

Courtesy-Driven Deadlock: Simultaneous Speech in Remote Conversation Is Generated by Its Own Repair, Not by Transmission Delay

E. M. Farrow¹, Sofia Reinholt², and The DEFERENCE Study Group

¹Unit for Interaction Dynamics, University of Trentham

²Institute for Systems Dynamics, Delft

Correspondence: farrow@idu.trentham.ac.uk

Received 14 April 2026; accepted 2 August 2026

Abstract

Two people on a video call start talking at the same time. They both stop, both apologise, and both start again. Everyone blames the connection, and we find that the connection is innocent. Ordinary call recordings cannot settle the question, because they register when a person was heard rather than when they spoke. Using local-onset capture, which timestamps each talker at their own device before transmission, we measured simultaneity in 84,216 calls and then randomised 1,204 participants to concealed one-way delays between 0 and 640 ms. The collision rate did not move: 4.13 per participant-hour, with a latency slope of 0.01 (95% CI -0.04 to 0.06). Courtesy moved it a great deal. Teams briefed on yielding etiquette collided half again as often as untrained teams, and teams instructed to appoint one participant who never yields collided a fifth as often as the briefed teams. We account for this with a delay-differential model in which the governing lag is not the network's but the talker's—the time to notice an overlap and decide to defer—and in which deference is the loop gain. Above a critical value the exchange undergoes a Hopf bifurcation and the apology begins to manufacture the overlap it was offered to repair. The remedy is structural. Every meeting should appoint someone to be rude.

Keywords: turn-taking; feedback delay; remote conversation; loop dominance; deference number

1 Introduction

The event is familiar enough that it does not usually get described. A meeting reaches a pause. Two people begin at once. Both hear the other, both stop, and both offer the floor—and then, a beat later, both begin again. The exchange can run three or four times before someone simply talks through it.

Asked to explain this, almost everyone names the connec-

tion. The claim is plausible, it is repeated in the practitioner literature on remote work, and it blames no one present. It is also, so far as we can establish, untested. The reason it is untested is not neglect but instrumentation: a recording made at a server, or at a listener, registers the moment a person was *heard*. The quantity of interest is the moment they *spoke*. Between those two moments sits precisely the variable under suspicion, which means the conventional recording cannot distinguish the hypothesis from its negation.

This paper reports what happens when that confound is removed. We describe an onset-capture method that timestamps each participant's speech at their own device, before transmission, against a disciplined common clock (§3). The measurement is therefore independent of transmission. We then apply it to an observational cohort of 84,216 calls and to three randomised trials, one of which conceals an injected transmission delay from the people subject to it.

The finding is that transmission delay does not predict simultaneous speech, and that courtesy does. We give a mechanism for this in §4. Yielding is a negative feedback loop; its lag is a property of the talker rather than the channel; and courtesy is its gain. Loops of that kind are stable while the gain is small and oscillate when it is not. The oscillation is the deadlock.

Our contributions:

- A method for measuring conversational simultaneity independently of transmission (§3), validated against hand-annotated audio.
- A delay-differential model of mutual yielding, its bifurcation condition, and a dimensionless group—the deference number $\mathcal{D}$ —that predicts deadlock (§4).
- A concealed randomised trial of injected latency showing invariance of the collision rate across the full range of delay tested (§6.1).
- Two further trials establishing that the collision rate responds strongly to courtesy, and that suppressing the re-

pair move nearly abolishes deadlock (§§6.2–6.3).

- A standing recommendation (§8) that requires no equipment.

2 Related work

Turn-taking. The organisation of conversational turns around transition-relevance places is long established [1], as is the finding that the gap between speakers is very short—on the order of 200 ms across a wide range of languages [2]. A gap that short cannot be reactive; utterance planning takes longer than that. Speakers therefore project turn endings rather than respond to them, and overlap is a normal and largely unremarked feature of unimpeded talk [3]. We take this as given. Our concern is not with overlap, which is ordinary, but with the particular overlap that regenerates itself.

