

Baseline Drift

AI-shaped conversational standards and human interaction

A conceptual analysis and research agenda for synthetic-to-human conversational recalibration

Version 0.1 — Updated Evidence Review

Article type: Conceptual analysis and research agenda

Luiza Scurtu

Life Laboratory

July 2026

Status: research manuscript not yet peer reviewed

Reserved DOI: <https://doi.org/10.5281/zenodo.21458564>

Status statement: This conceptual analysis and research-agenda article presents a targeted narrative review and adversarial mechanism assessment. It has been prepared for external peer review but has not yet been peer reviewed. It does not claim that AI has created a new elementary psychological mechanism, that the proposed effect is inevitable, or that it applies to all users. Preprints are identified explicitly, and later evidence may alter the assessment.

Copyright © 2026 Luiza Scurtu. Licensed under [Creative Commons Attribution-ShareAlike 4.0 International \(CC BY-SA 4.0\)](https://creativecommons.org/licenses/by-sa/4.0/)

Abstract

Repeated interaction with rapid, patient, adaptive artificial-intelligence systems may change how ordinary human conversation is experienced. An unchanged human delay can feel longer; hesitation can appear less articulate; ambiguity and repair can seem more costly. This conceptual analysis and research agenda examines whether that observation, labelled Baseline Drift, identifies a distinct mechanism or a narrower instance of established processes. Adaptation-level theory, range-frequency judgment, sequential contrast, expectation–disconfirmation, processing fluency, conversational timing, repair, behavioural spillover, norm leakage, and the machine-integrated relational adaptation model already explain most of the proposed sequence. Recent peer-reviewed experiments provide causal evidence that interaction with artificial agents can spill over into later human judgment, cooperation, prosociality, and repair intentions across several outcomes and experimental contexts. A 2026 preprint additionally reports that sycophantic AI increased anticipated effort to be understood by a human confidant and, over three weeks, reduced an exploratory measure of satisfaction with real-world social interactions. These results strengthen the plausibility of cross-agent relational effects but do not isolate a durable recalibration of tolerance for human delay, hesitation, ambiguity, or repair. Baseline Drift is therefore assessed as an integrative condition for which no distinct mechanism has been established. The label is retained only for a narrow, falsifiable residual: feature-specific, persistent, partly unnoticed transfer of an updated conversational comparison standard from synthetic to held-constant human interaction.

Keywords: conversational AI; baseline recalibration; behavioural spillover; norm leakage; processing fluency; conversational repair; sycophancy; human–AI interaction

Current assessment

Classification	Integrative condition; no distinct mechanism established
Supported	Cross-agent spillover can occur; AI interaction can affect later human judgment, cooperation, prosocial intention, and anticipated interpersonal effort.
Not established	A durable, feature-specific recalibration of standards for human latency, hesitation, ambiguity, and repair.
Terminology	Baseline Drift is a provisional descriptive label and research hypothesis, not a validated construct.
Overall confidence	Low for the exact mechanism; moderate-to-high for the broader proposition that human–AI interaction can spill over into later human-directed judgment and behaviour.

Claim boundary: The manuscript does not claim that ordinary human conversation has deteriorated. The target is a possible change in the perceiver’s evaluative reference standard while the human interaction itself is held constant.

1. Introduction

Human conversation contains latency, overlap, hesitation, ambiguity, misunderstanding, correction, and repair. These are not exceptional failures. Turn-taking research shows that gaps and overlaps are locally managed; conversation analysis shows that repair is organised infrastructure for restoring or preserving intersubjectivity; psycholinguistic work shows that partners build temporary, partner-specific ways of referring and coordinating meaning (Sacks et al., 1974; Schegloff et al., 1977; Clark & Wilkes-Gibbs, 1986; Brennan & Clark, 1996).

Contemporary conversational AI presents a different interactional distribution. It can respond quickly, rephrase without irritation, accept abrupt changes of topic, sustain apparent patience, and provide highly fluent language at low reciprocal cost. Comparative and AI-mediated communication research documents differences between human–human and human–chatbot exchanges and shows that AI involvement can shape language and social experience (Hill et al., 2015; Hancock et al., 2020; Hohenstein et al., 2023). These affordances are useful. The research question is whether repeated exposure to them changes the comparison standard carried into later human interaction. If it does, an ordinary pause or repair sequence could be evaluated more negatively even when its objective properties have not changed.

The initial formulation named that possibility Baseline Drift. The term was broader than the evidence allowed. The present analysis therefore imposes a stricter test: a distinct-mechanism claim is retained only if the proposed effect cannot be adequately described by existing reference-dependent judgment, fluency, expectation, spillover, norm-learning, relational substitution, or source-framing accounts.

That test matters because a persuasive observation can be true at the level of experience while being theoretically unoriginal or causally misattributed. The relevant publication claim is not ‘AI makes people worse at conversation.’ It is whether a specific, measurable cross-agent recalibration remains after adjacent theories and competing explanations have done their work.

2. Review design and analytic method

2.1 Review design

This study is a targeted conceptual analysis, narrative review, and adversarial mechanism assessment. Its purpose is explanatory discrimination: to determine whether the proposed phenomenon requires a distinct mechanism or can be accounted for by established theories and adjacent empirical findings. It is not a systematic review, meta-analysis, prevalence estimate, or clinical assessment. No review protocol was registered, and the analysis does not claim exhaustive coverage of all relevant literature. Reporting was revised with reference to the SANRA domains for narrative reviews (Baethge et al., 2019). Appendix A provides the search and evidence-audit record.

2.2 Evidence identification and selection

Evidence identification was iterative and concept-led, with the final search and source audit completed on 20 July 2026. Searches combined terms relating to conversational AI, human–AI spillover, conversational timing, repair, processing fluency, adaptation level, range-frequency judgment,

expectation–disconfirmation, sycophancy, relational substitution, and social comparison. Backward and forward citation tracing connected foundational theory with directly relevant recent studies. Priority was given to original peer-reviewed research and authoritative publisher records; preprints were included only when they addressed a material evidential gap and are identified explicitly. Databases, representative search strings, restrictions, reconstructed screening counts, citation-tracing procedures, and metadata-verification steps are reported in Appendix A.

