the participants of interactions.

Finally, what explains the lack of preference in the *Inanimate>Inanimate* events across species? Of all conditions, these event types were probably perceived as the most unnatural. Moreover, in most cases, the inanimate agent was most likely set in motion by a hidden animate agent (e.g., a ball thrown by someone), thus participants had incomplete information and might have struggled to reconcile the perceptual features of the (agent) object with fundamental features of an agent, such as self-propelledness and goal-directedness.<sup>18,56</sup> Alternatively, they might have imagined the impact of the invisible agent, with further complicated response patterns.

What was responsible for the universal preference for animate agents? One possibility is that it was chiefly driven by the *Animate>Inanimate* condition, the only condition in which the preference was fully robust for all participants. This condition might reveal a general template that was occasionally extended to events in the *Animate>Animate* condition, likening the animate patient to an inanimate entity because of its reduced activity. Animates indeed were perceptually more salient than inanimates. For example, changes affecting animate elements in complex natural scenes were detected faster and more accurately by human subjects than changes affecting inanimate elements.<sup>57</sup> In line with this is the fact that, in the *Inanimate>Inanimate* condition, the agent preference disappeared in all species (Figure 1A, fourth panel). However, if animacy were the sole factor guiding choice, subjects should have shown a clear preference for the patient in the *Inanimate>Animate* condition, which was not the case. In addition, this account would predict that the agent preference in the *Animate>Animate* condition would be reduced most strongly in agonistic events, where the patient is least active. Yet the largest reduction was observed in play events, where both activities were more balanced.

Another possibility is that the preference for agents (albeit moderated by various factors) was due to a phylogenetically old but general bias to perceive events in terms of causal forces,<sup>42</sup> preferring for example, self-propelled entities in abstract visual displays (as shown, e.g., for newly hatched chickens<sup>56</sup>). Under this account, one would expect the preference to extend to all inanimate agents, but this is not what we found. Moreover, agent preference seems specific to decision-making, while sensitivity to causal forces holds for any perceptual processing. For example, in an ongoing companion study using eye-tracking, we found that great apes and humans, watching how events unfold over time, alternated their gaze between the agents and patients, suggesting they were trying to understand the causal nature of the interaction.<sup>58</sup> In the absence of a request for decision, this gaze alternation did not show a general bias for agents.<sup>58</sup> Similarly, the diversity of agents encountered across events makes it unlikely that participants relied on a set of complex cues (such as face recognition, facial expression, or position) to make their choices. In addition, such cues would arguably separate animate and inanimate entities more than agents and patients; yet we found no evidence that choices were reducible to preferences for animates or inanimates.

Importantly, our experiment was based on a free-choice paradigm with no specific training or instructions that could have biased preferences toward the agents or patients of events, in any of the species we tested. The fact that participants, across species, still showed consistent degrees of agent preference indicates the presence of a shared, evolutionarily old cognitive predisposition that predates language.

### Limitations of the study

A frequent argument in studies such as this one is that, in the absence of detailed instruction, human participants try to guess the aim of the task, which may affect the nature of their choices. Additionally, apes may carry over biases from previous experiments.<sup>59</sup> Our ape participants however were fairly new to touchscreen studies and certainly never had experienced anything akin to this study. If anything, any bias would have obstructed the emergence of the observed patterns, rather than produced them.

### Conclusions

Our experiments reveal a universal preference for animate agents across the tested hominids. The preference seems specific to the kind of decision-making that our task required and that also underlies the agent preference in human sentence processing and still picture apprehension. This suggests that the agent preference is a specific mechanism combining agency with animacy. It is conceivable that this mechanism has been selected for in highly encephalized species, due to its enhancement effect on rapid decision-making in event cognition, allowing individuals to quickly recognize and predict the potential impact that an animate entity can have on the perceiver, e.g., as a predator or as a competitor. Our data are consistent with the interpretation that this cognitive prior has evolved considerably earlier than modern humans. Whether or not the same or similar preferences for animate agents are present in other groups of animals is a topic for further research.

### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCES AVAILABILITY
  - Lead contact
  - Material availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Ethic statement
  - Participants and study site
- METHODS DETAILS
  - General procedure
  - Stimuli
  - Training of the apes
  - Testing procedure for the great apes
- QUANTIFICATION AND STATISTICAL ANALYSIS

### SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2024.109996>.

### ACKNOWLEDGMENTS

For support and assistance at Basel Zoo, we thank Adrian Baumeier, Fabia Wyss, Stephan Lopez, Gaby Rindlisbacher, Rene Buob, Roland Kleger, Flurin Baer, Markus Beutle, Nicole Fischer, Dominic Hohler, Reto Lehmann, Jonas Schaub, Amanda Spillmann, Patrick Wyser, Corinne Zollinger, Stephan Argast, and the technician team. For providing footage from apes and humans, we thank Cat Hobaiter, Adrian Soldati, Judith Curran and the Orangutan Jungle School, Season #1, NHNZ Worldwide, Sebastian Sauppe, Caroline Andrews, and Arrate Isasi-Isasmendi. We are also grateful to the NCCR@LiRi team, and especially Erik Ringen, for their advice. This study was supported by fundings

from the Swiss National Science Foundation grants number 310030\_185324 (K.Z.) and 100015\_182845 (B.B.), as well as by the National Center for Competence in Research “Evolving Language”, SNSF agreement number 51NF40\_180888 (B.B. and K.Z.).