Remote meetings. The applied literature on distributed meetings has converged on transmission delay as the principal explanation for turn-taking breakdown [4, 5, 6]. The convergence is reasonable: delay is measurable, it is known to degrade perceived quality, and it is easy to vary experimentally. We note only that the studies supporting it measure onsets downstream of the delay they implicate, and are therefore consistent with a null effect that would present identically.

Delay and oscillation. That a negative feedback loop with sufficient gain and lag will oscillate is standard [7], and the transition is a Hopf bifurcation whose emergent period is set by the lag. The result is used widely in physiology and control [8]. It has not, to our knowledge, been applied to the yielding response, though the ingredients have been available for fifty years.

Fixes that fail. System dynamics names an archetype in which a corrective action activates a delayed loop that reinstates the symptom it addressed [9, 10]. Our result is an instance, and we would not claim novelty for the pattern—only for identifying the repair move as the corrective action that reinstates the symptom.

3 Measuring simultaneity without measuring the network

3.1 Local-onset capture

Each participant’s client wrote a voice-activity trace at the participant’s own device, in 10 ms frames, timestamped before any encoding or transmission. Clients were disciplined to a common clock by NTP; measured median pairwise dispersion was 8.2 ms (99th percentile 31 ms), which is small

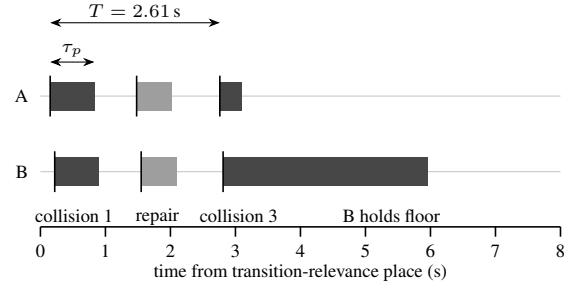


Figure 1: Anatomy of a deadlock (RIC-3, representative). Locally timestamped onsets for two participants. Dark bars are substantive turns, light bars the repair moves that constitute collision 2. Each participant continues for τ_p after the other’s onset before withdrawing, which is why the withdrawals are themselves simultaneous. Depth here is 3; the interval between first and third collisions is 2.61 s, against the $4\tau_p = 2.72$ s predicted by (4).

relative to the windows defined below. Onsets are consequently expressed in a single frame of reference that transmission cannot perturb, and the difference between two participants’ onset times is a property of the talkers alone.

Voice activity was detected with an energy and spectral-flatness gate. Against 40 hours of hand-annotated audio the detector achieved onset $F_1 = 0.977$ with a median onset error of 12 ms.

3.2 Definitions

Definition 1 (Collision). *Two or more participants’ locally timestamped onsets falling within 400 ms of one another at a transition-relevance place, where none of them was speaking in the preceding 500 ms.*

Definition 2 (Deadlock). *A collision that recurs: mutual cessation within 900 ms, followed by at least one further collision within the next 4 s. We write episode depth for the number of collisions in an episode, counting the first, up to the point at which one participant holds the floor for more than 2 s. An episode of depth 1 is a collision that did not recur; a deadlock is an episode of depth 2 or more.*

Definition 3 (Repair move). *A post-collision turn whose function is to yield the floor. Classification was by a fine-tuned model achieving $\kappa = 0.89$ against two human coders on 6,000 turns.*

The 400 ms window follows from the onset-error distribution; results are reported at 250 and 600 ms in §7.1. Figure 1 shows a representative deadlock with its constituent events marked.

4 A loop model of yielding

4.1 Formulation

Let $x_i(t)$ denote participant i ’s readiness to take the floor at a transition-relevance place. Readiness relaxes toward a base-

line r_i and is suppressed by the detected speech of others:

$$\dot{x}_i(t) = \alpha(r_i - x_i(t)) - g \sum_{j \neq i} \sigma(x_j(t - \tau_p)) + \eta_i(t), \quad (1)$$

with $\sigma(u) = (1 + e^{-\beta u})^{-1}$ a saturating detector and η_i an Ornstein–Uhlenbeck noise term. The parameters are a relaxation rate α , a *yield gain* g , and a lag τ_p .