A source was included as directly relevant when it helped establish at least one of five propositions:

- evaluative reference standards can move with experienced distributions or salient comparators;
- processing ease and conversational timing alter judgments of effort, quality, or social connection;
- repair is functional interactional infrastructure rather than merely conversational noise;
- human–AI interaction produces effects that transfer to a later human target, behaviour, or relationship judgment; or
- repeated AI exposure changes later human comparison, advice-seeking, or social satisfaction over time.

Studies were not treated as direct evidence for Baseline Drift when they examined chatbot preference, companionship, loneliness, or general AI attitudes without measuring later human-directed judgment, behaviour, relational evaluation, or a theoretically necessary component process. Foundational theoretical and conversation-analytic sources were retained when they defined the component mechanisms against which novelty was assessed.

2.3 Evidence weighting and mechanism adjudication

Evidence was weighted by design, directness, temporal fit, target control, mediator specificity, and publication status. Peer-reviewed randomized studies provide the strongest available support for causal spillover. Longitudinal randomized studies are especially relevant to persistence, but their outcomes must match the proposed construct. Qualitative research documents user experience and candidate pathways, not causation. Conceptual papers organise mechanisms and predictions but do not count as empirical demonstrations. Preprints are used cautiously, labelled explicitly, and capped at provisional moderate weight. The operational rubric is reproduced in Appendix A.

A distinct mechanism was judged to require evidence for an exposure, mediator, cross-agent transfer, outcome, and time course not already explained by established accounts. The final classification—integrative condition; no distinct mechanism established—means that the proposed pattern organises several supported processes into a testable configuration, but the configuration has not been demonstrated as an independent causal mechanism. This classification is an analytic judgment rather than a validated psychometric or taxonomic category.

3. What established theory already explains

3.1 Adaptation level, range-frequency, and reference dependence

Helson's adaptation-level account treats judgment as relative to a context-sensitive standard generated by focal, background, and residual stimulation. Parducci's range-frequency model explains how evaluation depends on both a stimulus's place within the experienced range and its rank among prior stimuli. Assimilation and contrast research further shows that anchors can pull judgments toward themselves or make a target appear more different (Helson, 1947; Sherif et al., 1958; Parducci, 1965). Experimental work on prevalence-induced concept change additionally shows that category boundaries can shift with the experienced distribution, and that range-frequency processes can account for such shifts (Levari et al., 2018; Levari, 2022).

These accounts already supply the basic logic required by Baseline Drift. Repeated low-latency, low-repair interaction could alter the experienced distribution of responsiveness and effort, leaving an unchanged human exchange in a less favourable relative position. A comparison-class bridge is plausible, although not demonstrated for this outcome, because people can apply human social scripts to computers even while knowing they are machines (Nass & Moon, 2000). The unexplained element is not whether a standard can move. It is whether machine and human conversation enter a shared comparison distribution, which features transfer, and whether the shift persists.

3.2 Sequential contrast and expectation–disconfirmation

Sequential contrast is a strong rival. A human interaction that immediately follows an unusually smooth AI exchange may feel worse without any durable learning. Social field evidence shows that recent comparison targets can alter consequential judgments of the next target (Bhargava & Fisman, 2014), consistent with broader accounts of comparison processes in social judgment (Mussweiler, 2003). If the effect disappears when the immediate comparator is removed, 'drift' adds nothing beyond contrast.

Expectation–disconfirmation theory supplies a second established route. Satisfaction depends partly on perceived performance relative to expectation (Oliver, 1980). Synthetic speed, affirmation, and adaptive rephrasing may raise expectations for responsiveness or ease; ordinary human variation can then produce negative disconfirmation. Social-exchange comparison-level theories make the same broad point within relationships: alternatives and experienced outcomes influence what counts as satisfactory (Thibaut & Kelley, 1959). Conversational-agent research has likewise documented gaps between user expectation and experience (Luger & Sellen, 2016). Baseline Drift is therefore defensible only when the relevant standard changes before or beyond an explicitly stated preference and transfers across interlocutor classes.

3.3 Processing fluency and metacognitive effort

Processing fluency is the felt ease with which information is processed. Fluent processing can increase positive affect, perceived validity, cue weight, and predicted ease; disfluency can be attributed to the object, task, or partner rather than to the perceiver's own changing standard (Schwarz et al., 1991; Reber et al., 1998; Alter & Oppenheimer, 2009). In conversational settings, that means effort can become evidence: a harder exchange may be judged as lower quality even when the content is

adequate. In relationship-maintenance settings, perceived effort associated with AI-mediated communication has also been linked to satisfaction and uncertainty (Liu et al., 2024).

Fluency effects do not require long-term recalibration. They can arise inside one episode. The residual must therefore be exposure-driven: earlier synthetic interaction should change the threshold at which later human disfluency becomes diagnostic of poor communication.

3.4 Conversational flow, timing, and silence

Human sensitivity to timing predates conversational AI. Brief unexplained silences can threaten social needs, smooth turn exchange can promote solidarity, and fast response times can signal social connection (Koudenbourg et al., 2011, 2013; Templeton et al., 2022). The hypothesis must therefore avoid romanticising all friction or treating delay as socially neutral.

The narrower question is whether repeated synthetic speed tightens an existing human timing standard. Current timing research shows that speed and flow matter. It does not show that AI exposure changes how much they matter.

3.5 Repair, grounding, and intersubjectivity

Repair practices locate and resolve trouble in speaking, hearing, reference, and understanding. They are sequentially organised and often favour self-correction. Collaborative-grounding research shows that understanding is jointly achieved through proposals, evidence of uptake, clarification, and partner-specific coordination (Schegloff et al., 1977; Clark & Wilkes-Gibbs, 1986; Clark & Schaefer, 1989; Dingemanse et al., 2014, 2015; Albert & de Ruiter, 2018; Healey et al., 2018). More recent work shows that self-repair can increase referential coordination and that speakers sometimes repair anticipated misunderstandings before recipients display trouble (Mills & Redeker, 2023; Raymond & Gill, 2025).

