

MIT Press • Open Encyclopedia of Cognitive Science

# The Interaction Engine

Stephen C. Levinson<sup>1</sup>

<sup>1</sup>Max Planck Institute for Psycholinguistics

MIT Press

**Published on:** Aug 19, 2025

**DOI:** <https://doi.org/10.21428/e2759450.e3df24b2>

**License:** [Creative Commons Attribution-NonCommercial 4.0 International License \(CC-BY-NC 4.0\)](#).

The *interaction engine* is a label for the specialized capacities involved in human social interaction. The interaction engine comprises a suite of specific abilities that enable communication, for example, the capacity for rapid conversational turn-taking using multimodal signals, with systematic contingency between initiating signal and response—in which that contingency is often inference based rather than conventional in character. The properties of informal interaction subsumed within the broader “interaction engine” label appear to be largely universal, in contrast to the diversity of languages and many other aspects of social behavior. They seem to provide a platform that makes it possible for children to acquire their particular language and culture.

## History

The term *interaction engine* was proposed to draw out the theoretical consequences for the cognitive sciences of many observations made about human interaction by social scientist pioneers such as Ray Birdwhistell, Jerome Bruner, Erving Goffman, Adam Kendon, Harvey Sacks, and Emmanuel Schegloff ([Levinson, 2006](#)). These scholars drew attention to the orderly regularity and complexity of the organization of informal talk. Subsequent work showed that many aspects of this organization are invariant across cultures and are acquired very early in human development, suggesting the hypothesis that an interactional foundation makes possible the diversity of human languages ([Levinson, 2025](#)). The cognitive demands of rapid language use in an interactional context has proved challenging to current models of language processing ([Levinson & Torreira, 2015](#)) [see [Psycholinguistics](#)]. Recent work has also explored overlaps between human social interaction and that of other primates, making a case for some strong continuities in the communicative behavior of humans and other apes ([Heesen & Fröhlich, 2022](#)).

## Core concepts

Human informal interaction (as in a friendly chat) is premised on the capacity for *joint attention* and *joint action* ([Tomasello, 2022](#)), the kind of coordination involved when we play tennis or a duet together, but it also has many more specific properties ([Levinson, 2025](#)) [see [Social Learning](#) and [Shared Intentionality](#)]. Prominently, it has characteristic temporal qualities, with a rapid exchange of short *conversational turns*, each on average about 2 seconds long, with gaps between them of only around 200 ms. ([Levinson & Torreira, 2015](#)). These turns are typically contingent on a prior turn (as in an answer to a question) involving a mapping of an utterance (verbal and nonverbal signals) onto actions like requesting, promising, proposing, and so on [see [Conversation](#)]. This mapping is inferential and often quite indirect rather than strictly rule governed, although guided by the form of the utterance. For example, a question form like “Why are you doing that?” can be an information query, but it can also be a challenge or a request to desist ([Drew & Couper-Kuhlen, 2014](#); [Levinson, 2013](#)). The contingencies can be stacked, producing embedded sequences ([Levinson, 2013](#); [Schegloff, 2007](#)), as in a conversation between two individuals, A and B:

A: Did you go?

B: Where?

A: To Harry's party.

B: You mean last weekend?

A: Yeah.

B: I went but didn't stay.

In face-to-face conversation, turns are typically multimodal, involving vocalization, gesture, gaze, facial expressions, and body posture. The orchestration of multiple articulators into a coherent message poses cognitive challenges for both production and comprehension ([Holler & Levinson, 2019](#)). Uniquely, humans are also able to shift the whole burden of communication from the vocal organs to the hands, as in sign languages [see [Sign Language](#)].

These and related properties of human interaction are, in contrast to language itself, relatively robust to cultural variation. First, turn-taking timing seems to be consistently in the 200- to 400-ms range across cultures ([Stivers et al., 2009](#)). Second, the same structure of embedded contingencies (as in the example above) can be found everywhere ([Kendrick et al., 2020](#)). Third, the mechanisms for repairing utterances not heard or understood seem to be universal in organization ([Dingemanse et al., 2015](#)). Further, all cultures make use of the full multimodal range of signals, with expressive gestures and facial expressions. Even the close huddle with frequent mutual gaze is characteristic of human informal interaction everywhere.