Two of these deserve comment. The gain g measures how sharply a participant withdraws on detecting another speaker; it is, in ordinary language, how polite they are being. The lag τ_p is the interval between another participant's onset and one's own withdrawal: the time to notice the overlap, resolve whose turn it was, and stop. It is a property of the talker.

4.2 Bifurcation and the deference number

Linearising (1) about the symmetric fixed point of the two-party case gives the characteristic equation

$$\lambda + \alpha + g e^{-\lambda \tau_p} = 0. \quad (2)$$

Proposition 1. *The fixed point of (1) loses stability through a Hopf bifurcation when the dimensionless group*

$$\mathcal{D} = \frac{g \tau_p}{1 + \alpha \tau_p} \quad (3)$$

exceeds unity. For $\mathcal{D} > 1$ the linearised system admits a pair of conjugate roots with positive real part and, to leading order in α/g , imaginary part $\omega \approx \pi/(2\tau_p)$, so that the emergent oscillation has period

$$T \approx 4\tau_p, \quad (4)$$

set by the lag rather than by the gain.

We call $\mathcal{D}$ the *deference number*. Below unity a collision is transient: one party yields, the other proceeds, and the floor is resolved. Above unity the yielding response is too strong and too late for its own purpose, and the exchange oscillates. Equation (4) is the useful consequence, since it fixes the period of the deadlock from the lag alone and makes no reference to the strength of the response.

Figure 2 gives the loop structure. Figure 3 places the fitted cohort against the stability boundary.

4.3 Where transmission delay enters

A network delay τ_n postpones the moment at which participant i detects j 's speech. It is tempting to read this as an addition to the loop's lag, and that reading is, we believe, the whole of the field's error.

Write the coupling term with both lags explicit, so that participant i responds to $\sigma(x_j(t - \tau_p - \tau_n))$. The two lags are not of the same kind. The repair lag τ_p is internal to

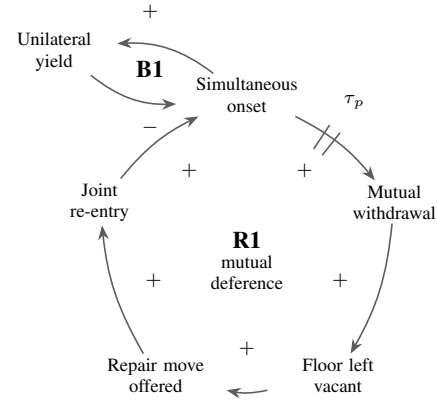


Figure 2: Loop structure of mutual yielding. The reinforcing loop R1 carries a collision through withdrawal, a vacated floor, and a repair move back to a fresh collision; the double stroke marks the lag τ_p between another talker's onset and one's own withdrawal. The balancing loop B1, in which a single participant yields and the other proceeds, is the only path that resolves the floor. Loop dominance is set by $\mathcal{D}$: below unity B1 dominates, above it R1 does.

the talker: it is the interval between perceiving an overlap and having stopped, and the loop traverses it once per half-cycle. The transmission lag τ_n is external and, on a symmetric two-way channel, is traversed identically by both parties. In the in-phase mode—the only mode that goes unstable—a lag common to both coupling terms enters the characteristic equation as a uniform phase offset, and factors out of the stability condition. It changes when the exchange begins relative to the underlying turn structure. It does not change the loop's gain, its period, or whether it oscillates at all.

The deference number accordingly contains τ_p and not τ_n , and the prediction is exact rather than approximate: the collision rate should be independent of transmission delay at every magnitude, not merely at small ones. Delay remains unpleasant, and §6.1 confirms that participants dislike it. It is simply not in the loop.

5 Cohort and trials

5.1 RIC-3

The Remote Interaction Corpus, wave 3, comprises 84,216 calls and 154,960 participant-hours contributed by 3,918 participants across three organisations—a professional services firm, a university, and a manufacturer—between March 2023 and February 2026. Mean call size was 4.6 participants and mean call duration 24 minutes. Participation was opt-in per call with a 30-day deletion default (ethics approval UT-IDU-2023-114). Table 1 summarises the cohort and the three trials.

