Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Psychology

Review

Negotiation impasses as meaningful outcomes: How AI is changing the cost structure of walking away

Martin Schweinsberg[1], Stefan Thau[2] and Madan M. Pillutla[3]

Artificial intelligence (AI) is often expected to improve negotiation by strengthening information processing and analytical consistency. We argue that its effects on impasses are more selective. Extending the impasse cause, type, and resolution (ICTR) model, we distinguish two pathways. The first pathway is analytical augmentation: AI can reduce unwanted impasses caused by misunderstanding, poor option generation, and coordination failure. The second pathway runs through behavioral flexibility: Humans sometimes reach mutually acceptable agreements through aspiration adjustment, face-saving, and relational repair. Systems configured to apply stable decision thresholds and optimize focal-party outcomes without relational objectives reduce this flexibility and can thereby increase forced impasses. Which pathway dominates depends on AI's role as advisor or counterpart; its objective, training, and prompt; and whether agreement primarily requires analytical problem solving or interpersonal flexibility. These predictions concern system behavior, not whether AI experiences fatigue or social pressure: systems can be designed to respond to relational cues without experiencing them. Wanted impasses depend on both parties' underlying preferences, which AI does not change; AI instead helps negotiators recognize incompatibility sooner and justify walking away more openly. AI therefore changes the composition and meaning of impasses rather than uniformly increasing or decreasing agreement. Remaining impasses may reveal genuine incompatibility, but they may also reflect misspecified objectives or failures to encode principals' relational and economic costs.

Addresses
[1] ESMT Berlin, Schlossplatz 1, 10178 Berlin, Germany
[2] INSEAD, 1 Ayer Rajah Avenue, 138676, Singapore
[3] Indian School of Business, Gachibowli, Hyderabad, Telangana, 500 111, India

Current Opinion in Psychology 2027, **73**:102419

This review comes from a themed issue on VSI: Current Directions and New Horizons in Negotiation Research

Edited by Maurice Schweitzer, Brian Gunia, and Nazli M Bhatia

For complete overview about the section, refer Current Directions and New Horizons in Negotiation Research

Available online 26 August 2026

https://doi.org/10.1016/j.copsyc.2026.102419

Corresponding author: Schweinsberg, Martin (martin.schweinsberg@esmt.org)

Impasses, reconsidered

Negotiation scholars have often treated impasses as failures to be minimized through better information, incentives, and decision making [1–3]. This view reflects paradigms in which negotiators are given positive zones of agreement: when both parties prefer an available deal to non-agreement, an impasse is a mistake by design [4–6].

However, agreement is not the appropriate benchmark in every negotiation. In field settings, no mutually acceptable deal may exist, alternatives may be superior, or parties may derive strategic or relational value from non-agreement. Recent negotiation reviews therefore treat impasses as potentially meaningful and sometimes appropriate outcomes rather than uniformly failed agreements [7,8]. The relevant evaluative question is not simply whether negotiators agree, but whether agreement or non-agreement better serves the parties' preferences after economic, relational, temporal, emotional, and cognitive consequences are considered [6].

The ICTR model distinguishes three impasse types by the parties' preference configuration [6]. A *wanted impasse* occurs when both parties prefer non-agreement to the feasible agreements: for example, when both have superior alternatives or when agreement would violate commitments neither will compromise. It may protect alternatives or principles, although the process can still consume time or damage relationships. A *forced impasse* occurs when one party prefers non-agreement while the other prefers an available agreement. For example, a seller might reject a price below its reservation value that the buyer would accept. It imposes the loss of agreement on the deal-seeking party and may generate reputational or relational costs for both. An *unwanted impasse* occurs when both would prefer an available agreement but misunderstanding, emotion, or coordination failure prevents it, such as when parties overlook a mutually beneficial tradeoff, leading both parties to forgo value. The typology therefore shifts attention from whether an impasse occurs to whose preferences it reflects, why it occurred, and what it cost.