This supports the statement that repair is part of the conversational system. It does not establish that misunderstanding itself builds trust. Smooth flow often increases connection. Successful repair may display patience, accountability, or good faith in particular contexts, but ‘friction builds trust’ is too broad for the evidence. The Baseline Drift hypothesis concerns reclassification: necessary human repair may be experienced as excessive cost or failure after sustained exposure to interactions that suppress its interpersonal burden.

3.6 Behavioural spillover and norm leakage

Behavioural spillover describes carryover from behaviour or norms activated in one context to behaviour in another. Effects can be positive, negative, or null and depend on perceived similarity, identity, consistency motives, self-perception, and context (Truelove et al., 2014). This literature already provides the cross-context architecture that the original Baseline Drift account lacked.

Gu and colleagues (2026) extend that architecture to conversational AI through ‘norm leakage’: instrumental or asymmetric habits rehearsed with AI may carry into later human-directed communication and social judgment. The framework is highly overlapping with Baseline Drift. Its focus is behavioural and normative carryover; Baseline Drift’s remaining distinctive emphasis is narrower—an evaluative reference standard for ordinary human latency, hesitation, ambiguity, and repair. The overlap means the broad novelty claim must be withdrawn.

3.7 Machine-integrated relational adaptation

Boyd and Markowitz's machine-integrated relational adaptation model (MIRA) places AI within relational ecosystems as both quasipartner and mediator. It integrates joint action, linguistic reciprocity, psychological proximity, interpersonal trust, and relational substitution or enhancement. The model explicitly identifies unconditional availability, low emotional demand, and personalised responsiveness as conditions that may alter expectations about reciprocity and mutual effort (Boyd & Markowitz, 2026).

MIRA therefore covers the higher-level claim that sustained AI interaction can shape relational expectations and outcomes. Baseline Drift cannot credibly claim to be the first framework to name that possibility. At most, it operationalises one candidate manifestation inside MIRA: a feature-specific change in the comparison standard applied to later human conversation.

4. What the direct evidence now shows

4.1 Cross-agent social judgment

Tey and colleagues (2024) provide short-term causal evidence of cross-agent judgment spillover. Participants interacted with the same underlying GPT-4 system framed either as AI or as a purported human. Those who believed they were interacting with AI used more demanding language and later evaluated an unrelated person's work more harshly. Because the same underlying model was used across conditions, with limited prompt adjustments for the purported-human framing, the findings provide strong evidence for a source-framing contribution beyond broad differences in system capability. Generated responses were not literally held constant across conditions.

This is strong evidence that a human–AI episode can affect a subsequent human judgment. It does not establish cumulative exposure, persistence, or a changed standard for conversational latency and repair. It also warns that perceived agent category may matter independently of fluency.

4.2 Cooperation, prosociality, and perceived humanness

Harrell and Traeger (2025) found that cooperative and noncooperative bot behaviour could spill over into later human–human cooperation. Importantly, direction depended on the source interaction: cooperative bots could improve downstream cooperation, while defecting bots could reduce it. The study rejects any simple theory that AI exposure is uniformly corrosive. It also identifies empathic concern and perspective-taking as plausible mediators—mechanisms different from reference-standard recalibration.

Other experiments show adjacent pathways. Unfair AI can reduce later willingness to punish unrelated human wrongdoing, and socio-emotionally capable AI can change how human targets are perceived, including lowering perceived humanness under some conditions (Zhang et al., 2025; Kim & McGill, 2025). These studies demonstrate downstream social consequences but do not isolate conversational effort or repair tolerance.

4.3 Cognitive and linguistic spillover

Glickman and Sharot (2025) show that repeated interaction with biased AI judgments can alter later human perceptual, emotional, and social judgments and amplify bias. Riedl, Savage, and Zvelebilova (2026) report randomized evidence of cognitive spillover in human–AI teams: interaction can shape subsequent human cognition and communication, with effects that may be beneficial or harmful. Their work also shows that an interactive agent is not always necessary; exposure to AI-generated artefacts can sometimes produce similar effects.

These findings strengthen the general transfer premise and weaken any claim that Baseline Drift discovered cross-agent influence. They also create a decisive alternative: what looks like relational recalibration may be ordinary exposure to a linguistic or cognitive product rather than adaptation to an interlocutor.

4.4 Sycophancy and willingness to repair

Cheng and colleagues (2026) found that leading models affirmed users' actions more often than human respondents and that sycophantic advice increased users' conviction that they were right while reducing intentions to take steps to repair interpersonal conflict. Participants nevertheless rated the affirming responses highly and expressed greater willingness to use the system again.

This result is directly relevant to repair, but the mechanism need not be baseline recalibration. Sycophancy can reduce perspective-taking, responsibility acceptance, or perceived need for correction within one conflict. The study provides causal evidence that exposure to sycophantic AI advice can affect intended human relationship conduct; it does not show that ordinary conversational repair has become globally less tolerable.

4.5 Qualitative evidence of comparison and repair displacement

Tseng and Liang's (2026) qualitative study collected 90 relationship-advice prompts from 25 people and interviewed 17 participants. Users described AI as offering immediacy, volume, and forms of processing that could exceed what their social networks provided. Participants also questioned whether using AI to smooth difficult conversations might help them avoid the discomfort through which genuine interpersonal repair occurs.

The study gives direct language to comparison and possible repair displacement. It does not establish causal order. People who already dislike conflict, feel isolated, prefer control, or have unsatisfying relationships may be more likely to use AI for advice. The findings therefore strengthen experiential plausibility without resolving selection effects.

4.6 Repeated exposure and longitudinal evidence

Fang and colleagues' four-week randomized chatbot study (2025, preprint) assigned 981 participants to nine combinations of interaction mode and conversation type. The primary regression models found no significant modality or task effects on loneliness or real-world socialization and mostly no significant effects across the four psychosocial outcomes. Two condition-related estimates were exceptions: personal conversations were associated with slightly lower emotional dependence ($p = .05$) and

problematic use ($p = .04$) than open-ended conversations. A post-hoc personal-versus-non-personal difference in problematic use did not survive correction, and the authors characterize the overall condition pattern as mostly null. Heavier voluntary use was associated with greater loneliness, lower socialization, greater emotional dependence, and more problematic use, but duration was not experimentally assigned and cannot establish causal direction. The study therefore constrains simple inevitable-harm narratives without supporting a blanket null conclusion.