### ***Cognitive implications***

Responses in conversation are at the speed of the minimum human response time (c. 200 ms). In order to respond at this speed, brain imaging has shown that interlocutors begin processing their response midway during comprehension of the incoming turn ([Bögels et al., 2015](#)). This planning is cognitively demanding, especially as language production and comprehension use much of the same apparatus [see [Language Production](#)]. Studies show that although infants respond quickly in nonlinguistic ways to caretaker input ([Casillas & Frank, 2017](#)), it takes children until middle childhood before they can respond linguistically at adult-like speed ([Stivers et al., 2018](#)). The processing of multimodal signals must also be cognitively demanding, since the parts that go together are not necessarily synchronous.

The embedding of contingencies, as illustrated above, has been claimed by generative linguists to be a property that we owe to linguistic structure, but in fact, center embedding occurs stacked much deeper (six or more embeddings within embeddings) in interaction structure than in linguistic structure. It was thought that the practical limits in linguistic structure are due to memory limitations, but interactional data shows this cannot be the case ([Levinson, 2013](#)).

## ***Innate or learned abilities?***

As with so many aspects of human behavior, the degree to which the interaction engine is genetically endowed is uncertain. There are, however, a number of strong arguments in favor: a striking overlap with the precise timing of responses with the gestural exchange of great apes, the manifestation of many of these abilities in early infancy in “proto-conversation,” and the precise neural control not only of the tongue but of the face and the hands, which play an important role in human communication. Above all, there is a striking cross-cultural universality to the properties of the interaction engine as exhibited in informal conversation, which contrasts markedly with the diversity of human languages and cultures.

The interaction engine concept brings together numerous properties of interaction: timing, contingency, inference to intent, multimodality including gaze and gesture, repair organization, and so forth. One legitimate question is whether these actually cohere as a close-knit constellation of abilities. Here, the study of clinical syndromes is telling: the diagnostics of autism includes impairment on each of these dimensions [see [Autism](#)]. Similarly, *The Diagnostic and Statistical Manual of Mental Disorders* ([American Psychiatric Association, 2024](#)) lists *social (pragmatic) communication disorder* as a newly defined clinical syndrome that covers all the ingredients of the interaction engine while not being tied to general intelligence. From a clinical perspective, the interaction engine binds together these interactive abilities in such a way that they are likely to be collectively impaired. The current opinion in genetics is that autism and related syndromes are linked firmly to genetic predispositions. However, there is also evidence that interactional abilities develop through practice interacting with others in the environment ([Rubio-Fernandez, 2024](#)).

## ***Evolutionary implications***

The interaction engine offers an explanatory bridge between other primate communication systems and human language and communication. Many primate species are vocal turn-takers, but the great apes show the closest similarity to human interaction patterns in a gestural mode of communication ([Call & Tomasello, 2007](#)) [see [Primate Communication](#)]. There is a close match between great ape gestural timing and the timing of human turn-taking ([Fröhlich et al., 2016](#)) and a similar contingency between signals mapped to actions ([Rossano, 2013](#); [Rossano & Liebal, 2014](#)); these ape gestural signals seem also to be at least partially learned and flexibly used. Given that all the great ape species are gestural communicators in intimate interaction, it seems likely the common ancestor with humans also was ([Levinson, 2023](#)). That offers an account of the pancultural association of gesture and speech in humans, in line with the theory that language had its origin in gesture (the *gestural origin* hypothesis).