We extend this logic to AI-mediated negotiation. Our central argument separates two pathways. Analytical augmentation can reduce *unwanted impasses* by improving preference inference, option generation, and coordination. The second pathway operates through diminished behavioral flexibility: AI systems might use stable thresholds, optimize focal-party outcomes — the outcomes of the party the system represents - and neglect relational objectives, increasing *forced impasses*. The net effect is conditional on the system's role, objectives, training, and responsiveness to relational cues, as well as on what the negotiation requires. AI should therefore redistribute impasse types rather than uniformly raise or lower agreement rates.

The cost structure of walking away

This redistribution is partially explained by the extent to which these costs are borne by the negotiating agent or by the represented principal. Human negotiators directly experience effort, fatigue, conflict, and face concerns. AI can lower the marginal effort of preparing, generating, and evaluating offers, but it does not eliminate the principal's delay, opportunity, reputational, relational, or economic costs of non-agreement. Whether an AI system responds to those costs depends on how they are represented in its objectives and constraints [9,10].

Reported impasses illustrate why these preference and cost distinctions matter. In one survey of 202 negotiators, 17% of reported impasses were *wanted*, 39% *forced*, and 44% *unwanted* [6]. These figures describe the composition of observed impasses in that sample rather than the prevalence of each type across negotiations. Moreover, because admitting a preference for non-agreement carries social costs, negotiators may report wanted impasses as forced or unwanted. The cost structure thus shapes not only impasse behavior but how impasses are described. Related evidence shows that weak alternatives and hence high costs of impasses, can change who persists and who performs well in distributive bargaining [11].

The same human constraints that make prolonged disagreement costly can also facilitate agreement. Fatigue or social pressure may induce a mistaken concession, but aspiration adjustment, face-saving, and relational repair may enable a rational, positive-surplus settlement [12–14]. Historical accounts likewise suggest that social lubricants such as alcohol can lower interpersonal barriers, even though they may also impair judgment [15]. AI changes these processes unevenly, as Figure 1 summarizes.

How AI changes the cost structure

AI can change negotiation through two related but distinct mechanisms. First, analytical tools can generate and compare offers, simulate alternatives, identify integrative tradeoffs, and help negotiators estimate reservation values. By lowering preparation and information-processing costs, these capabilities can reduce *unwanted impasses* caused by misunderstanding, cognitive overload, or poor option generation. They do not, however, eliminate the principal's costs of delay or non-agreement, and inaccurate inputs or objectives can generate new errors.

Second, delegation can change the behavioral route to settlement. Humans sometimes concede because they tire, lower aspirations, seek face, respond to relational discomfort, or repair a strained interaction. Neither fatigue nor rapport makes an agreement mistaken by itself. The ex-ante criterion is whether, given the information reasonably available at the time and including economic, relational, temporal, and emotional consequences, the expected value of agreement exceeds the no-agreement alternative. Fatigue that induces acceptance below a properly specified reservation value is an agreement mistake; fatigue that merely lowers an aspiration and enables a positive-surplus agreement is not necessarily one. Rapport may reveal or create value, but it can also distort judgment. AI can reduce both erroneous concessions and legitimate settlement flexibility.

These mechanisms can therefore work in opposite directions. Analytical augmentation should reduce *unwanted impasses* when failures of understanding or coordination are decisive. *Forced impasses* may increase when systems reinforce focal-party aspirations, apply stable thresholds, or disregard relational costs; they need not increase when systems are trained to seek joint gains, preserve relationships, or accommodate excessively. The relevant theoretical variable is the system's decision policy and objectives, not whether it subjectively experiences fatigue, face concerns, or social pressure.

AI can also alter counterparts' reactions. Small, delayed, or infrequent concessions increase impasse rates partly because they can appear hostile or uncooperative [16]. A rigid AI-supported concession pattern may therefore provoke rejection, whereas diplomatic framing or transparent explanations may make the same substantive limit easier to accept.

