UC Davis

UC Davis Electronic Theses and Dissertations

Title

Towards a Communicative Process Design for Human-AI Team Collaboration

Permalink

https://escholarship.org/uc/item/0vw9r6z7

ISBN

9798193266433

Author

Chang, Rong-Ching

Publication Date

2026-09-08

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Towards a Communicative Process Design for Human-AI Team Collaboration

By

RONG-CHING CHANG

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Computer Science

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Hao-Chuan Wang, Chair

Susan R. Fussell

Seth Frey

Committee in Charge

2026

Abstract

As AI systems shift from individual tools to collaborative team members, the communicative processes through which humans and AI partners coordinate remain poorly understood. Decades of research have established how human interlocutors accommodate, ground, and coordinate in teamwork, but whether these patterns extend to human-AI teams, how they vary across cultures and team configurations, and whether designed support can serve diverse users equally are open questions. This dissertation develops a communicative process design approach for human-AI team collaboration, asking how human-AI communicative processes are, and can be, shaped by culture, team structure, and designed support.

Four studies across collaborative tasks, from group brainstorming to negotiation, were designed, conducted and integrated through three cumulative stages. The first stage is a characterization stage that aims to understand how communicative processes such as accommodation and grounding operate in human-AI interaction and how they vary across cultural populations: an analysis of 8,011 user-agent conversations from 1,500 participants spanning 75 countries reveals that linguistic accommodation varies by cultural population, with AI agents that accommodate well on average failing specific cultural groups (Chapter 3). Separating affective accommodation from grounding within the same negotiation sample ($N = 40$) establishes that these operate as distinguishable constructs (Chapter 4). The second stage is a formation stage that explores how team formation shapes the communicative processes that the characterization stage reveals. A 2×2 factorial experiment ($N = 168$) crossing human-AI team size and formation of AI behaviors within the team finds that user-parameterized formation creates qualitatively distinct communicative modes,

shifting the AI from co-generation to a coaching profile, while a second AI erases the formation effect by raising group output and reducing human authorship (Chapter 5). In the intervention stage that aims to improve human-AI communication and collaboration, multi-agent coaching for cross-cultural negotiation ($N = 80$) improves message persuasiveness, but whether this gain reaches the task outcome depends on horizontal collectivism, an individual-level cooperative orientation, not regional category (Chapter 6). The convergent finding across the communicative process design approach is that population variation identified in characterization returns as the moderation that determines who benefits from designed communicative support, and that human communicative behavior shifts through role redistributions rather than uniform amplification.

These results motivate a conceptual framework for understanding alignment in human-AI collaboration. The dissertation proposes an iceberg framework that separates visible AI behavior from hidden user assumptions and model processes, and traces three interaction design paradigms, free-form interaction, behavioral engineering, and human-AI co-configuration, along a trajectory in which communicative processes become increasingly revisable during collaboration (Chapter 7).

The dissertation makes three contributions to the understanding and programmatic design of communicative processes in human-AI teams. First, it establishes a characterize-before-you-design principle: effective communicative process design must account for population variation rather than assume a universal user, and culturally aware systems should infer cooperative orientation from interaction rather than assign support by demographic category. Second, the iceberg framework and design paradigms provide a conceptual vocabulary for the field: the iceberg identifies where alignment interventions operate, and the co-configuration paradigm reframes alignment as an interaction property maintained through ongoing mutual grounding rather than being fixed at design time. Third, the measurement architecture separates human from AI authorship, affective accommodation from grounding, global from local diversity, and message-level gains from task outcomes, providing reusable constructs for benchmarking communicative dynamics across future model generations. The gap between initial parameterization and iterative co-configuration is the central design frontier that this work opens.

Contents

Abstract **ii**

List of Figures **xi**

List of Tables **xiv**

Acknowledgments **xv**

1 Introduction **1**

1.1 Motivation: From Copilots to Teammates 1

1.2 Gaps in the Current Understanding 3

1.3 Contributions and Dissertation Overview 5

2 Theoretical Background and Related Work **9**

2.1 Communicative Accommodation and Grounding 10

2.1.1 Communication Accommodation Theory (CAT) 10

2.1.2 Grounding in Communication 11

2.1.3 Accommodation and Grounding as Complementary Aspects of Communication 12

2.2 Cultural Variation in Communication 12

2.3 Collaborative Task Domains 14

2.3.1 Group Brainstorming and Process Losses 15

2.3.2 Persuasion and Negotiation 16

2.4 Human-AI Communication and Collaboration 17
2.4.1 Individual AI Productivity and the Shift to Teams 18
2.4.2 AI-Mediated Coaching and Communicative Support 19