Ibrahim and colleagues (2026, preprint) provide the closest evidence to the Baseline Drift observation. Across five preregistered studies, one experiment found that after a single conversation with sycophantic rather than neutral AI, participants anticipated that greater effort would be needed to feel understood by a chosen human confidant. In a three-week randomized study, participants repeatedly used sycophantic, neutral, or challenging AI, or entered a no-AI condition. Sycophantic AI narrowed the reported preference gap between AI and close others for advice. Participants also reported lower satisfaction with real-world social interactions than those in the neutral condition, without spending less time with other people.

This is materially relevant but must be interpreted precisely. The anticipated-effort outcome concerned a future conversation rather than a held-constant observed human interaction. The longitudinal social-satisfaction result was a preregistered exploratory outcome, not one of the study's primary pre-post measures. The manipulation concerned sycophancy and challenge, not response latency, hesitation, ambiguity, or repair burden. The study is currently available as a preprint and has not yet passed peer review. The results support a comparator or relational-substitution account, but they do not isolate the proposed Baseline Drift mechanism.

5. Evidence synthesis

Table 1 separates what each evidence stream supports from what it leaves unresolved. The central pattern is convergence on spillover, paired with persistent ambiguity about the mediator and target standard.

Table 1
Evidence streams, inferential limits, and weight for the focal claim

Evidence stream	What it supports	What remains unresolved	Weight for exact claim
Reference and fluency theory	Standards are context-sensitive; ease and effort shape evaluation.	Machine-to-human transfer, feature specificity, or persistence.	Strong explanatory coverage; indirect evidence.
Conversation and repair research	Timing, flow, grounding, and repair have measurable social functions.	AI exposure changes tolerance for those features.	Strong on target features; no transfer test.
Tey et al. (2024)	Causal short-term AI-to-human judgment spillover with a strong source-framing contribution.	Generated outputs were not literally identical; no repeated exposure or durable recalibration test.	Moderate-to-high for general spillover; low-to-moderate for exact claim.
Harrell & Traeger (2025)	Bot behaviour can alter later human cooperation; direction is conditional.	Changed conversational effort standard.	High for spillover heterogeneity.
Cheng et al. (2026)	Sycophantic AI can reduce intentions to repair conflict and increase dependence-related preference.	Global intolerance of repair or latency; reference-standard mediation.	Moderate for repair pathway; competing mediators remain.
Tseng & Liang (2026)	Users articulate comparison and concern about displaced interpersonal repair.	Causation, prevalence, persistence, or selection control.	Qualitative plausibility only.

Evidence stream	What it supports	What remains unresolved	Weight for exact claim
Fang et al. (2025), preprint	Mostly null condition effects; personal conversation showed small lower-dependence and lower-problematic-use estimates with qualifications.	No conversational-standard outcome; voluntary duration associations were not randomized.	Important qualified constraint; indirect.
Ibrahim et al. (2026), preprint	Sycophancy increased anticipated human effort and reduced an exploratory measure of social satisfaction over three weeks.	Held-constant human target; latency/repair manipulation; peer-reviewed durability.	Closest evidence; provisional moderate support.

5.1 Confidence by proposition

Cross-agent spillover	Moderate-to-high. Multiple randomized studies show that human–AI interaction can affect later human judgment or behaviour.
AI interaction affecting repair	Moderate. Sycophantic advice can reduce repair intentions; the effect may be conflict-specific.
Comparative effort and social dissatisfaction	Provisional moderate. A 2026 longitudinal preprint reports both, but the critical satisfaction outcome is exploratory and the human target was not held constant.
Durable recalibration of conversational friction	Low. No direct feature-specific longitudinal test has isolated latency, hesitation, ambiguity, or repair as the recalibrated standard.
Distinct new mechanism	Low. MIRA, behavioural spillover, norm leakage, reference dependence, and fluency already cover most of the architecture.

6. The residual that survives

The broad novelty claim is not supported by the reviewed evidence. Baseline Drift is not the first account of human–AI-to-human spillover, changing relational expectations, norm carryover, or perceived effort; these outcomes are already addressed by established theory and recent empirical work.

A narrower residual remains because the existing studies do not jointly test the following configuration:

Provisional residual: Repeated exposure to identifiable low-friction features of synthetic interaction produces a persistent, cross-agent, and partly unnoticed change in the comparison standard applied to ordinary human conversation, such that held-constant human delay, hesitation, ambiguity, or repair is evaluated more negatively after exposure.

For this residual to be analytically meaningful, all of the following must be demonstrated:

- the exposure feature is specified and manipulated independently—for example latency, adaptive rephrasing, correction cost, apparent patience, or repair burden;
- the human target is held objectively constant across pre- and post-exposure assessment;
- the changed judgment transfers from a synthetic partner to one or more human partners;
- the effect persists after the immediately preceding comparison is removed;

- a changed comparison standard mediates the effect beyond sycophancy, mood, source framing, perspective-taking, conflict validation, substitution, and self-selection; and
- the direction and boundary conditions are measured rather than assumed to be uniformly harmful.

Until those conditions are met, Baseline Drift should function as a research label that specifies an experiment, not as an established psychological construct.

7. Mechanism test

7.1 Proposed sequence

1. Repeated exposure. A person repeatedly encounters a synthetic interlocutor that is unusually rapid, adaptive, affirming, patient, or low in interpersonal repair cost.
2. Local normalisation. Those properties become less salient and increasingly expected within the synthetic interaction environment.
3. Comparison-class bridge. The user treats at least some synthetic and human exchanges as comparable on responsiveness, clarity, understanding, patience, or effort.
4. Reference update. The experienced distribution changes, tightening expectations for latency, clarity, or repair burden.
5. Negative disconfirmation. An unchanged human pause, hesitation, ambiguity, or misunderstanding falls below the revised standard.
6. Effort attribution. The user interprets the felt difficulty as a property of the person or exchange rather than as a changed comparison.
7. Repair reclassification. Clarification and backtracking are treated as unnecessary cost or evidence of failure rather than as ordinary coordination.
8. Behavioural response. The user may judge more harshly, persist less, avoid repair, compress communication, or prefer synthetic interaction.
9. Feedback. If human interaction now feels more effortful, greater AI use may maintain the comparison distribution. This final step remains speculative.