The predicted result is not a uniform change in agreement but a conditional redistribution. *Unwanted impasses* should decline when analytical assistance corrects genuine coordination failures. *Forced impasses* should rise when rigid focal-party optimization displaces positive-surplus agreements that humans would otherwise reach. *Wanted impasses* may become easier to identify or justify when AI clarifies genuine incompatibility, but their frequency depends on both parties' underlying preferences. Figure 2 applies these predictions across three configurations.

Figure 1

Analytical-augmentation pathway	Behavioral-flexibility pathway
AI-supported preference inference, option generation, and information processing can reduce misunderstanding and coordination failure.	Systems configured to apply stable thresholds and optimize focal-party outcomes may become less responsive to settlement-producing social and relational cues.
Expected effect: fewer unwanted impasses when analytical failures previously prevented mutually beneficial agreement.	Expected effect: more forced impasses when agreement previously depended on aspiration adjustment, face-saving, or relational repair.
Ambiguity about interests or terms can obstruct coordination; strategically useful ambiguity can also enable face-saving and settlement.	The relevant issue is whether a system's policy responds to relational cues—not whether the system experiences them.
Which pathway dominates depends on whether the negotiation primarily requires analytical problem solving or interpersonal flexibility.	It also depends on the system's role, objective, training, prompt, and whether relational and reputational costs are encoded.
When analytical augmentation dominates	When rigid focal-party optimization dominates
Likely pattern: sharper reduction in misunderstanding-based unwanted impasses	Likely pattern: greater risk of forced impasses and lost positive-surplus agreements
Most negotiations contain both pathways; predictions are conditional, not properties of AI in general.	
Analytical and coordination failures	**Threshold rigidity and strategic deadlock**
Unwanted impasses caused by cognitive limits	Forced impasses under stable own thresholds
Misunderstanding-based unwanted impasses	Positive-surplus agreements lost to over-optimization
Coordination failures	Relationship-damaging paths to impasse

Current Opinion in Psychology

Two pathways through which AI can redistribute impasses.

Human–human negotiations (with AI support)

In the most prevalent configuration, humans negotiate with one another while using AI as advisor, drafter, or analyst. AI can reduce preparation costs and improve evaluation, but humans continue to bear relational, reputational, and emotional consequences. Effects therefore depend on how negotiators use the advice and on whether the system is optimized for focal-party value, joint gain, or another objective.

Analytical support should most clearly reduce *unwanted impasses*. Sato et al. [17] found that training negotiators to infer counterpart preferences through a visualized appraisal process improved preference inference, although it did not consistently improve integrative outcomes. Hart et al. [18] show that job candidates overestimate the risk that negotiating will jeopardize an offer; one potential role for AI advice is to identify and correct such misperceptions. These findings do not establish that AI advice produces agreement, but they illustrate analytical failures that suitable support might reduce.

The same support can contribute to *forced impasses* when it validates aggressive aspirations or treats reservation estimates as certain [19]. A negotiator who might otherwise accept a positive-surplus agreement above their reservation value but below their aspiration may continue bargaining when AI repeatedly recommends demanding better terms. Conversely, more accurate analysis can lower an inflated threshold and facilitate agreement. The

Figure 2

	Human-Human + AI support *(advisory AI)*	Human–AI *(AI as counterpart)*	AI–AI *(autonomous AI)*
Unwanted impasses (neither party preferred non-agreement)	Likely decreases when AI corrects genuine information-processing and coordination failures while humans retain relational judgment.	Mixed effects: AI may reduce some analytical errors but also introduce misinference, prompt sensitivity, or new communication failures.	Human cognitive failures may decline; system-specific failures can emerge from misspecified rewards, prompts, and deployment context.
Forced impasses (one party preferred non-agreement)	May increase under focal-party optimization that reinforces aspirations or treats uncertain reservation estimates as fixed.	May increase when the AI uses stable thresholds and lacks relational objectives; may decrease under joint-gain or accommodating policies.	Depends on objectives and protocol: persistence and flexibility can be explicitly engineered into autonomous bargaining.
Wanted impasses (both parties preferred non-agreement)	May become easier to disclose or justify when both parties genuinely prefer non-agreement; blame deflection may also change reporting.	Classification depends on principal-agent alignment: the agent's policy may reject a deal its human principal preferred.	Frequency depends on underlying feasibility and encoded principal preferences, not optimization alone.