## AUTHOR CONTRIBUTIONS

Conceptualization, S.B., V.A.D.W., B.B., and K.Z.; investigation, S.B. and C.B.; methodology, S.B., V.A.D.W., B.B., and K.Z.; data curation, S.B.; resources, S.B., C.B., B.B., and K.Z.; formal analysis, S.B., V.A.D.W., B.B., and K.Z.; software, S.B.; visualization, S.B., B.B., and K.Z.; writing – original draft, S.B.; writing – review and editing, S.B., V.A.D.W., C.B., B.B., and K.Z.; supervision, B.B. and K.Z.; project administration, S.B.; funding acquisition, B.B. and K.Z.;

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 6, 2023

Revised: March 15, 2024

Accepted: May 14, 2024

Published: May 16, 2024

## REFERENCES

- Altmann, G.T.M., and Mirković, J. (2009). Incrementality and Prediction in Human Sentence Processing. *Cogn. Sci.* 33, 583–609. <https://doi.org/10.1111/j.1551-6709.2009.01022.x>.
- Marslen-Wilson, W. (1973). Linguistic Structure and Speech Shadowing at Very Short Latencies. *Nature* 244, 522–523. <https://doi.org/10.1038/244522a0>.
- Christiansen, M.H., and Chater, N. (2016). The Now-or-Never bottleneck: A fundamental constraint on language. *Behav. Brain Sci.* 39, e62. <https://doi.org/10.1017/S0140525X1500031X>.
- Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M., and Sedivy, J.C. (1995). Integration of Visual and Linguistic Information in Spoken Language Comprehension. *Science* 268, 1632–1634. <https://doi.org/10.1126/science.7777863>.
- Bornkessel-Schlesewsky, I., and Schlesewsky, M. (2019). Toward a Neurobiologically Plausible Model of Language-Related, Negative Event-Related Potentials. *Front. Psychol.* 10, 298. <https://doi.org/10.3389/fpsyg.2019.00298>.
- Su, Y., MacGregor, L.J., Olasagasti, I., and Giraud, A.-L. (2023). A deep hierarchy of predictions enables online meaning extraction in a computational model of human speech comprehension. *PLoS Biol.* 21, e3002046. <https://doi.org/10.1371/journal.pbio.3002046>.
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. <https://doi.org/10.1038/nrn2787>.
- Knill, D.C., and Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci.* 27, 712–719. <https://doi.org/10.1016/j.tins.2004.10.007>.
- Haupt, F.S., Schlesewsky, M., Roehm, D., Friederici, A.D., and Bornkessel-Schlesewsky, I. (2008). The status of subject-object reanalyses in the language comprehension architecture. *J. Mem. Lang.* 59, 54–96. <https://doi.org/10.1016/j.jml.2008.02.003>.
- Demiral, Ş.B., Schlesewsky, M., and Bornkessel-Schlesewsky, I. (2008). On the universality of language comprehension strategies: Evidence from Turkish. *Cognition* 106, 484–500. <https://doi.org/10.1016/j.cognition.2007.01.008>.
- Wang, L., Schlesewsky, M., Bickel, B., and Bornkessel-Schlesewsky, I. (2009). Exploring the nature of the ‘subject’-preference: Evidence from the online comprehension of simple sentences in Mandarin Chinese. *Lang. Cogn. Process.* 24, 1180–1226. <https://doi.org/10.1080/01690960802159937>.
- Erdocia, K., Laka, I., Mestres-Missé, A., and Rodriguez-Fornells, A. (2009). Syntactic complexity and ambiguity resolution in a free word order language: Behavioral and electrophysiological evidences from Basque. *Brain Lang.* 109, 1–17. <https://doi.org/10.1016/j.bandl.2008.12.003>.