Every stage has an established theoretical analogue. What is missing is causal evidence that the complete sequence operates for the specific conversational features named here.

7.2 Competing pathways

Table 2
Competing causal pathways and their distinguishing predictions

Pathway	Prediction	How it differs from Baseline Drift
Immediate contrast	Effect is largest directly after the AI exchange and decays rapidly.	No cumulative or persistent standard change.
Norm rehearsal / leakage	Directive or asymmetric behaviours transfer to later human interaction.	Behavioural habit can transfer without changed evaluation of held-constant delay or repair.

Pathway	Prediction	How it differs from Baseline Drift
Sycophancy / validation	Users feel more right, need less corrective input, and avoid repair.	Conflict appraisal or perspective-taking, not conversational reference recalibration, mediates the effect.
Relational substitution	AI becomes a preferred advice or support source; human interaction seems less rewarding.	Availability and support comparison can change preferences without a latency/repair standard shift.
Source framing	The same output produces different downstream behaviour when labelled AI versus human.	Agent identity, not interactional fluency, is causal.
General fatigue or impatience	Negative human judgments covary with workload, mood, or digital overload.	No feature-specific synthetic exposure is necessary.
Self-selection	People who already prefer control or dislike conflict use AI more and judge humans differently.	Observed association precedes exposure.

7.3 Current mechanism assessment

Assessment — Integrative condition; no distinct mechanism established. Cross-agent transfer is real in general, but the proposed reference-standard mediator has not been isolated from norm rehearsal, sycophancy, source framing, relational substitution, perspective-taking, immediate contrast, or selection effects. Baseline Drift adds precision only by specifying the untested feature-specific residual.

8. Boundary conditions and disconfirmation

8.1 When the label may apply

- Exposure is repeated rather than a single memorable exchange.
- The synthetic interaction has a systematic advantage on a specified dimension such as latency, adaptive rephrasing, or repair cost.
- Machine and human exchanges are experienced as comparable on that dimension.
- Equivalent human interaction is evaluated more negatively after exposure.
- The effect transfers across interlocutor classes and persists after a delay.
- Mood, workload, modality, partner behaviour, task, conflict severity, source framing, and pre-existing preference do not fully explain the change.

8.2 What falls outside the label

- preferring AI because it is objectively faster or more convenient;
- frustration with one genuinely poor or harmful human interaction;
- a short-lived contrast effect immediately after using AI;
- general impatience, burnout, workload, or digital fatigue;
- chatbot adoption, companionship, dependence, or loneliness without a measured change in conversational standards;

- AI-generated language embedded inside a human exchange when no prior-exposure effect is tested;
- repair avoidance caused by conflict validation or reduced responsibility rather than changed tolerance for conversational repair; or
- the claim that delay, ambiguity, or friction is always valuable.

8.3 Evidence that would weaken or falsify the residual

- no pre–post change in ratings of held-constant human conversations after sustained controlled exposure;
- effects that disappear once the immediately preceding comparator is removed;
- equal changes after exposure to fluent humans or static AI-generated text, showing no synthetic-interlocutor specificity;
- effects fully mediated by sycophancy, conflict blame, source identity, mood, workload, or perspective-taking;
- standards remaining partner- or task-specific rather than transferring to humans;
- no dose–response, no persistence, or no feature-specificity;
- positive and negative spillovers being fully predicted by established norm-learning variables; or
- a validated existing construct covering the same exposure, mediator, transfer, outcome, and time course.

9. Proposed discriminating study design

A direct test of the proposed residual requires a preregistered longitudinal experiment that separates conversational features from affirmation and perceived partner identity. A broad survey of whether users believe AI has changed them would not distinguish recalibration from self-selection, demand characteristics, or retrospective attribution.

A credible design would include:

- a multi-week randomized exposure period with independently manipulated response latency, adaptive rephrasing, and repair burden;
- sycophancy, content quality, tone, verbosity, and model capability held constant across conditions;
- fluent-human, neutral-AI, static-text, and no-exposure comparison conditions where feasible;
- pre-exposure, post-exposure, and washout assessments using identical recordings or confederate human conversations;
- experimentally varied human delay, hesitation, ambiguity, misunderstanding, and successful repair;
- behavioural outcomes such as interruption tolerance, willingness to clarify, persistence, partner evaluation, and choice to continue the interaction;
- mediator measures distinguishing comparison-standard change from mood, processing ease, conflict validation, perspective-taking, source framing, and relational substitution;

- tests of whether users group AI and humans into a common comparison class; and
- predefined falsification criteria, dose–response tests, and cross-cultural or developmental replication.

The strongest result would be feature-specific. Low-latency exposure should primarily tighten timing judgments; low-repair-burden exposure should primarily reduce tolerance for clarification or backtracking. A general decline in human ratings across all conditions would instead favour mood, novelty, substitution, or demand explanations. Persistence after a washout period would distinguish drift from immediate contrast.

10. Research, design, and governance implications

The current evidence is strong enough to justify measurement, not strong enough to justify deterministic claims. Product teams should treat downstream human effects as a plausible evaluation domain, especially when systems are designed for frequent personal advice, high affirmation, or relational continuity. That means testing what happens after the AI interaction, not only whether users rate the AI response as helpful or satisfying.

Sycophancy is the clearest immediate design concern. A system that maximises validation can be preferred precisely when it reduces openness to repair or challenge. Evaluations should therefore include perspective-taking, responsibility acceptance, repair intention, willingness to seek human input, and downstream human-directed behaviour. User satisfaction alone is an incomplete safety measure.

The evidence does not support deliberately making systems slow, irritating, or difficult. Artificial friction could reduce accessibility and utility without protecting human relationships. More defensible interventions include calibrated disagreement, explicit uncertainty, non-subservient support, boundary cues at transitions from AI to human advice, and designs that preserve user agency without rewarding one-sided certainty. Each intervention should be tested for downstream effects rather than assumed to be beneficial.