The evolutionary origins of language are contested and remain uncertain. However, paleontological analysis suggests a gradual switch of the burden of communication from gesture to speech around a million years ago, with multiple adaptations of the vocal and auditory systems already in place by the common ancestor of modern humans and Neanderthals around 600,000 to 700,000 years ago ([Dediu & Levinson, 2013](#)). The relatively recent nature of human language is also suggested by the fact that most of

the structure and content of languages are not innately specified but outsourced to culture, making us the only species with a radically diversified communication system. The interaction engine arguably makes this possible by providing an interactional system infants can use to learn their natal language.

## Questions, controversies, and new developments

The interaction engine concept offers a new way of approaching old questions concerning the nature and origin of human communication and language in particular. It also draws attention to the relatively neglected study within the cognitive sciences of the study of human social interaction, with recent advances showing, for example, how processes of language comprehension and production must work in parallel. In particular, it draws attention to the importance of cross-cultural and cross-species comparison. Much further work can therefore be done in, for example, the study of child development of communication across cultures, in the cross-cultural study of interactive process, and in the hyper scanning (simultaneous neuroimaging) of the brain processes of interacting individuals.

The interaction engine also raises controversial issues by suggesting, for example, that language might have grown out of a slowly accumulated set of interactional abilities rather than the reverse idea, that language evolved suddenly by a set of chance mutations, thereby facilitating human interaction. Similarly, one can ask whether children are bootstrapped into their native language through the interaction engine or whether it is language-specific learning mechanisms that facilitate the accumulation of communicative competence. Is the interaction engine really a coherent suite of abilities, or is it rather an auxiliary concept that will dissolve into multiple subprocesses on closer examination? The association of most of the traits in autism do suggest an organic coherence, which future genetic studies may clarify.

## Broader connections

The concept offers broad connections as mentioned above to such topics as primate communication systems, the evolution of language, language diversity, and conversational structure. The idea of an interactional foundation underlying the diversity of languages and cultures also draws attention to the central role of child development in bootstrapping individuals into their cultural milieu [see [Language Acquisition](#); [Cognitive Development](#)]. The concept also suggests that the cognitive sciences could benefit from wider connections to the social sciences and anthropology.

## Acknowledgments

The research on which this article is based was funded by the Max Planck Society and the European Research Council. I am indebted to generations of students and colleagues, too numerous to list, who did so much of the research work.

## Further reading

- Heesen, R., & Fröhlich, M. (2022). Revisiting the human “interaction engine”: Comparative approaches to social action coordination. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, 377(1859), 20210092. <https://doi.org/10.1098/rstb.2021.0092>
- Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., de Ruiter, J. P., Yoon, K. E., & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences of the United States of America*, 106 (26), 10587-10592. <https://doi.org/10.1073/pnas.0903616106>
- Levinson, S. C. (2019). Interactional foundations of language: The interaction engine hypothesis. In P. Hagoort (Ed.), *Human language: From genes and brain to behavior* (pp. 189-200). MIT Press. <https://doi.org/10.7551/mitpress/10841.003.0018>

## References

- American Psychiatric Association. (2024). *The diagnostic and statistical manual of mental illnesses* (5th ed.). American Psychiatric Association.

↩

- Bögels, S., Magyari, L., & Levinson, S. C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. *Scientific Reports*, 5, 12881. <https://doi.org/10.1038/srep12881>

↩

- Call, J., & Tomasello, M. (Eds). (2007). *The gestural communication of apes and monkeys*. Psychology Press.

↩

- Casillas, M., & Frank, M. C. (2017). The development of children’s ability to track and predict turn structure in conversation. *Journal of Memory and Language*, 92, 234–253. <https://doi.org/10.1016/j.jml.2016.06.013>

↩

- Dediu, D., & Levinson, S. C. (2013). On the antiquity of language: The reinterpretation of Neandertal linguistic capacities and its consequences. *Frontiers in Language Sciences*, 4, 397. <https://doi.org/10.3389/fpsyg.2013.00397>

↩

- Dingemanse, M., Roberts, S. G., Baranova, J., Blythe, J., Drew, P., Floyd, S., Gisladdottir, R. S., Kendrick, K. H., Levinson, S. C., Manrique, E., Rossi, G., & Enfield, N. J. (2015). Universal principles in the repair of communication problems. *PLoS One*, 10(9), e0136100. <https://doi.org/10.1371/journal.pone.0136100>

↩

- Drew, P., & Couper-Kuhlen, E. (Eds.) (2014). *Requesting in social interaction*. John Benjamins.