Current Opinion in Psychology

How AI may redistribute impasses across negotiation configurations.

prediction is therefore conditional on whether AI corrects or amplifies strategic overconfidence.

AI advice may also change how non-agreement is explained. When both parties genuinely prefer non-agreement, algorithmic analysis can make a *wanted impasse* easier to justify. When only one party prefers it, blame deflection does not change the outcome into a wanted impasse; it may instead reduce the reputational cost of imposing a *forced impasse*.

Human–AI negotiations

In human-AI negotiation, a person bargains with an autonomous or semi-autonomous AI counterpart acting for another principal. The human directly experiences relational and psychological pressures, whereas the AI counterpart does not. Yet the represented principal may still bear delay, reputational, relational, and economic costs. The important asymmetry is therefore not that one side has no costs, but that the AI responds only to costs encoded in its objective, constraints, or policy.

Effects on *unwanted impasses* are mixed: an AI counterpart may process offers consistently and generate options rapidly, but it may also misinfer intent, rely on misspecified rewards, or respond sensitively to prompts and context. AI replaces some human failure modes with system-specific ones rather than simply eliminating error.

Forced impasses should become more likely when an AI counterpart applies stable reservation thresholds, optimizes focal-party value, and lacks relational objectives. They may become less likely when the system is trained or prompted to seek joint gain or accommodate the human counterpart. Shah et al. [20] report extreme anchoring by LLM negotiators, while Erlei et al. [21] show that people incur economic costs to avoid bargaining with AI, suggesting another route to premature exit.

Available evidence confirms that system design can reverse the prediction. Greater reasoning capability can produce harder reservation behavior [22], whereas RLHF can produce over-accommodation [23]; personality prompts likewise generate impasse rates ranging from near zero to over half [24]. AI is therefore not a single treatment. Its effects depend on model, training, prompt, objective, and interface. Classification also requires a stated perspective: an AI may reject a deal consistently with its encoded policy even when the human principal would have preferred agreement. From the principal's perspective, that is an *unwanted*, not a *wanted*, impasse.

AI–AI negotiations

When both parties delegate to autonomous agents, neither negotiating process is constrained by human fatigue or face concerns. Marginal computational bargaining costs may be low, but principals still bear delay, opportunity, relational, and economic consequences. Those consequences affect behavior only if the systems' objectives or protocols represent them.

Some human sources of *unwanted impasse* may decline because agents can exchange information and iteratively compare proposals. New unwanted impasses can arise from reward misspecification, prompt-dependent interpretation, deployment artifacts, and generalization failure. Consistent with this conditionality, prompt changes have shifted AI—AI deal rates from 27% to 89% [25], deployment language can matter more than model architecture [26], and interaction styles resembling warmth and dominance shape outcomes even in automated bargaining [27].

Forced impasses do not lose all social meaning merely because agents conduct the exchange. The principals may interpret a walkaway as hostile or consequential, and repeated-market reputation can matter even when agents do not experience it. Systems such as ASTRA [28] and ASTRO [29] instead show that persistence, cooperation, and walkaway behavior can become explicit design parameters. Engineering choices therefore relocate, rather than eliminate, the social and strategic consequences of impasse.

Wanted impasses arise in AI—AI bargaining only when both principals prefer non-agreement to the feasible deals, or when that preference has been validly delegated. Optimization by itself does not make such impasses routine; it may simply identify incompatibility more quickly. If agent objectives diverge from principal preferences, the same terminal outcome may be agent-consistent but principal-unwanted. This principal-agent distinction is necessary before AI-mediated impasses can be interpreted.