- Bickel, B., Witzlack-Makarevich, A., Choudhary, K.K., Schlesewsky, M., and Bornkessel-Schlesewsky, I. (2015). The Neurophysiology of Language Processing Shapes the Evolution of Grammar: Evidence from Case Marking. *PLoS One* 10, e0132819. <https://doi.org/10.1371/journal.pone.0132819>.
- Huber, E., Sauppe, S., Isasi-Isasmendi, A., Bornkessel-Schlesewsky, I., Merlo, P., and Bickel, B. (2024). Surprisal from language models can predict ERPs in processing predicate-argument structures only if enriched by an Agent Preference principle. *Neurobiol. Lang.* 5, 167–200. [https://doi.org/10.1162/nol\\_a\\_00121](https://doi.org/10.1162/nol_a_00121).
- Sauppe, S., Naess, A., Roversi, G., Meyer, M., Bornkessel-Schlesewsky, I., and Bickel, B. (2023). An Agent-First Preference in a Patient-First Language During Sentence Comprehension. *Cogn. Sci.* 47, e13340. <https://doi.org/10.1111/cogs.13340>.
- Fillmore, C.J. (1968). The Case for Case. In *Universals in Linguistic Theory*, E. Bach and R. Harms, eds. (Holt, Rinehart, and Winston).
- Silverstein, M. (1976). Hierarchy of features and ergativity. In *Grammatical categories in Australian languages* (Australian Institute of Aboriginal Studies), pp. 112–171. <https://doi.org/10.5281/ZENODO.4688088>.
- Rissanen, L., and Majid, A. (2019). Thematic roles: Core knowledge or linguistic construct? *Psychon. Bull. Rev.* 26, 1850–1869. <https://doi.org/10.3758/s13423-019-01634-5>.
- Sauppe, S., Choudhary, K.K., Giroud, N., Blasi, D.E., Norcliffe, E., Bhattamishra, S., et al. (2021). Neural signatures of syntactic variation in speech planning. *PLoS Biol.* 19, e3001038. <https://doi.org/10.1371/journal.pbio.3001038>.
- Egurtzegi, A., Blasi, D.E., Bornkessel-Schlesewsky, I., Laka, I., Meyer, M., Bickel, B., and Sauppe, S. (2022). Cross-linguistic differences in case marking shape neural power dynamics and gaze behavior during sentence planning. *Brain and Language* 230, 105127. <https://doi.org/10.1016/j.bandl.2022.105127>.
- Dobel, C., Gummior, H., Bölte, J., and Zwitserlood, P. (2007). Describing scenes hardly seen. *Acta Psychol.* 125, 129–143. <https://doi.org/10.1016/j.actpsy.2006.07.004>.
- Sauppe, S. (2016). Verbal Semantics Drives Early Anticipatory Eye Movements during the Comprehension of Verb-Initial Sentences. *Front. Psychol.* 7, 95. <https://doi.org/10.3389/fpsyg.2016.00095>.
- Ünal, E., Richards, C., Trueswell, J.C., and Papafragou, A. (2021). Representing agents, patients, goals and instruments in causative events: A cross-linguistic investigation of early language and cognition. *Dev. Sci.* 24, e13116. <https://doi.org/10.1111/desc.13116>.
- Wilson, F., Papafragou, A., Bunger, A., and Trueswell, J. (2011). Rapid Extraction of Event Participants in Caused Motion Events. In *Proc. 33rd Annu. Conf. Cogn. Sci. Soc.*
- Cohn, N., and Paczynski, M. (2013). Prediction, events, and the advantage of Agents: The processing of semantic roles in visual narrative. *Cognit. Psychol.* 67, 73–97. <https://doi.org/10.1016/j.cogpsych.2013.07.002>.
- Hafri, A., Trueswell, J.C., and 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>.
- Isasi-Isasmendi, A., Andrews, C., Flecken, M., Laka, I., Daum, M.M., Meyer, M., Bickel, B., and Sauppe, S. (2023). The Agent Preference in Visual Event Apprehension. *Open Mind* 7, 240–282. [https://doi.org/10.1162/opmi\\_a\\_00083](https://doi.org/10.1162/opmi_a_00083).
- Greenberg, J.H. (1963). *Some Universals of Grammar with Particular Reference to the*