- Fröhlich, M., Kuchenbuch, P., Müller, G., Fruth, B., Furuichi, T., Wittig, R. M., & Pika, S. (2016). Unpeeling the layers of language: Bonobos and chimpanzees engage in cooperative turn-taking sequences. *Scientific Reports*, 6, 25887. <https://doi.org/10.1038/srep25887>



- Heesen, R., & Fröhlich, M. (2022). Revisiting the human “interaction engine”: Comparative approaches to social action coordination. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, 377(1859), 20210092. <https://doi.org/10.1098/rstb.2021.0092>



- Holler, J., & Levinson, S. C. (2019). Multimodal language processing in human communication. *Trends in Cognitive Sciences*, 23(8), 639-652. <https://doi.org/10.1016/j.tics.2019.05.006>



- Kendrick, K. H., Brown, P., Dingemanse, M., Floyd, S., Gipper, S., Hayano, K., Hoey, E., Hoymann, G., Manrique, E., Rossi, G., & Levinson, S. C. (2020). Sequence organization: A universal infrastructure for social action. *Journal of Pragmatics*, 168, 119-138. <https://doi.org/10.1016/j.pragma.2020.06.009>



- Levinson, S. C. (2006). On the human “interaction engine.” In N. J. Enfield & S. C. Levinson (Eds.), *Roots of human sociality: Culture, cognition and interaction* (pp. 39-69). Berg. <https://doi.org/10.4324/9781003135517-3>



- Levinson, S. C. (2013). Action formation and ascription. In T. Stivers & J. Sidnell (Eds.), *The handbook of conversation analysis* (pp. 103-130). Wiley-Blackwell. <https://doi.org/10.1002/9781118325001.ch6>



- Levinson, S. C. (2013). Recursion in pragmatics. *Language*, 89(1), 149-162. <https://doi.org/10.1353/lan.2013.0005>



- Levinson, S. C. (2023). Gesture, spatial cognition and the evolution of language. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, 378(1875), 20210481. <https://doi.org/10.1098/rstb.2021.0481>



- Levinson, S. C. (2025). *The interaction engine: language in social life and human evolution*. Cambridge University Press.



- Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, 6, 731. <https://doi.org/10.3389/fpsyg.2015.00731>



- Rossano, F. (2013). Sequence organization and timing of bonobo mother-infant interactions. *Interaction Studies*, 14(2), 160-189. <https://doi.org/10.1075/is.14.2.02ros>



- Rossano, F., & Liebal, K. (2014). “Requests” and “offers” in orangutans and human infants. In Drew, P. & Couper-Kuhlen, E. (Eds.), *Requesting in social interaction* (pp. 333-362). <https://doi.org/10.1075/slsi.26.13ros>



- Rubio-Fernandez, P. (2024). Cultural evolutionary pragmatics: Investigating the codevelopment and coevolution of language and social cognition. *Psychological Review* 131(1), 18-35. <https://doi.org/10.1037/rev0000423>



- Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis*. Cambridge University Press.



- Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., de Ruiter, J. P., Yoon, K. E., & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences of the United States of America*, 106(26), 10587-10592. <https://doi.org/10.1073/pnas.0903616106>



- Stivers, T., Sidnell, J., & Bergen, C. (2018). Children's responses to questions in peer interaction: A window into the ontogenesis of interactional competence. *Journal of Pragmatics*, 124, 14-30. <https://doi.org/10.1016/j.pragma.2017.11.013>



- Tomasello, M. (2022). The coordination of attention and action in great apes and humans. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, 377, 20210093. <https://doi.org/10.1098/rstb.2021.0093>