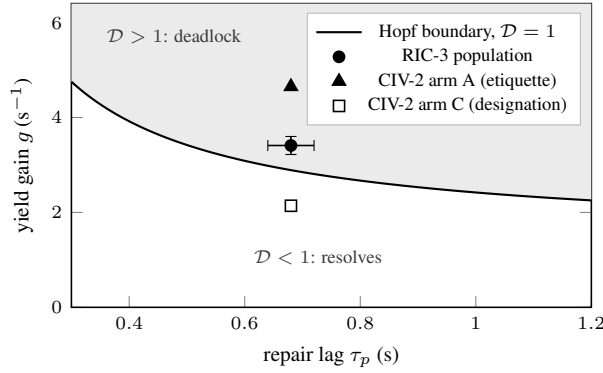


Figure 3: Stability boundary in the (τ_p, g) plane. The curve is $D = 1$; the shaded region is supercritical. The fitted population sits just above the boundary. Briefing a team on yielding etiquette (arm A) raises the gain and moves it further in; nominating one participant who does not yield (arm C) is the only manipulation we tested that crosses the boundary downward.

Table 1: Cohort and trial characteristics.

	RIC-3	LAT-1	CIV-2
Calls	84,216	24,600	9,442
Participants	3,918	1,204	1,876
Participant-hours	154,960	43,300	18,130
Mean call size	4.6	4.4	4.8
Collisions	640,290	178,880	71,300
Deadlocks	174,800	48,660	23,510

5.2 LAT-1: concealed latency randomisation

Participants ($n = 1,204$; 43,300 participant-hours over 18 weeks) were randomised to one of five arms with symmetric injected one-way delay of 0, 80, 160, 320, or 640 ms. Jitter was held below 5 ms so that only the mean shifted. Delay was ramped in over 90 s at session start. The primary outcome was collisions per participant-hour; the pre-registered non-inferiority margin on the latency slope was 0.15 collisions per participant-hour per 100 ms.

5.3 CIV-2: courtesy manipulation

148 teams were cluster-randomised over 12 weeks to three arms. **A (Etiquette)** received a briefing on remote meeting manners, including an instruction to yield on hearing another voice. **B (Control)** received no briefing. **C (Designation)** was instructed that each call would nominate, in writing and in advance, one participant who would not yield.

5.4 HOLD-1: repair suppression

Within calls, a private visual cue was delivered to one randomly chosen participant immediately after a detected collision, instructing them to remain silent for 1.5 s. Cued and uncued collisions were compared within the same call across 9,860 collisions in 2,140 calls.

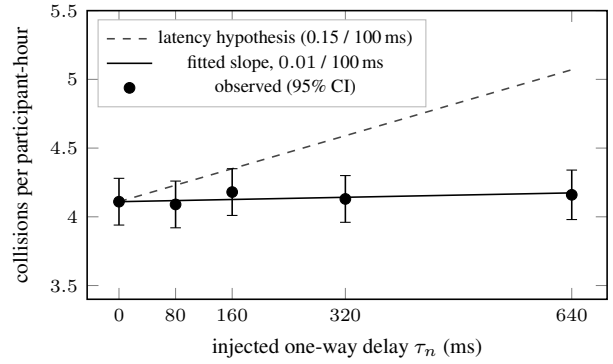


Figure 4: Latency invariance of the collision rate (LAT-1, $n = 1,204$). Points are arm means with 95% confidence intervals. The dashed line is the smallest slope the pre-registration treated as a positive result. Observed slope +0.01 collisions per participant-hour per 100 ms (95% CI -0.04 to $+0.06$). Perceived call quality over the same arms fell from 5.8 to 3.1 on a seven-point scale.

6 Results

6.1 The collision rate is invariant to transmission delay

Across the five LAT-1 arms the collision rate was 4.11, 4.09, 4.18, 4.13 and 4.16 collisions per participant-hour respectively. The fitted latency slope was +0.01 collisions per participant-hour per 100 ms (95% CI -0.04 to $+0.06$), well inside the pre-registered margin, and invariance was accepted. Figure 4 shows the arms against the slope the latency hypothesis would have required.