- Order of Meaningful Elements. In *Universals of Language* (MIT Press), pp. 73–113.
29. Goldin-Meadow, S., So, W.C., Ozyürek, A., and Mylander, C. (2008). The natural order of events: How speakers of different languages represent events nonverbally. *Proc. Natl. Acad. Sci. USA* 105, 9163–9168. <https://doi.org/10.1073/pnas.0710060105>.
  30. Schouwstra, M., and De Swart, H. (2014). The semantic origins of word order. *Cognition* 131, 431–436. <https://doi.org/10.1016/j.cognition.2014.03.004>.
  31. Futrell, R., Hickey, T., Lee, A., Lim, E., Luchkina, E., and Gibson, E. (2015). Cross-linguistic gestures reflect typological universals: A subject-initial, verb-final bias in speakers of diverse languages. *Cognition* 136, 215–221. <https://doi.org/10.1016/j.cognition.2014.11.022>.
  32. Dryer, M.S. (2013). Order of Subject, Object and Verb. In *World Atlas Lang. Struct. Online* (Zenodo). <https://doi.org/10.5281/zenodo.7385533>.
  33. Napoli, D.J., and Sutton-Spence, R. (2014). Order of the major constituents in sign languages: implications for all language. *Front. Psychol.* 5, 376. <https://doi.org/10.3389/fpsyg.2014.00376>.
  34. Givón, T. (1983). *Topic Continuity in Discourse: A Quantitative Cross-Language Study* (John Benjamins Publishing Company), p. 492. <https://doi.org/10.1075/tsl.3>.
  35. Haig, G., and Schnell, S. (2016). The discourse basis of ergativity revisited. *Language* 92, 591–618. <https://doi.org/10.1353/lan.2016.0049>.
  36. Ariel, M. (2014). *Accessing Noun-Phrase Antecedents* (RLE. Linguistics B: Grammar). Routledge 280. <https://doi.org/10.4324/9781315857473>.
  37. MacDonald, M.C. (2013). How language production shapes language form and comprehension. *Front. Psychol.* 4, 226. <https://doi.org/10.3389/fpsyg.2013.00226>.
  38. Christianson, K., and Ferreira, F. (2005). Conceptual accessibility and sentence production in a free word order language (Odawa). *Cognition* 98, 105–135. <https://doi.org/10.1016/j.cognition.2004.10.006>.
  39. Meir, I., Aronoff, M., Börstell, C., Hwang, S.-O., Ilkbasaran, D., Kastner, I., Lepic, R., Lifshitz Ben-Basat, A., Padden, C., and 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>.
  40. Nordlinger, R., Rodriguez, G.G., and Kidd, E. (2022). Sentence planning and production in Murrinhpatha, an Australian “free word order” language. *Language* 98, 187–220. <https://doi.org/10.1353/lan.2022.0008>.
  41. Strickland, B. (2017). Language Reflects “Core” Cognition: A New Theory About the Origin of Cross-Linguistic Regularities. *Cogn. Sci.* 41, 70–101. <https://doi.org/10.1111/cogs.12332>.
  42. Michotte, A. (1963). *Perception of Causality Basic Books* (Basic Books).
  43. Wilson, V.A.D., Zuberbühler, K., and Bickel, B. (2022). The evolutionary origins of syntax: Event cognition in nonhuman primates. *Sci. Adv.* 8, eabn8464. <https://doi.org/10.1126/sciadv.abn8464>.
  44. Egelkamp, C.L., and Ross, S.R. (2019). A review of zoo-based cognitive research using touchscreen interfaces. *Zoo Biol.* 38, 220–235. <https://doi.org/10.1002/zoo.21458>.
  45. McElreath, R. (2020). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan Second Edition* (CRC Press).
  46. Wolff, P., Jeon, G.-H., and Li, Y. (2009). Causers in English, Korean, and Chinese and the individuation of events. *Lang. Cogn.* 1, 167–196. <https://doi.org/10.1515/LANGCOG.2009.009>.
  47. De Vevey, M., Bouchard, A., Soldati, A., and Zuberbühler, K. (2022). Thermal imaging reveals audience-dependent effects during cooperation and competition in wild chimpanzees. *Sci. Rep.* 12, 2972. <https://doi.org/10.1038/s41598-022-07003-y>.
  48. McFarland, R., Roebuck, H., Yan, Y., Majolo, B., Li, W., and Guo, K. (2013). Social Interactions through the Eyes of Macaques and Humans. *PLoS One* 8, e56437. <https://doi.org/10.1371/journal.pone.0056437>.
  49. Jäger, G. (2007). Evolutionary Game Theory and Typology: A Case Study. *Language* 83, 74–109. <https://doi.org/10.1353/lan.2007.0020>.
  50. Meulman, E.J.M., and Van Schaik, C.P. (2013). Orangutan tool use and the evolution of technology. In *Tool Use in Animals*, C.M. Sanz, J. Call, and C. Boesch, eds. (Cambridge University Press), pp. 176–202. <https://doi.org/10.1017/CBO9780511894800.012>.
  51. Assersohn, C., Whiten, A., Kiwede, Z.T., Tinka, J., and Karamagi, J. (2004). Use of leaves to inspect ectoparasites in wild chimpanzees: a third cultural variant? *Primates* 45, 255–258. <https://doi.org/10.1007/s10329-004-0092-x>.
  52. Humle, T., and Matsuzawa, T. (2002). Ant-dipping among the chimpanzees of Bossou, Guinea, and some comparisons with other sites. *Am. J. Primatol.* 58, 133–148. <https://doi.org/10.1002/ajp.10055>.