3 Characterization: Linguistic Accommodation in Human-AI Interaction Across Cultures 20

3.1 Introduction 21
3.2 Related Work 23
3.2.1 Communication Accommodation Between Users and Conversational Agents . 23
3.2.2 Direction of Communication Accommodation 23
3.3 Dataset 24
3.4 Method 27
3.4.1 Impact of Cultural Dimensions on Communication Accommodation 28
3.4.2 Impact of Conversation Length on Communication Accommodation 29
3.4.3 Direction of Communication Accommodation 30
3.5 Results 30
3.5.1 RQ1: Do national cultures of different placements along Hofstede's cultural dimensions influence linguistic style-matching between users and LLM-powered conversational agents? 30
3.5.2 RQ2: Does communication accommodation between users and conversational agents increase as the conversation is prolonged? 31
3.5.3 RQ3: Do users adapt their communication styles to the conversational agents and vice versa? 34
3.6 Discussion 36

4 Characterization: Affective Accommodation and Grounding in Human-AI Negotiation 39

4.1 Introduction 40

4.2 The Current Study . . . 42
4.3 Methods . . . 45
4.3.1 Dataset and Task . . . 45
4.3.2 Operationalizing Language-Mediated Affective Accommodation . . . 45
4.3.3 Operationalizing Conversational Affective Grounding . . . 46
4.4 Results . . . 46
4.4.1 Language-Mediated Affective Accommodation . . . 46
4.4.2 Conversational Affective Grounding . . . 47
4.5 Discussion . . . 48

5 Formation: AI Team Size and Behavioral Configuration in Human-AI Brainstorming 50

5.1 Introduction . . . 52
5.2 Related Work . . . 56
5.3 Method . . . 58
5.3.1 Design and Participants . . . 58
5.3.2 Task and Platform . . . 58
5.3.3 Procedure . . . 59
5.3.4 AVG and CUS Formations . . . 59
5.3.5 AI Implementation . . . 61
5.3.6 Measures . . . 61
5.4 Results . . . 62
5.4.1 Does Adding More AI Help Productivity? (RQ1) . . . 62
5.4.2 Does Adding More AI Increase Output Quality? (RQ2) . . . 67
5.5 Post-Hoc Analyses and Discussions . . . 69
5.5.1 Group output gains do not transfer to human authorship . . . 70
5.5.2 A second AI partner increases local differentiation more than global breadth 76

5.5.3 Implications . . . 78
5.5.4 Limitations . . . 80

6 Intervention: Multi-Agent Coaching for Cross-Cultural Negotiation 82

6.1 Introduction . . . 83
6.2 Background . . . 87
6.2.1 Persuasion and Strategic Negotiation . . . 87
6.2.2 Culture and Negotiation Behavior . . . 87
6.2.3 LLM Agents in Coaching . . . 88
6.2.4 Study Design . . . 89
6.3 Task-Setting: The Fundraising Game as a Negotiation Environment . . . 89
6.4 The PersuaBridge Framework: AI-mediated Negotiation Training System . . . 91
6.4.1 Culturally Aware Coaching . . . 92
6.4.2 Negotiation Counter Partner . . . 96
6.4.3 Negotiation Partner Architecture . . . 96
6.5 Experiment . . . 97
6.5.1 Participants and Procedure . . . 98
6.6 Results . . . 100
6.6.1 RQ1: How Do Humans Perform Against AI Counterparts in Negotiation? . . . 100
6.6.2 RQ2: How Do Coaching and Cultural Alignment Jointly Shape Outcomes? . . . 100
6.6.3 RQ3: What Is the Relationship between an AI Partner's Reasoning Effort and a Participant's Persuasive Effort? . . . 103
6.7 Post-Hoc Analyses and Discussion . . . 105
6.7.1 RQ1: Regional Differences in Negotiation Performance . . . 106
6.7.2 RQ2a: How Coaching Relates to Persuasiveness Across Regions . . . 106
6.7.3 RQ2b: Why Coaching Affects Donation Differently Across Regions . . . 107
6.7.4 RQ3: AI Reasoning Effort and Participant Persuasive Effort . . . 108

6.8 Implications for Human-AI Negotiation Support . . . 109

7 General Discussion **111**

7.1 Summary of Findings . . . 111

7.2 The Three-Stage Cumulative Structure . . . 113

7.3 From Initial Parameterization to Iterative Co-Configuration . . . 114

7.4 General Limitations . . . 118

7.5 Future Directions . . . 120

7.6 Conclusion . . . 121

A Supplementary Tables for Chapter 5: AI Team Size and Behavioral Formation **123**