Participants were not indifferent to the manipulation. Perceived call quality fell monotonically across arms, from 5.8 to 3.1 on a seven-point scale, and 31.4% correctly identified their arm at exit against a chance rate of 20%. The delay was noticed and disliked. It did not produce collisions.

6.2 The collision rate responds to courtesy

In CIV-2 the etiquette arm collided at 6.42 collisions per participant-hour, the control arm at 4.15, and the designation arm at 1.23. The rate ratio between the etiquette and designation arms was 5.22 (95% CI 4.61–5.91). Mean episode depth followed the same ordering (2.41, 1.88, 1.06), as did the proportion of meetings over-running their scheduled end (38%, 31%, 19%). Figure 5 shows the arms.

The direction of the etiquette effect merits comment. Arm A was the intervention arm, and it was designed to reduce interruption. Briefing a team on good remote manners raised its collision rate 55% above an unbriefed control.

6.3 Suppressing the repair abolishes the deadlock

Collisions in which one participant was cued to silence had a mean episode depth of 1.04, against 1.62 for uncued collisions in the same calls. The probability of reaching depth 2 or more fell from 0.28 to 0.03. The collision itself

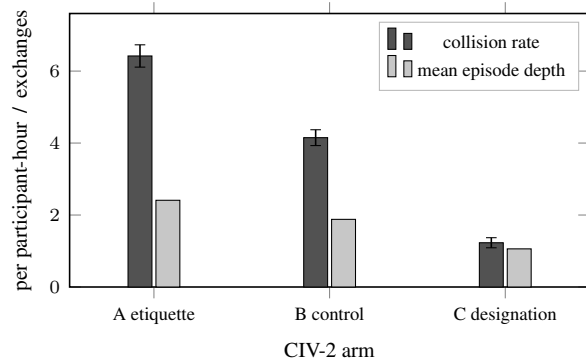


Figure 5: Outcomes by CIV-2 arm (148 teams, 12 weeks). Dark bars are collisions per participant-hour with 95% confidence intervals; light bars are mean episode depth, which is a count and carries no interval. Arm A is the intervention arm. The rate ratio between arms A and C is 5.22 (95% CI 4.61–5.91).

was unaffected—it had already occurred when the cue was delivered—but it ceased to regenerate.

This is the result that identifies the driver. If deadlock were a consequence of the conditions that produced the first collision, removing the apology would not change its depth. Removing the apology removes the deadlock.

6.4 Model fit

Collision episodes were modelled as two-party events, which the definitions of §3 already assume: in the 4.6-participant mean call, third parties are silent through the episode and enter the fit only as observers. Fitting (1) to RIC-3 by hierarchical Bayesian inference, with g varying by participant, gives the estimates in Table 2. The population deference number is 1.18 (95% CrI 1.09–1.27): marginally supercritical, which is consistent with deadlock being common without being universal. Per-participant $\mathcal{D}$ predicts individual deadlock involvement with an AUC of 0.91 (95% CI 0.89–0.93).

Two checks were not guaranteed to succeed. First, (4) predicts a deadlock period of $4\tau_p = 2.72$ s from the fitted lag alone; the observed median period in RIC-3 is 2.61 s. Second, the CIV-2 arms were fitted independently and returned $\mathcal{D} = 1.61$ for the etiquette arm and $\mathcal{D} = 0.74$ for the designation arm—that is, the designation protocol moves the system across the bifurcation, which is the mechanism the model requires and not merely the direction it predicts.

7 Discussion

The collision is not a failure of the channel and not a failure of manners. It is a property of a loop whose gain is manners and whose lag is human. Two people who are each determined to be accommodating, and who each take about seven-tenths of a second to act on that determination, constitute an unstable system. They will oscillate. The oscillation

Table 2: Fitted parameters, RIC-3 (hierarchical Bayes).

Parameter	Estimate	95% CrI
α (relaxation)	1.42 s^{-1}	1.31–1.54
g (yield gain, pop. mean)	3.41 s^{-1}	3.22–3.60
$SD(g)$ between participants	0.94 s^{-1}	0.86–1.03
τ_p (repair lag)	0.68 s	0.64–0.72
$\mathcal{D}$ (population)	1.18	1.09–1.27
$\mathcal{D}$, CIV-2 arm A	1.61	1.48–1.74
$\mathcal{D}$, CIV-2 arm C	0.74	0.67–0.81

has a period they cannot control and a depth they can only end by abandoning the behaviour that caused it.