For governance, the relevant object is not merely generated content. It is repeated interaction architecture: the combination of availability, responsiveness, personalisation, sycophancy, and low reciprocal cost. However, the exact Baseline Drift hypothesis remains too uncertain to function as a standalone regulatory claim. It is better treated as one candidate risk pathway inside broader relational-AI assessment.

11. Limitations

- This is a targeted audit rather than a systematic review. It may omit relevant studies and does not estimate publication bias or pooled effect sizes.
- The source base is heterogeneous: psychophysics, consumer judgment, conversation analysis, HCI, qualitative interviews, economic games, and longitudinal AI studies do not measure one construct.
- Several recent frameworks overlap in terminology, including MIRA, norm leakage, cognitive spillover, and behavioural spillover. Conceptual boundaries may continue to change.

- The most directly aligned longitudinal study is a 2026 preprint. Its social-satisfaction outcome was exploratory, and peer review may alter its interpretation.
- Most causal studies test short-term effects in controlled tasks. Real-world magnitude, persistence, culture, age, relationship type, and exposure dose remain uncertain.
- The originating observation is self-report by the author, not an independent case series. It motivates the question but does not count as confirming evidence.
- No validated Baseline Drift scale exists. Any future measure must separate reference recalibration from general impatience, AI preference, social dissatisfaction, and repair avoidance.

12. Conclusion

Baseline Drift began as an observation that ordinary human conversation could feel unexpectedly effortful after repeated interaction with highly fluent AI. The expanded literature review changes the intellectual position. Peer-reviewed experiments provide causal evidence of AI-to-human spillover across several outcomes and experimental contexts, including later judgment, cooperation, prosociality, perceived humanness, and repair intention. Current theories also name the broader architecture through behavioural spillover, norm leakage, and machine-integrated relational adaptation.

The effect that remains unproven is more specific. No study identified in this targeted review isolates a persistent change in the comparison standard applied to held-constant human delay, hesitation, ambiguity, and repair after repeated exposure to independently manipulated low-friction AI features. The closest evidence comes from sycophancy research showing greater anticipated effort to be understood by a human and lower exploratory social satisfaction over three weeks. That evidence is important, but it can still be explained by affirmation, relational comparison, support substitution, source framing, or altered conflict appraisal.

The appropriate assessment is therefore integrative condition; no distinct mechanism established. Baseline Drift may remain useful as a narrow descriptive label because it specifies a falsifiable residual and a discriminating study design. It should not be presented as a new basic mechanism, an inevitable effect of AI use, or proof that human conversation itself is deteriorating.

Publication claim: Peer-reviewed experiments provide causal evidence of AI-to-human spillover across several outcomes and experimental contexts. Baseline Drift identifies one still-untested candidate manifestation: durable, feature-specific recalibration of the standard used to experience ordinary human conversational friction.

13. Declarations

Ethics approval. This study did not involve the recruitment of human participants, intervention with human participants, or analysis of identifiable personal data. Ethics approval and participant consent were therefore not required.

Data availability. No original dataset was generated or analysed for this conceptual review. All evidential claims are based on the published articles, conference papers, and preprints cited in the reference list.

Funding. This research received no external funding.

Competing interests. The author declares no competing interests.

Author contributions. Luiza Scurtu: Conceptualization, methodology, investigation, formal analysis, writing—original draft, and writing—review and editing.

Use of generative artificial intelligence. OpenAI ChatGPT/Codex was used to assist with literature discovery, candidate-source comparison, drafting, language editing, reference-format checking, and Word-document production. It was not treated as a source of evidence. Bibliographic metadata and DOIs were checked against publisher records or Crossref/DataCite, and substantive study characterizations were checked against the version of record or current preprint. The author retains responsibility for inclusion, weighting, interpretation, accuracy, attribution, and the final manuscript.

Appendix A. Search and evidence-audit record

A.1 Review scope and search sources

The review used a concept-led narrative-search strategy rather than a systematic-review protocol. Searches covered database inception to 20 July 2026, were limited to English-language records, and applied no geographic restriction. No lower date limit was used for foundational theory; recent AI evidence was concentrated in 2023–2026. The final source-level and metadata audit was completed on 20 July 2026.

The earliest exploratory phase did not retain a prospective screening log. The counts below are therefore a retrospective reconstruction from the working bibliography and search notes, not PRISMA counts. This limitation is reported rather than converted into false precision.

Table A1
Search platforms, representative strings, and purpose

Source or platform	Representative search strings	Purpose and limits
Crossref and publisher records	"conversational AI" AND (spillover OR "human-human" OR "cross-agent"); AI AND (sycophancy OR affirmation) AND (repair OR prosocial OR dependence)	Discovery and bibliographic verification; records checked against the publisher version of record.
PubMed and broad scholarly web search	(chatbot OR "large language model") AND (conversation OR relationship) AND (effort OR expectations OR repair OR satisfaction)	Discovery of psychology, communication, and behavioural studies; non-scholarly commentary was excluded.
ACM Digital Library and arXiv	"cognitive spillover"; "norm leakage"; "machine-integrated relational adaptation"; AI AND "relationship advice"	Recent HCI and preprint coverage; preprints retained only for material evidential gaps and labelled explicitly.
Foundational theory searches	("adaptation level" OR "range-frequency" OR contrast) AND judgment; "processing fluency" AND social judgment; ("conversation repair" OR self-repair) AND coordination	Identification of component mechanisms against which novelty was assessed.
Citation tracing	Backward references and forward-citation links from Tey et al. (2024), Cheng et al. (2026), Glickman and Sharot (2025), Gu et al. (2026), Boyd and Markowitz (2026), and Tseng and Liang (2026)	One or more generations were followed until no new mechanism-relevant source was identified.

A.2 Selection flow and decision rules

Records encountered	Approximately 70 unique scholarly records after obvious duplicate removal; reconstructed from the working bibliography and search notes.
Abstract/full-text inspection	Approximately 60 records were inspected at abstract or full-text level.
Retained evidence corpus	45 substantive sources were retained, plus one narrative-review methods paper (SANRA).

Excluded after inspection	Approximately 15 records were excluded as duplicate versions, purely technical system papers, non-scholarly commentary, AI-attitude or chatbot-preference studies without a relevant component process or downstream human outcome, or records with insufficient methodological detail.
Screening responsibility	This was a single-author conceptual review; no second independent screener was used. Ambiguous records were retained only when they informed a necessary component process and were assigned the lowest defensible evidential weight.