A.1 Complete OLS Models Predicting Human Board Items . . . 123

A.1.1 Social/Filler Model . . . 123

A.1.2 Generative Model . . . 124

A.1.3 Evaluative Model . . . 124

A.1.4 Questioning Model . . . 125

B Supplementary Materials for Chapter 6: Multi-Agent Coaching **126**

B.1 Validating Agent Culturalization and Value Embedding via Survey Comparison . . . 126

B.2 Validating Cultural Value in Generated Messages via Human Ratings . . . 127

B.3 Supplementary Model Tables . . . 128

List of Figures

1.1 Idea flow of communicative process design in this dissertation. 6

3.1 Geographic distribution of participants across Hofstede's cultural dimensions (1 of 2). Each map shows country-level quantile assignment and sample size. 25
3.2 Geographic distribution of participants across Hofstede's cultural dimensions (2 of 2). 26
3.3 Relative presence of LIWC *category groups* showing convergence versus divergence in prolonged conversations. 32
3.4 Heatmap of communication adaptation across Hofstede's cultural dimensions. 34

4.1 Study design and analysis framework. 40
4.2 Summary statistics and lexical distributions from the donation negotiations, including asymmetric and model-dependent affective accommodation. 44

5.1 Overview of the experimental design. The study crosses AI team size (one versus two partners) with AI formation (AVG versus CUS), alongside a human-human reference condition. 51
5.2 Collaborative brainstorming workspace. Human participants and AI partners exchanged messages in real time while contributing ideas to a persistent shared board. 52

5.3 Productivity and human-authorship results. Panel (a) shows group board items by condition with the 40-item quota. Panel (b) shows human-authored board items per person. Panel (c) shows human messages per human board item. Panel (d) shows the OLS model linking AI board volume and social/filler talk to human board items. 65

5.4 Paired human and AI message-act composition by condition. Stacked bars show the share of generative, evaluative, questioning, and social/filler messages. HH has only a human bar. 67

5.5 Global and local differentiation of human-authored focal board items. Error bars show 95% confidence intervals. The two-AI versus one-AI difference is detected locally but not globally. 69

6.1 Participant-facing PersuaBridge interaction. The fundraiser exchanges persuasive messages and counteroffers with a culturally aligned or misaligned AI donor while a private AI coach identifies strategy, provides feedback, and suggests improved phrasing. 83

6.2 Fundraising negotiation task. Participants are assigned a fundraising scenario (counterbalanced), negotiate over multiple turns with an AI donor whose opening offer is randomized between $10 and $40, and receive a final donation. 84

6.3 Overview of the PersuaBridge architecture. The framework integrates two LLM agents: a culturally grounded *coach* that provides real-time, turn-level feedback, and a *counter player* that simulates the donor in multi-turn negotiation. The coach performs adaptation checking, strategy identification, and culturally aware feedback generation to support user learning, while the counter player combines an interpretation module and a response module to assess persuasiveness, reason internally through a five-step chain-of-thought, and generate culturally coherent replies. In combination, the modules realize a global, culturally conditioned counterpart and support user adaptation across culturally varied agents. 91

6.4 Opening-relative round outcomes by region. Asia participants finish above the opening in 80% of rounds versus 46% for US participants. 100
6.5 Mean persuasiveness score by region and coaching condition. 102
6.6 Coaching × HC interaction on final donation. At low HC, uncoached participants donate more ($34) than coached ($27); at mid and high HC, coached participants donate slightly more, consistent with the HC-by-coaching interaction ($\beta = +3.58$, $p = .006$). 103
6.7 AI reasoning effort by region and CoT field. Asia-born participants consistently elicit more reasoning effort across all five fields (all $p < .001$). Only User Strategy Assessment shows a significant coaching-by-region interaction (+9.687, $p = .007$). . . 103

7.1 The iceberg framework contrasts visible AI behavior with hidden user assumptions and model processes, creating asymmetric inference where AI can often infer users better than users can infer AI commitments. 116
7.2 Three interaction design paradigms for human-AI interactions, from free-form interaction through behavioral engineering to human-AI co-configuration. 118

B.1 AI Cultural Survey Results comparing baseline vs. finetuned models under individualistic vs. collectivist role assignments. Ratings reflect average survey response on a 1–9 scale. 127