  53. McGrew, W.C., and Collins, D.A. (1985). Tool use by wild chimpanzees (*Pan troglodytes*) to obtain termites (*Macrotermes* herus) in the Mahale Mountains, Tanzania. *Am. J. Primatol.* 9, 47–62. <https://doi.org/10.1002/ajp.1350090106>.
  54. Boesch, C., and Boesch, H. (1989). Hunting behavior of wild chimpanzees in the Tai National Park. *Am. J. Phys. Anthropol.* 78, 547–573. <https://doi.org/10.1002/ajpa.1330780410>.
  55. Van Lawick-Goodall, J. (1968). The Behaviour of Free-living Chimpanzees in the Gombe Stream Reserve. *Anim. Behav. Monogr.* 1, 161–IN12. [https://doi.org/10.1016/S0066-1856\(68\)80003-2](https://doi.org/10.1016/S0066-1856(68)80003-2).
  56. Mascalzoni, E., Regolin, L., and Vallortigara, G. (2010). Innate sensitivity for self-propelled causal agency in newly hatched chicks. *Proc. Natl. Acad. Sci. USA* 107, 4483–4485. <https://doi.org/10.1073/pnas.0908792107>.
  57. New, J., Cosmides, L., and Tooby, J. (2007). Category-specific attention for animals reflects ancestral priorities, not expertise. *Proc. Natl. Acad. Sci. USA* 104, 16598–16603. <https://doi.org/10.1073/pnas.0703913104>.
  58. Wilson, V.A.D., Sauppe, S., Brocard, S., Ringen, E., Daum, M.M., Wermelinger, S., Gu, N., Andrews, C., Isasi-Isasmendi, A., Bickel, B., and Zuberbühler, K. (2023). Primate origins of human event cognition. Preprint at bioRxiv. <https://doi.org/10.1101/2023.11.23.568086>.
  59. Déaux, E.C., Baumeyer, A., and Zuberbühler, K. (2021). Food calls enhance visual discrimination learning in chimpanzees (*Pan troglodytes verus*). *J. Comp. Psychol.* 135, 420–429. <https://doi.org/10.1037/com0000286>.
  60. The MathWorks Inc (2017). MATLAB version: 9.2 (R2017a) (Natick, Massachusetts: The MathWorks Inc). <https://www.mathworks.com>.
  61. R Core Team (2021). R: A language and environment for statistical computing (Vienna, Austria: R Foundation for Statistical Computing). <https://www.R-project.org/>.
  62. Carpenter, B., Gelman, A., Hoffman, M.D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M.A., Guo, J., Li, P., and Riddell, A. (2017). Stan: A Probabilistic Programming Language. *J. Stat. Softw.* 76, 1. <https://doi.org/10.18637/jss.v076.i01>.
  63. Brainard, D.H. (1997). The Psychophysics Toolbox. *Spat. Vis.* 10, 433–436. <https://doi.org/10.1163/156856897X00357>.
  64. Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* 10, 437–442. <https://doi.org/10.1163/156856897X00366>.
  65. Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *J. Stat. Softw.* 80, 1–28. <https://doi.org/10.18637/jss.v080.i01>.
  66. Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *R J.* 10, 395. <https://doi.org/10.32614/RJ-2018-017>.
  67. Bürkner, P.-C. (2021). Bayesian Item Response Modeling in R with brms and Stan. *J. Stat. Softw.* 100, 1–54. <https://doi.org/10.18637/jss.v100.i05>.
  68. Piironen, J., and Vehtari, A. (2017). Comparison of Bayesian predictive methods for model selection. *Stat. Comput.* 27, 711–735. <https://doi.org/10.1007/s11222-016-9649-y>.
  69. Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* 27, 1413–1432. <https://doi.org/10.1007/s11222-016-9696-4>.
  70. Yao, Y., Vehtari, A., Simpson, D., and Gelman, A. (2018). Using Stacking to Average Bayesian Predictive Distributions (with Discussion). *Bayesian Anal.* 13, 917–1007. <https://doi.org/10.1214/17-BA1091>.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Deposited data</b>		
Dataset	This paper	<a href="https://osf.io/5vrc7/?view_only=3519ecd7e366403d9f8e9e52cb17eb71">https://osf.io/5vrc7/?view_only=3519ecd7e366403d9f8e9e52cb17eb71</a>
Data analyses	This paper	<a href="https://osf.io/5vrc7/?view_only=3519ecd7e366403d9f8e9e52cb17eb71">https://osf.io/5vrc7/?view_only=3519ecd7e366403d9f8e9e52cb17eb71</a>
<b>Software and algorithms</b>		
MATLAB (version R2017a)	The MathWorks Inc., 2017 <sup>60</sup>	<a href="https://www.mathworks.com">https://www.mathworks.com</a>
R (version 4.0.5)	R Core Team, 2021 <sup>61</sup>	<a href="https://www.r-project.org">https://www.r-project.org</a>
Stan	Carpenter et al., 2017 <sup>62</sup>	<a href="https://doi.org/10.18637/jss.v076.i01">https://doi.org/10.18637/jss.v076.i01</a>
<b>Other</b>		
Video stimuli	This paper	<a href="https://osf.io/5vrc7/?view_only=3519ecd7e366403d9f8e9e52cb17eb71">https://osf.io/5vrc7/?view_only=3519ecd7e366403d9f8e9e52cb17eb71</a>
Formal definition of the statistical models	This paper	<a href="https://osf.io/5vrc7/?view_only=3519ecd7e366403d9f8e9e52cb17eb71">https://osf.io/5vrc7/?view_only=3519ecd7e366403d9f8e9e52cb17eb71</a>