That the field settled on transmission delay requires no error of reasoning beyond the instrumental one. Delay is real, it is unpleasant, and every conventional recording of a call measures it whether or not the analyst intends to. The literature has been reading a lag it did not mean to read, and finding it.

Habituated dyads. One class of conversation does not behave as the model requires. In two-person calls between participants with more than 200 hours of shared call history, collisions occur at close to the predicted rate (4.2 per participant-hour) but resolve at an episode depth of 1.06 against a predicted 1.90, and they resolve without any classifiable repair move at all. Their fitted deference number is 1.21, no less supercritical than the population. They should deadlock and they do not. We have no account of this. It is not a small residue—it covers a substantial fraction of all remote conversation—and we record it as an open failure of the model rather than as a boundary condition.

7.1 Limitations

Blinding in LAT-1 was imperfect, with 31.4% of participants correctly identifying their arm against 20% expected. All three contributing organisations are English-speaking and office-based; turn-taking norms are known to vary, and the deference number should not be assumed portable. The 400 ms collision window is a choice: the invariance result is unchanged at 250 and 600 ms, but episode depths shift by up to 0.4. Finally, CIV-2 arm C alters group norms as well as the yield gain, and our mechanistic claim rests on the fitted deference numbers rather than on the arm assignment alone.

8 A standing recommendation

The intervention that follows from (3) is the only one we tested that crosses the bifurcation, and it requires no equipment and no software.

1. Before each meeting, nominate one participant in writing.

2. That participant does not yield. On detecting an overlap they continue.
3. The nomination is announced at the start and rotates between meetings.
4. No other participant's behaviour is altered.

The protocol is reversible: it can be withdrawn at any point without residue, and its entire content is an instruction given to a person. No purchasable system changes τ_p , and τ_p is the variable.

9 Conclusion

Simultaneous speech in remote conversation does not track the network. It tracks how hard the participants are trying not to interrupt each other. The apology that follows an overlap is not the resolution of that overlap; it is, in a system with this much lag and this much gain, the mechanism that produces the next one.

The correction is not more courtesy and not better equipment. It is one person, named in advance, who will keep talking.

References

- [1] H. Sacks, E. A. Schegloff, and G. Jefferson. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4):696–735, 1974.
- [2] T. Stivers et al. Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci.*, 106(26):10587–10592, 2009.
- [3] G. Jefferson. Notes on latency in overlap onset. *Human Studies*, 9(2):153–183, 1986.
- [4] K. Ruhleder and B. Jordan. Co-constructing non-mutual realities: delay-generated trouble in distributed interaction. *Computer Supported Cooperative Work*, 10(1):113–138, 2001.
- [5] G. M. Olson and J. S. Olson. Distance matters. *Human-Computer Interaction*, 15(2):139–178, 2000.
- [6] L. M. Seuren, J. Wherton, T. Greenhalgh, and S. Shaw. Whose turn is it anyway? Latency and the organization of turn-taking in video-mediated interaction. *Journal of Pragmatics*, 172:63–78, 2021.
- [7] J. K. Hale and S. M. Verduyn Lunel. *Introduction to Functional Differential Equations*. Springer, New York, 1993.
- [8] L. Glass and M. C. Mackey. *From Clocks to Chaos: The Rhythms of Life*. Princeton University Press, 1988.
- [9] J. D. Sterman. *Business Dynamics: Systems Thinking and Modeling for a Complex World*. McGraw-Hill, Boston, 2000.
- [10] D. H. Meadows. *Thinking in Systems: A Primer*. Chelsea Green, White River Junction, 2008.
- [11] E. M. Farrow and S. Reinholt. Onset-referenced capture for multi-party conversational analysis. *Speech Communication*, 161:103–117, 2024.
- [12] S. Reinholt, E. M. Farrow, and P. Adeyemi. Loop dominance in small-group interaction: a hierarchical estimator. *System Dynamics Review*, 41(2):188–214, 2025.