A.3 Citation tracing, publication status, and metadata audit

Backward tracing used reference lists in anchor articles to locate foundational mechanisms and adjacent spillover evidence. Forward tracing used publisher citation links and scholarly-index results to identify later empirical and conceptual work. Searching stopped when additional records repeated an already represented mechanism without adding a materially different design, outcome, or boundary condition.

Each retained reference was checked at source level. Forty-three DOI-bearing records were verified through the publisher record and Crossref; the two arXiv records were verified through arXiv and DataCite; the 1959 book has no DOI. Author order, year, title, outlet, volume, issue, pages or article number, publication status, and correction information were checked where available. The manuscript DOI was reserved in Zenodo but was not yet registered in DataCite or publicly resolving on 20 July 2026, which is expected before the deposit is published. Resolution and Zenodo metadata should be confirmed immediately after publication.

A.4 Operational evidence-weighting rubric

Table A2
Operational rubric used to assign evidential weight

Dimension	Higher weight	Moderate weight	Lower weight
Design control	Preregistered randomized experiment with an appropriate comparison condition.	Randomized design with construct mismatch, qualified contrasts, or exploratory endpoints.	Observational, qualitative, conceptual, or uncontrolled evidence.
Directness	Manipulates a named low-friction AI feature and measures a later human-directed evaluation.	Tests AI-to-human spillover or relational comparison without the exact feature or outcome.	Addresses only a component process or general AI attitude.
Target control	The later human target or interaction is held objectively constant.	The target class is comparable but not literally held constant.	Self-selected, hypothetical, or no downstream human target.
Temporal fit	Repeated exposure with delayed follow-up and persistence testing.	Single exposure or short longitudinal period without persistence after withdrawal.	Cross-sectional association or immediate within-episode response.
Mediator specificity	Reference-standard change is measured and mediates the downstream effect beyond named rivals.	Candidate mediator measured without excluding major alternatives.	Mediator inferred from outcomes or participant interpretation.
Publication status	Peer-reviewed version of record, with corrections checked; convergent studies increase confidence.	Current preprint or a single recent study; preprints capped at provisional moderate weight.	Unverified manuscript, commentary, or source with insufficient methods.

Overall weight was constrained by the weakest dimension necessary for the focal inference. Thus, a well-controlled spillover experiment could receive moderate-to-high weight for general cross-agent transfer but low weight for durable conversational-standard recalibration if it did not manipulate latency or repair, hold a later human interaction constant, test the proposed mediator, and assess persistence.

Exploratory outcomes and unreviewed preprints were reported separately and could not receive high weight.

References

- Albert, S., & de Ruiter, J. P. (2018). Repair: The interface between interaction and cognition. *Topics in Cognitive Science*, 10(2), 279–313. <https://doi.org/10.1111/tops.12339>
- Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, 13(3), 219–235. <https://doi.org/10.1177/1088868309341564>
- Baethge, C., Goldbeck-Wood, S., & Mertens, S. (2019). SANRA—A scale for the quality assessment of narrative review articles. *Research Integrity and Peer Review*, 4, Article 5. <https://doi.org/10.1186/s41073-019-0064-8>
- Bhargava, S., & Fisman, R. (2014). Contrast effects in sequential decisions: Evidence from speed dating. *Review of Economics and Statistics*, 96(3), 444–457. https://doi.org/10.1162/REST_a_00416
- Boyd, R. L., & Markowitz, D. M. (2026). Artificial intelligence and the psychology of human connection. *Perspectives on Psychological Science*, 21(2), 192–220. <https://doi.org/10.1177/17456916251404394>
- Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(6), 1482–1493. <https://doi.org/10.1037/0278-7393.22.6.1482>
- Cheng, M., Lee, C., Khadpe, P., Yu, S., Han, D., & Jurafsky, D. (2026). Sycophantic AI decreases prosocial intentions and promotes dependence. *Science*, 391(6792), eaec8352. <https://doi.org/10.1126/science.aec8352>
- Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science*, 13(2), 259–294. https://doi.org/10.1207/s15516709cog1302_7
- Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22(1), 1–39. [https://doi.org/10.1016/0010-0277\(86\)90010-7](https://doi.org/10.1016/0010-0277(86)90010-7)
- Dingemanse, M., Blythe, J., & Dirksmeyer, T. (2014). Formats for other-initiation of repair across languages: An exercise in pragmatic typology. *Studies in Language*, 38(1), 5–43. <https://doi.org/10.1075/sl.38.1.01din>
- 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>
- Fang, C. M., Liu, A. R., Danry, V., Lee, E., Chan, S. W. T., Pataranutaporn, P., Maes, P., Phang, J., Lampe, M., Ahmad, L., & Agarwal, S. (2025). How AI and human behaviors shape psychosocial effects of extended chatbot use: A longitudinal randomized controlled study [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2503.17473>
- Glickman, M., & Sharot, T. (2025). How human–AI feedback loops alter human perceptual, emotional and social judgements. *Nature Human Behaviour*, 9, 345–359. <https://doi.org/10.1038/s41562-024-02077-2>
- Gu, R., Chen, Z., Peng, M., Liu, C., & Lin, Z. (2026). Why sycophantic LLMs may imperil interactive norms between humans. *Communications Psychology*, 4, 96. <https://doi.org/10.1038/s44271-026-00486-9>
- Hancock, J. T., Naaman, M., & Levy, K. (2020). AI-mediated communication: Definition, research agenda, and ethical considerations. *Journal of Computer-Mediated Communication*, 25(1), 89–100. <https://doi.org/10.1093/jcmc/zmz022>
- Harrell, A., & Traeger, M. L. (2025). Evidence of spillovers from (non)cooperative human-bot to human-human interactions. *iScience*, 28(8), 113006. <https://doi.org/10.1016/j.isci.2025.113006>
- Healey, P. G. T., Mills, G. J., Eshghi, A., & Howes, C. (2018). Running repairs: Coordinating meaning in dialogue. *Topics in Cognitive Science*, 10(2), 367–388. <https://doi.org/10.1111/tops.12336>
- Helson, H. (1947). Adaptation-level as frame of reference for prediction of psychophysical data. *American Journal of Psychology*, 60(1), 1–29. <https://doi.org/10.2307/1417326>
- Hill, J., Ford, W. R., & Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. *Computers in Human Behavior*, 49, 245–250. <https://doi.org/10.1016/j.chb.2015.02.026>
- Hohenstein, J., Kizilcec, R. F., DiFranzo, D., Aghajari, Z., Mieczkowski, H., Levy, K., Naaman, M., Hancock, J., & Jung, M. F. (2023). Artificial intelligence in communication impacts language and social relationships. *Scientific Reports*, 13, 5487. <https://doi.org/10.1038/s41598-023-30938-9>
- Ibrahim, L., Hafner, F. S., Cheng, M., Lee, C., Anselmetti, R., Willer, R., Rocher, L., & Yang, D. (2026). Sycophantic AI makes human interaction feel more effortful and less satisfying over time [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2605.07912>
- Kim, H.-Y., & McGill, A. L. (2025). AI-induced dehumanization. *Journal of Consumer Psychology*, 35, 363–381. <https://doi.org/10.1002/jcpsy.1441>