### RESOURCES AVAILABILITY

#### Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact: Sarah Brocard ([sarah.brocard1@gmail.com](mailto:sarah.brocard1@gmail.com)).

#### Material availability

This study did not generate reagents.

#### Data and code availability

- Data have been deposited at [https://osf.io/5vrc7/?view\\_only=3519ecd7e366403d9f8e9e52cb17eb71](https://osf.io/5vrc7/?view_only=3519ecd7e366403d9f8e9e52cb17eb71) and are publicly available as of the date of publication.
- All original code, analyses and stimuli have been deposited at [https://osf.io/5vrc7/?view\\_only=3519ecd7e366403d9f8e9e52cb17eb71](https://osf.io/5vrc7/?view_only=3519ecd7e366403d9f8e9e52cb17eb71) and is publicly available as of the date of publication. All custom MATLAB scripts will be available upon request.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

### EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

#### Ethic statement

This study received approval from the ethics committee of the University of Neuchâtel (project 66-2020-B), the Basel cantonal veterinary office (permit 3077), and the Animal Welfare Officer at Basel Zoo.

#### Participants and study site

##### Humans

We tested two age groups: (i) adults (*Animate*>*Animate* condition:  $N = 19$ ; 11 females; mean  $29.95 \pm 12.04$  years old (an extra participant was removed as she was considered as an age outlier); range [19; 56]; other conditions:  $N = 20$ ; 13 females; mean  $24.55 \pm 7.88$  years old; range [18; 56]), and (ii) children from 2 to 12 years old (*Animate*>*Animate* condition:  $N = 50$ ; 28 females; mean  $5.64 \pm 2.03$  years old; range [2; 11]; other conditions:  $N = 50$ ; 23 females; mean  $5.84 \pm 2.26$  years old; range [2; 12]).

The children and some of the adults were recruited from visitors of the primate house at Basel Zoo (Switzerland), through the zoo's social media advertisements, posters in the zoo or direct approach. Other adult participants were recruited from amongst students at the University of Neuchâtel via email. All participants were Caucasian.

Prior to the experiment, all volunteer participants were given basic information about the procedure and policies, without knowing the goal of the experiment. They were instructed to watch the short videos and choose one actor by touching the screen when the image paused. Participants were also informed that there was no right or wrong answer. By the end of the experiment, all participants or their legal guardians received a detailed information sheet (background, procedure, confidentiality policy) and had to sign an informed consent form and respond to a short demographic questionnaire. Children also received a participation "certificate".

### Great apes

Participants were chimpanzees, gorillas and orangutans housed at Basel Zoo. During the study period, the zoo housed up to 15 chimpanzees, eight gorillas and nine orangutans. Chimpanzees and gorillas were always maintained as respective cohesive social groups and orangutans were kept as adult pairs with their immature offspring. Four chimpanzees (two females; mean  $7.25 \pm 7.18$  years old; range [3; 18]), four gorillas (three females; mean  $13.75 \pm 12.26$  years old; range [5; 31]) and five orangutans (three females; mean  $15.40 \pm 5.08$  years old; range [8; 20]) voluntarily participated in the study (Table S6). All species had access to indoor and outdoor enclosures (total holding area for chimpanzees:  $767.43 \text{ m}^2$ ,  $4,957.03 \text{ m}^3$ ; gorillas:  $753.22 \text{ m}^2$ ,  $4,173.89 \text{ m}^3$ ; orangutans:  $813.81 \text{ m}^2$ ,  $1,0691.79 \text{ m}^3$ ). All enclosures were equipped with ropes, hammocks and climbing structures; material to build nests was also provided every day. The roof of all enclosures was made of sliding glass windows assuring natural lightning throughout the day. All species were fed a mix of fruit and vegetables supplemented with small amounts of protein, with several "meals" distributed throughout the day. The study did not pose any risk and participation was fully voluntary. Subjects were never food or water deprived at any point, but received food rewards for participating, in accordance with their diet.

## METHODS DETAILS

### General procedure

Training of the great apes and testing data were collected every weekday from 8:00 to 11:00 for the great apes and from 9:00 to 16:00 for human participants. We used touchscreen devices (Iiyama ProLite T1931SR, 19", 48 cm and  $1280 \times 1024$  resolution, using resistive technology) connected to a laptop (Dell Latitude 5580). The experiments were programmed and displayed using MATLAB software (The MathWorks Inc. (2022), version R2017a) and specifically the Psychophysics Toolbox Version 3 (PTB-3) extensions.<sup>63,64</sup> The setup for great apes was protected by Plexiglas boxes; chimpanzee and gorilla setups were fixed in the enclosure, the one used with the orangutans was installed on a trolley allowing the experimenter to move between the enclosures (Figure S4).