Research agenda

Our framework shifts the question from whether AI increases or decreases agreement to which impasse types it changes, through which mechanism, and from whose perspective. Empirical attention is uneven: human-AI negotiation attracts substantial attention, while human—human negotiation with AI support remains thinly studied, even though it's likely the most prevalent use-case in the near future. AI—AI research often treats agreement efficiency as the outcome without measuring ICTR preference configurations.

One priority is to manipulate system objectives directly. Systems optimized for focal-party value, joint gain, relationship preservation, or agreement should produce different impasse distributions even when model architecture is held constant. The tension between harder reasoning-based reservation behavior and RLHF-based accommodation [23] illustrates why model labels alone are theoretically insufficient. Research should measure the objectives, prompts, constraints, and decision rights that produce a walkaway and cross these system properties with negotiations that primarily require analytical processing versus interpersonal flexibility.

How an impasse is reached should be studied as a dimension orthogonal to its type. The same wanted, forced, or unwanted impasse can follow hostile escalation and abrupt withdrawal or transparent explanation and respectful closure. These paths should have different consequences for procedural fairness, relationship damage, reputation, and future bargaining [6,16]. AI may reduce relational damage through neutral language, consistent procedures, or face-saving explanations; it may increase damage through rigidity, depersonalization, inappropriate tone, or perceived procedural illegitimacy.

Research must also specify whose preferences define the outcome. As delegation grows, principal and agent objectives may be only partially aligned. Studies should therefore measure both the principal's ranking of agreement versus non-agreement and the agent's encoded or elicited objective. Digital traces make concession patterns, thresholds, and persistence observable at scale, but observation alone does not reveal whether an impasse was wanted, forced, or unwanted. That classification requires preference evidence at the relevant level of agency.

Conclusion

Impasses are not uniformly failures, and agreement is not uniformly success. AI changes negotiation through at least two separable pathways: analytical augmentation can reduce unwanted impasses caused by misunderstanding and poor option generation, while systems configured for stable focal-party optimization can reduce the behavioral flexibility that enables some positive-surplus settlements and thereby increase forced impasses. Neither effect is a property of AI in general. It depends on the system's role, objective, training, prompt, and responsiveness to encoded relational costs, as well as on what the negotiation requires. Interpreting the resulting impasse also requires identifying whose preferences count: an agent's policy may reject a deal that its principal would have preferred. AI will therefore make some impasses more diagnostically informative and create others that reveal design and delegation failures rather than genuine incompatibility.

CRediT statement

Martin Schweinsberg: Conceptualization, Methodology, Writing - Original draft, Writing - Review & editing, Visualization.

Stefan Thau: Conceptualization, Methodology, Writing - Original draft, Writing - Review & editing, Visualization.

Madan M. Pillutla: Conceptualization, Methodology, Writing - Review & editing, Visualization.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the authors used Claude (Anthropic) to assist with literature search, source verification, copy-editing, and manuscript preparation. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

References of particular interest have been highlighted as:

* of special interest
** of outstanding interest

1. Bazerman HM, Neale MA: *Negotiating rationally*. New York: Free Press; 1992.
2. Fisher R, Ury WL: *Getting to YES*. London: Penguin; 1981.
3. Raiffa H: *The art and science of negotiating*. Cambridge, MA: Harvard University Press; 1982.
4. Thompson L, Hastie R: **Social perception in negotiation**. *Organ Behav Hum Decis Process* 1990, **47**:98–123, https://doi.org/10.1016/0749-5978(90)90048-E.
5. Bazerman MH, Curhan JR, Moore DA, Valley KL: **Negotiation**. *Annu Rev Psychol* 2000, **51**:279–314, https://doi.org/10.1146/annurev.psych.51.1.279.
6. Schweinsberg M, Thau S, Pillutla MM: **Negotiation impasses: types, causes, and resolutions**. *J Manag* 2022, **48**:49–76, https://doi.org/10.1177/01492063211021657.
7. Boothby EJ, Cooney G, Schweitzer ME: **Embracing complexity: a review of negotiation research**. *Annu Rev Psychol* 2023, **74**: 299–332, https://doi.org/10.1146/annurev-psych-033020-014116.
8. Schweitzer M, Krueger K, Boothby E, Cooney G: **Negotiation**. In *Handbook of social psychology*. Edited by Gilbert D, Fiske S, Finkel E, Mendes W. 6th ed., Situational Press; 2025, https://doi.org/10.70400/NUZE7621.
9. Walton RE, McKersie RB: *A behavioral theory of labor negotiations*. McGraw-Hill; 1965.
10. Pruitt DG, Rubin J: *Social conflict: escalation, stalemate, and settlement*. New York: Random House; 1986.
11. ** Ma A, Ponce De Leon R, Rosette AS: **Asking for less (but receiving more): women avoid impasses and outperform men when negotiators have weak alternatives**. *J Appl Psychol* 2024, **109**:1145–1158, https://doi.org/10.1037/apl0001138.
12. Simon HA: **A behavioral model of rational choice**. *Q J Econ* 1955, **69**:99, https://doi.org/10.2307/1884852.
13. March JG, Simon HA: *Organizations*. New York: Wiley; 1958.
14. Pruitt DG: *Negotiation behavior*. New York: Academic Press; 1981.
15. Sciolino E, Cohen R: **In U.S. eyes, "good" Muslims and "bad" Serbs did a switch**. *N Y Times* 1995.
16. Mertes M, Kunz D, Hüffmeier J: **A deep dive into distributive concession making and the likelihood of impasses in negotiations**. *Collabra: Psychology* 2023, **9**, 88929, https://doi.org/10.1525/collabra.88929.
17. * Sato M, Terada K, Gratch J: **Teaching reverse appraisal to improve negotiation skills**. *IEEE Trans Affective Comput* 2024, **15**:872–884, https://doi.org/10.1109/TAFFC.2023.3285931.
18. Hart E, Bear JB, Ren B: **But what if I lose the offer? Negotiators' inflated perception of their likelihood of jeopardizing a deal**. *Organ Behav Hum Decis Process* 2024, **181**, 104319, https://doi.org/10.1016/j.obhdp.2024.104319.
19. Maaravi Y, Heller B, Levy A: **Low power, first offers, and reservation prices: weak negotiators are self-anchored by their own alternatives**. *Negot J* 2023, **39**:7–34, https://doi.org/10.1111/nejo.12423.
20. Shah C, Agarwal A, Garg K, Heddaya M. LLM rationalis? Measuring bargaining capabilities of AI negotiators 2025. [Preprint]. https://doi.org/10.48550/ARXIV.2512.13063.
21. Erlei A, Das R, Meub L, Anand A, Gadiraju U: *For what it's worth: humans overwrite their economic self-interest to avoid bargaining with AI systems. CHI conference on human factors in computing systems*. New Orleans LA USA: ACM; 2022:1–18, https://doi.org/10.1145/3491102.3517734.
22. * Kitadai A, Rico Lugo SD, Tsurusaki Y, Fukasawa Y, Nishino N. *Can AI with high reasoning ability replicate human-like decision making in economic experiments? Group Decis Negot*, vol. 34; 2025:1303–1326, https://doi.org/10.1007/s10726-025-09946-9.
23. ** Araujo DKG, Uhlig H: *How does AI distribute the pie? Large language models and the ultimatum game*. Cambridge, MA: National Bureau of Economic Research; 2026, https://doi.org/10.3386/w34919.
24. * Kwon D, Shrestha K, Han B, Lee EH, Lucas G: *Evaluating behavioral alignment in conflict dialogue: a multi-dimensional comparison of LLM agents and humans. Proceedings of the 2025 conference on empirical methods in natural language processing*. Suzhou, China: Association for Computational Linguistics; 2025:16366–16380, https://doi.org/10.18653/v1/2025.emnlp-main.828.
25. * Xia T, He Z, Ren T, Miao Y, Zhang Z, Yang Y, *et al*.: *Measuring bargaining abilities of LLMs: a benchmark and a buyer-enhancement method. Findings of the association for computational linguistics ACL 2024, Bangkok, Thailand and virtual meeting*. Association for Computational Linguistics; 2024: 3579–3602, https://doi.org/10.18653/v1/2024.findings-acl.213.
26. * Sinha S, Kumar H, Mandapati AR, Sakhuja R, Kumar D. The language of bargaining: linguistic effects in LLM negotiations 2026. [Preprint]. https://doi.org/10.48550/ARXIV.2601.04387.
27. ** Vaccaro M, Caosun M, Ju H, Aral S, Curhan JR: **Advancing AI negotiations: a large-scale autonomous negotiation competition**. *Proc Natl Acad Sci USA* 2026, **123**, e2521774123, https://doi.org/10.1073/pnas.2521774123.
28. Kwon D, Hae J, Clift E, Shamsoddini D, Gratch J, Lucas G: *ASTRA: a negotiation agent with adaptive and strategic reasoning via tool-integrated action for dynamic offer optimization. Proceedings of the 2025 conference on empirical methods in natural language processing*. Suzhou, China: Association for Computational Linguistics; 2025:16228–16249, https://doi.org/10.18653/v1/2025.emnlp-main.821.
29. Hu Y, Huang C, Lei W: *ASTRO: automatic strategy optimization for non-cooperative dialogues. Findings of the association for computational linguistics: ACL 2025*. Vienna, Austria:

Association for Computational Linguistics; 2025:388–408, https://doi.org/10.18653/v1/2025.findings-acl.22.

Further information on references of particular interest

11. ** Across eight studies, including Shark Tank pitches where women reached impasse 40% of the time versus 48% for men, incorporating the economic costs of impasse into performance measures reverses the standard male advantage in distributive negotiation when alternatives are weak. Whether assertive behavior helps or hurts depends on the cost structure around non-agreement and this is exactly the variable AI most directly reshapes.

17. * Training negotiators to infer counterpart preferences from emotional expressions improved partner-model accuracy but not integrative outcomes. The dissociation shows that AI's analytical gains do not automatically translate into better deals.

22. * In Ultimatum Games, LLMs with stronger reasoning behaved closer to a strict reservation-point logic than weaker-reasoning models. Reasoning capacity is therefore a lever pushing AI bargaining toward harder forced impasses, in tension with alignment training pressures.

23. ** Demonstrates that alignment training can shift LLM bargaining toward over-accommodation, directly qualifying predictions of uniformly harder AI reservation behavior.

24. * In matched simulations of a multi-issue commercial dispute, impasse rates across four personality-prompted LLM dyads ranged from ~0 for Claude-3.7-Sonnet to 0.53 for Gemini-2.0-Flash, with GPT-4.1 variants in between and humans at 0.13. Model choice alone reshapes the distribution of agreements versus breakdowns at a scale dwarfing typical between-condition effects in human dispute studies.

25. * A prompt-based buyer-enhancement method moved AI–AI deal rates from 27% to 89% in a multi-turn benchmark. The shift, larger than typical between-model differences, indicates AI–AI impasse rates are downstream of scaffolding rather than model capability.

26. * In LLM negotiations, the language of bargaining instructions affected agreement rates more than model choice. The result reframes AI negotiation behavior as partly a property of the linguistic surface on which models were trained and prompted.

27. ** In an international competition producing over 180,000 autonomous AI–AI negotiations, warmth-associated language predicted deals and value while dominance-associated behavior predicted impasses.