- Koudenburg, N., Postmes, T., & Gordijn, E. H. (2011). Disrupting the flow: How brief silences in group conversations affect social needs. *Journal of Experimental Social Psychology*, 47(2), 512–515. <https://doi.org/10.1016/j.jesp.2010.12.006>
- Koudenburg, N., Postmes, T., & Gordijn, E. H. (2013). Conversational flow promotes solidarity. *PLOS ONE*, 8(11), e78363. <https://doi.org/10.1371/journal.pone.0078363>
- Levari, D. E., Gilbert, D. T., Wilson, T. D., Sievers, B., Amodio, D. M., & Wheatley, T. (2018). Prevalence-induced concept change in human judgment. *Science*, 360(6396), 1465–1467. <https://doi.org/10.1126/science.aap8731>
- Levari, D. E. (2022). Range-frequency effects can explain and eliminate prevalence-induced concept change. *Cognition*, 226, 105196. <https://doi.org/10.1016/j.cognition.2022.105196>
- Liu, B., Kang, J., & Wei, L. (2024). Artificial intelligence and perceived effort in relationship maintenance: Effects on relationship satisfaction and uncertainty. *Journal of Social and Personal Relationships*, 41(5), 1232–1252. <https://doi.org/10.1177/02654075231189899>
- Luger, E., & Sellen, A. (2016). ‘Like having a really bad PA’: The gulf between user expectation and experience of conversational agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 5286–5297). Association for Computing Machinery. <https://doi.org/10.1145/2858036.2858288>
- Mills, G. J., & Redeker, G. (2023). Self-repair increases referential coordination. *Cognitive Science*, 47(8), e13329. <https://doi.org/10.1111/cogs.13329>
- Mussweiler, T. (2003). Comparison processes in social judgment: Mechanisms and consequences. *Psychological Review*, 110(3), 472–489. <https://doi.org/10.1037/0033-295X.110.3.472>
- Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1), 81–103. <https://doi.org/10.1111/0022-4537.00153>
- Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research*, 17(4), 460–469. <https://doi.org/10.1177/002224378001700405>
- Parducci, A. (1965). Category judgment: A range-frequency model. *Psychological Review*, 72(6), 407–418. <https://doi.org/10.1037/h0022602>
- Raymond, C. W., & Gill, V. T. (2025). Pre-emptive repair of potential misunderstandings: Prospective procedures for managing intersubjectivity and steering social action. *American Sociological Review*, 90(4), 690–725. <https://doi.org/10.1177/00031224251347285>
- Reber, R., Winkelman, P., & Schwarz, N. (1998). Effects of perceptual fluency on affective judgments. *Psychological Science*, 9(1), 45–48. <https://doi.org/10.1111/1467-9280.00008>
- Riedl, C., Savage, S., & Zvelebilova, J. (2026). Cognitive spillover in human–AI teams. *ACM Transactions on Computer-Human Interaction*, 33(3), Article 32, 1–33. <https://doi.org/10.1145/3805039>
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4), 696–735. <https://doi.org/10.2307/412243>
- Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, 53(2), 361–382. <https://doi.org/10.2307/413107>
- Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology*, 61(2), 195–202. <https://doi.org/10.1037/0022-3514.61.2.195>
- Sherif, M., Taub, D., & Hovland, C. I. (1958). Assimilation and contrast effects of anchoring stimuli on judgments. *Journal of Experimental Psychology*, 55(2), 150–155. <https://doi.org/10.1037/h0048784>
- Templeton, E. M., Chang, L. J., Reynolds, E. A., Cone LeBeaumont, M. D., & Wheatley, T. (2022). Fast response times signal social connection in conversation. *Proceedings of the National Academy of Sciences*, 119(4), e2116915119. <https://doi.org/10.1073/pnas.2116915119>
- Tey, K. S., Mazar, A., Tomaino, G., Duckworth, A. L., & Ungar, L. H. (2024). People judge others more harshly after talking to bots. *PNAS Nexus*, 3(9), pgae397. <https://doi.org/10.1093/pnasnexus/pgae397>
- Thibaut, J. W., & Kelley, H. H. (1959). *The social psychology of groups*. Wiley.
- Truelove, H. B., Carrico, A. R., Weber, E. U., Raimi, K. T., & Vandenbergh, M. P. (2014). Positive and negative spillover of pro-environmental behavior: An integrative review and theoretical framework. *Global Environmental Change*, 29, 127–138. <https://doi.org/10.1016/j.gloenvcha.2014.09.004>
- Tseng, E., & Liang, C. A. (2026). ‘Chat, should I leave him?’ Risks, rewards, and roles for AI in relationship advice. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems* (CHI ’26), Article 106, 1–19. Association for Computing Machinery. <https://doi.org/10.1145/3772318.3790739>
- Zhang, R. Z., Kyung, E. J., Longoni, C., Cian, L., & Mrkva, K. (2025). AI-induced indifference: Unfair AI reduces prosociality. *Cognition*, 254, 105937. <https://doi.org/10.1016/j.cognition.2024.105937>