A daily testing session started when one individual approached and pressed the screen, then, participants saw approximately ten testing trials. As we did not want to induce any bias in the participants' choices, there was no sound feedback or food reward directly after participants made a choice. Thus, a single test trial could broadly be divided into three parts: participants (i) watched the clip, (ii) chose one actor on the last frame, and (iii) pressed on an image (a human hand holding a fruit for the great ape participants, and a trophy cup for the humans) to get a positive sound and a food reward, for the great apes only (Figure S1). If they pressed outside the image, a negative sound was emitted, and they had to try again until they succeeded. Once a participant left the testing area, the experimenter ended the test. If the participant returned later, the session resumed where it had stopped.

### Stimuli

Stimuli were short, soundless, colored natural video clips (see other [supplementary materials](#)), presented either in their original or mirrored version, to counterbalance the side of the agent across trials.

In the *Animate>Animate* condition, the two actors belonged to the same species but could have different perceptual features. The nature of the interaction (agonistic, play, affiliative or cooperative) varied as well, but the different event categories could be recognized by all participants, as affiliation and aggression share similar characteristics across species.<sup>48</sup> The clips were sorted into five blocks based on the species of the actors: chimpanzees, gorillas, humans, non-primates (e.g., dogs, elephants, birds ...) or orangutans.

All participants started with clips of conspecifics, as we expected it would be easier for them to detect goals and thus to understand the event. The remaining four blocks were randomised. Each block consisted of 15 events, except for the human block that only had 14 clips as one might have upset younger children (chimpanzee mean duration of the video clips ( $\pm$ SD):  $6.4\text{s} \pm 2.5$ ; gorilla:  $4.7\text{s} \pm 2.1$ ; human:  $4.3\text{s} \pm 0.9$ ; non-primates:  $6.0\text{s} \pm 1.8$  and orangutan:  $5.2\text{s} \pm 1.6$ ).

Before moving to the other conditions, great apes went through a ten trials session of *Inanimate>Inanimate* events, to ensure that they were able to generalise the task to inanimate actors. As the accuracy rate was approximately 85% (mean number of touches necessary to achieve a valid choice:  $1.9 \pm 2.9$ ; median = 1) we decided that additional training was unnecessary.

We presented 38 *Inanimate>Inanimate* events ( $4.0\text{s} \pm 1.7$ ) depicting different objects, such as balls, cars, trees, in different types of interaction (e.g., colliding, falling on). In the *Inanimate>Animate* condition we presented 15 clips ( $3.7\text{s} \pm 1.1$ ) of various objects acting on a human patient, such as a ball smashing into the face of a person. Finally, in the *Animate>Inanimate* condition we presented 25 clips ( $6.6\text{s} \pm 1.9$ ), where the animate actors, from the same species as in the *Animate>Animate* condition, manipulated various objects (e.g., a cow playing with a ball, an orangutan manipulating with Lego blocks).

In addition to the clips presenting interactions, we also showed control video clips that differed only insofar as there was no interaction between the two actors (*Animate/Animate*:  $N = 25$ , mean ( $\pm$ SD):  $4.3\text{s} \pm 1.4$ ; *Animate/Inanimate*:  $N = 7$ , mean ( $\pm$ SD):  $3.7\text{s} \pm 1.4$ ; *Inanimate/Inanimate*:  $N = 5$ , mean ( $\pm$ SD):  $3.0\text{s} \pm 0.7$ ), and allowed us to test the influence of the interaction on the choice.

The same great apes participated in all four conditions, whereas human adults and children participants were split, half of them were exclusively tested in the *Animate>Animate* condition, while the remaining half were tested in the three other conditions, to maximise our chance of recruiting participants from the zoo visitors. Great apes always started with the *Animate>Animate* condition to maximise their interest for the study and all species, except children, saw *Inanimate>Inanimate* events, followed by *Inanimate>Animate* events and finished by the *Animate>Inanimate* condition. These blocks were not randomised because we wanted to avoid any bias for the animate actor. Children saw the different conditions in a random order. Great ape and adult participants were tested on all the clips, while, for constraints of time and attention span (as participants were recruited amongst the zoo visitors), the children were tested with subsets (Table S7).

### Training of the apes

Only great ape participants received training before the experiment. The training consisted in watching short, soundless, and colored natural video clips, of a single unknown conspecific actor in his environment either at rest or moving. For each great ape species, we used a set of 92 clips (chimpanzee: mean ( $\pm$ SD) duration 4.9s  $\pm$ 0.6; gorilla: 3.8s  $\pm$ 1.4; and orangutan: 4.1s  $\pm$ 1.3). The clips were short to maximise the attention and interest of the subjects.

