**The limits of predictive looks: Eye tracking to assess interpretations of gaps in dialogue**

Mathias Barthel (mathias.barthel@uni-koeln.de)[1,2]

[1]Department of German Language and Literature, University of Cologne, [2]IDS Mannheim

Conversational turn-taking is characterised by a strong tendency to minimise gaps between speaking turns [5, 6, 11, 12, 15]. A key factor underlying this precise timing is the gap effect: Extended silences are often interpreted as meaningful, shaping how both the upcoming utterance [e.g., 7, 14] and the speaker [e.g., 1, 2, 16] are understood. The present study examines how overhearers' expectations about an upcoming response develop during a silent gap following a request or offer. Based on conversational preference organisation [13], listeners are likely to initially expect a preferred response (e.g., a granting following a request). However, it remains unclear how such expectations evolve as a gap extends. Expectations may either decay over time and get neutralised, or they may reverse, so that negative responses are expected when the gap becomes markedly long.

We conducted a reaction time and eye-tracking experiment with 64 native German speakers. All 240 trials began with a short written context showing the opening of a phone call, followed by an auditorily presented pair of turns in which the caller produced either a request ("Could you pick me up, then?"; 50%) or an offer ("Should I pick you up, then?"; 50%) that received a positive ("yes") or negative ("no") response after either a short (300–500 ms) or a long (1000–1200 ms) gap (2 $\times$ 2 design). Subjects indicated as fast as possible whether the response was "yes" or "no" via keypress. Trial-initially, a fixation cross was shown on the bottom of the screen. During the question, the screen was blank, with two coloured boxes (green/red) labelled "Ja" (yes) and "Nein" (no) appearing on the upper half of the screen at question offset, their positions being counterbalanced across participants. If listeners initially expect a positive response, reaction times should be faster for "yes" than for "no" after short gaps. If expectations dissipate over time, this advantage should disappear during long gaps. If expectations reverse, reactions to "no" should become even faster than reactions to "yes" after long gaps. Gaze is known to reflect ongoing cognitive processing [4, 9], and anticipatory looks can indicate predictions of upcoming linguistic input [3, 8, 10]. Hence, gaze patterns toward the yes and no areas were expected to index evolving response expectations: An initial bias toward "yes," followed by either neutralisation or reversal to a later bias toward "no" during the gap.

Reaction time data showed significantly faster responses to "Ja" than to "Nein", independent of gap length ($\beta$ = 25 ms; *SE* = 8 ms; $p < .01$), with no interaction between gap duration and response type (Fig. 1a). Contradicting our predictions, this suggests that participants maintained a preference for positive responses even after long gaps, challenging previous accounts suggesting reduced surprise for negative responses after longer gaps. However, invariant expectations should be visible in constantly high proportions of gazes to the yes-area during the gap. Yet, contrary to what can be predicted based on the eye-mind hypothesis, subjects did not look toward the yes/no-areas, but stayed focused on the region where the fixation cross was presented earlier (Fig. 1b). Thus, participants possibly did not engage in predictive processing of the question-answer pair but merely responded to the auditory stimulus of the answer, showcasing the limits of predictive dialogue processing.

In a currently running follow-up study, gaze movement in preparation of the manual response is made more relevant by shuffling the positions of the yes- and no-areas between trials. Results will be available by the time of the conference, resolving the riddle.

## References


[1] Barthel, M. (2025). In *Proceedings of the 29th Workshop on the Semantics and Pragmatics of Dialogue – Full Papers*, pages 135–145, Bielefeld, Germany.

[2] Blohm, S. and Barthel, M. (2025). In *Proceedings of the 29th Workshop on the Semantics and Pragmatics of Dialogue – Full Papers*, pages 125–134, Bielefeld, Germany. SemDial.

[3] Casillas, M. and Frank, M. C. (2017). *Journal of Memory and Language*, 92:234–253.

[4] Cooper, R. M. (1974). *Cognitive Psychology*, 6(1):84–107.

[5] Dingemanse, M. and Liesenfeld, A. (2022). In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5614–5633, Dublin, Ireland. Association for Computational Linguistics.

[6] Heldner, M. and Edlund, J. (2010). *Journal of Phonetics*, 38(4):555–568.

[7] Henetz, T. (2017). PhD Thesis, Stanford University, Stanford.

[8] Huettig, F., Rommers, J., and Meyer, A. S. (2011). *Acta Psychologica*, 137(2):151–171.

[9] Just, M. A. and Carpenter, P. A. (1980). *Psychological review*, 87(4):329–354.

[10] Kamide, Y., Altmann, G. T., and Haywood, S. L. (2003). *Journal of Memory and Language*, 49(1):133–156.

[11] Kendrick, K. H. and Torreira, F. (2014). *Discourse Processes*, 52(4):1–35.

[12] Levinson, S. C. and Torreira, F. (2015). *Frontiers in Psychology*, 6(731):10–26.

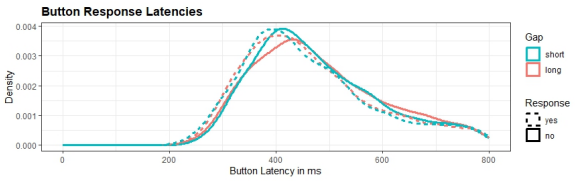
[13] Pomeranz, A. and Heritage, J. (2013). In Sidnell, J. and Stivers, T., editors, *The Handbook of Conversation Analysis*, pages 210–228. John Wiley & Sons, Ltd, Chichester, UK.

[14] Roberts, F. and Francis, A. L. (2013). *The Journal of the Acoustical Society of America*, 133(6):EL471–EL477.

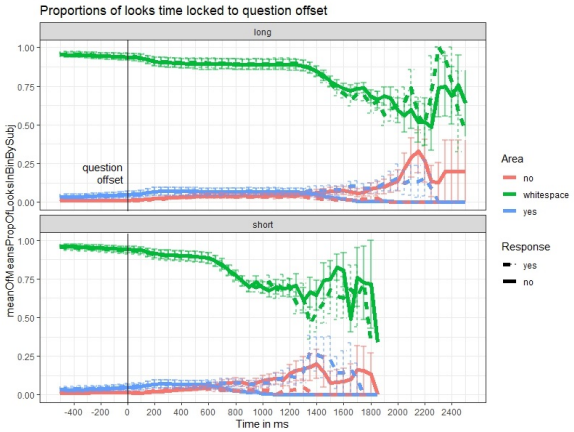
[15] Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., de Ruiter, J. P., Yoon, K.-E., and Levinson, S. C. (2009). *Proceedings of the National Academy of Sciences*, 106(26):10587–10592.

[16] Templeton, E. M., Chang, L. J., Reynolds, E. A., Cone LeBeaumont, M. D., and Wheatley, T. (2022). *Proceedings of the National Academy of Sciences*, 119(4):e2116915119.


## Figures and Tables




(a) Manual reaction times by gap length and answer (yes vs. no). Button presses were significantly faster for yes-answers than for no-answers ($\beta$ = 25 ms), irrespective of gap length.




(b) Proportions of looks towards areas of interest (yes/no) and whitespace in short-gap trials (top panel) and long-gap trials (bottom panel).