Our training was designed as a three-stage process. Each stage began with the participants watching a video clip, followed by a still image of the last frame remaining on screen. The content of the still image varied according to the training phase. In the first phase, the silhouette of the actor appeared on a white background. In the second, the background was blurred, while it remained unchanged in the third phase (Figure S5). The task difficulty also gradually increased throughout the training. In the first phase, participants were allowed to touch the screen as many times as needed to reach the target (the area of the actor), but the trial was considered as failed if more than one touch was necessary. To pass this first training phase, participants had to obtain an 80% success rate (or more) for two consecutive sessions of ten trials a day. The second training phase was quite similar, except the criteria of success was three consecutive sessions at 80% success rate. Finally, in the last phase, participants only had one chance to correctly touch the actor, if not the trial was a failure, and they moved on to the next one. The criterion of success was the same as for the second training phase. During the entire training, we used positive reinforcement based on sound feedback and food reward.

### Testing procedure for the great apes

A daily testing session, for the great apes, always started when one individual approached and pressed a green screen (followed by a food reward). Then, five warm-up trials (random images) were presented to help the participants to focus on work, followed by three training trials (phase three). Then, participants saw approximately ten testing trials before finishing with five cool-off trials (random images) and a red screen with three pieces of grapes (Figure S1). Human participants did not go through the warm-up and training trials and directly started with the test trials. They saw all clips in a row, with a 5-min break halfway through the session if needed.

## QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses were conducted in R (R Core Team, 2021, version 4.0.5) using Bayesian generalised linear models through the brms package<sup>65–67</sup> interface to Stan.<sup>62</sup> We chose a Bayesian framework because it is less dependent on sample size and allows richer assessment of uncertainty in parameter estimates.

Log-linear models were used to rule out that a choice was conditionally independent of the interaction between the actors, i.e., exchangeable across control and test videos. To compute this, we created a contingency table of summed choices and fitted Bayesian regression with a Poisson distribution (normally distributed priors:  $\mu = 0$ ,  $\sigma = 1.5$ ). In one model the type of test (control or test) was included and in the second one it was not (see key resources table for formal model definitions). We compared the expected log pointwise predictive density (elpd) under leave-one-out cross validation to select the best model. Higher elpd indicates higher out-of-sample predictive fit, thus a better model.

Bayesian Bernoulli models, with a logit link, were fitted (i) to model the agent choice according to the species of the participant and across conditions but also their interactions and (ii) model the agent choice regarding the species of the participant and the type of interaction in the *Animate>Animate* condition, as well as the interaction between the two. Normally distributed priors ( $\mu = 0$ ,  $\sigma = 1.5$ ) were used for the intercept and all population-level predictors (which correspond to a relatively flat distribution on the logit scale for each predictor), while exponentially distributed priors ( $\lambda = 1$ ) were used for the standard deviation of group-level effects. The outcome variable was the choice made and the predictors of interest were the species of the participant and the condition or the content of the event, as well as their interactions.

To control for potential confounding factors, both models included as fixed predictors the different features of the clip (e.g., the side of the choice, the side of the agent, the relative size difference between agent and patient, see Table S1 for details and coding of these features). These same predictors were also included as random slopes by participant and clip (stimulus). Formal model definitions are presented in the Supplementary Text.

Given the large number of potential confounds we performed variable selection using leave-one-out cross validation (*cv\_varef* function of the *projpred* package<sup>68</sup> with optimised parallelization) to trim the model down to only those covariates that are needed to retain the same out-of-sample predictive performance. From the 14 predictors only 8 were kept and used for all models (see Figure S2).

All models were checked for convergence based on trace plots, R-hat values and ESS diagnostics. We further controlled for the fit of the model using the *pp\_check* function, which compared the observed data to the data simulated from the model's posterior predictive distribution (see Figure S6). We also conducted sensitivity checks by fitting the models with different priors (Normal(0,1.5); Normal(0,0.5) and

Horseshoe(1)) and we did not get any noticeable differences, suggesting that the model learned parameters from the data and not the prior despite the relatively small sample size. Finally, we controlled for overly influential data points with Pareto  $k$  estimates.<sup>69,70</sup> All Pareto  $k$  were within the commonly accepted range ( $k < 0.5$ ), suggesting that there were no overly influential data points and that our  $elpd$  estimates were reliable.

For each parameter of interest, we reported its posterior estimate mean and standard error, as well as the proportion of the posterior estimates above a 0.5 choice (main text) and the 90% credible intervals, which we considered as the meaningful threshold (Table S2). We visualised this information with density plots of the posterior distribution. In these figures, in addition to the 90% CI we also reported the 30% and 60% CIs.