List of Tables

3.1 Example LIWC Categories and Sample Words 27

5.1 Factorial estimands for human-authored board items per person, using OLS estimates (N=120). 63

5.2 Human messages per human board item: pairwise contrasts across conditions. 64

5.3 CUS-versus-AVG AI behavioral profile contrasts by AI count. 66

6.1 Regional coaching interaction on message persuasiveness. 101

6.2 Regional coaching contrasts (derived from Table 6.1). 101

6.3 HC-by-coaching moderation of final donation. 102

6.4 Region-by-coaching model of AI reasoning effort (CoT length). 105

A.1 OLS model predicting human-authored board items with AI social/filler share as the focal predictor (R^2=.338). 123

A.2 OLS model predicting human-authored board items with AI generative share as the focal predictor (R^2=.315). 124

A.3 OLS model predicting human-authored board items with AI evaluative share as the focal predictor (R^2=.317). 124

A.4 OLS model predicting human-authored board items with AI questioning share as the focal predictor (R^2=.311). 125

B.1 Regional coaching interaction on message persuasiveness. The outcome is the AI-evaluator persuasiveness score. The model includes coaching condition, participant region, their interaction, alignment condition, game round, and the counter player's opening offer as fixed effects, with participant random intercepts (N = 1,613 scored messages, 80 participants). 128
B.2 Regional coaching interaction on final donation, anchor-adjusted. The model controls for the counter player's opening offer, alignment condition, and game round (N = 480 rounds, 80 participants). 129
B.3 Final donation conditional on current-message persuasiveness. The outcome is the round's final agreed donation amount, repeated on each scored participant-response row. This message-weighted specification controls for the persuasiveness score, within-round response order, and the counter player's opening offer (N = 1,613 response rows covering 480 rounds, 80 participants). 129
B.4 Regional coaching contrasts conditional on persuasiveness, derived from Table B.3. . 129
B.5 HC × coaching interaction with region control. Participant region is added as a covariate alongside the horizontal collectivism moderation (N = 480 rounds, 80 participants). 129
B.6 Internal AI chain-of-thought fields predicting next-human-message length. Each row is the focal coefficient from a separate multilevel model predicting the character count of the next visible participant message from one CoT field's character count, participant region, alignment, and turn number, with participant random intercepts (N = 1,161 pairs, 80 participants). Holm correction applied within the five-test family. 130
B.7 CoT field × Region interactions on next-human-message length. Five-test Holm family. Same specification as Table B.6 with additive control replaced by interaction. 130
B.8 CoT field × Alignment interactions on next-human-message length. Five-test Holm family. Same specification as Table B.6 with additive control replaced by interaction. 130

B.9 Region effect and coaching × region interaction by individual CoT field. Each row is from a separate multilevel model with one CoT field's character count as outcome. $N = 1{,}161$ pairs, 80 participants. 130

Acknowledgments

I am grateful to my advisor, Hao-Chuan Wang, for his guidance throughout this work. His research on intercultural communication and group creativity shaped the questions this dissertation asks, and his sustained investment in my development as a researcher made it possible to pursue them.

I thank my committee members, Susan R. Fussell and Seth Frey, for their insights and feedback. Their perspectives on grounding in communication and collective behavior strengthened this work at every stage.

This dissertation was supported in part by UC Davis' Graduate Group in Computer Science through its assistantships.

I dedicate this work to my family: Patricia Haseltine, Hui-Yun Yang, and Lan-Ting Chang, whose support carried me through this journey.

Chapter 1

Introduction

1.1 Motivation: From Copilots to Teammates

AI systems are increasingly deployed as collaborative partners rather than personal tools. The trajectory began with individual assistants that proved effective at speeding task completion and raising output quality in domains from writing [84] to customer support [13]. These successes have motivated multi-agent architectures in which AI systems participate not as personal tools but as collaborative team members, yet placing humans and AI in a shared workspace does not guarantee that the resulting conversation is effective or that the team coordinates well. Research on human collaboration has shown that effective coordination depends on communicative processes that go beyond exchanging messages: interlocutors accommodate to each other's styles, build mutual understanding through grounding, and repair misalignments as they arise [20, 41]. Whether these processes occur in human-AI teams, how team composition reshapes them, and whether designed support can serve diverse users equally remain poorly understood. These open questions motivate this dissertation's investigation across cultures, tasks, and team configurations.

Human-AI teams are already deployed across negotiation, education, and creative work: AI agents mediate retail deals [23, 102], coaching systems scaffold deliberate practice [2, 111], and multi-agent platforms position AI as a brainstorming colleague [65, 97]. These deployments produce

measurable output gains, but the dynamics differ from what decades of research on human teams would predict. In human teams, adding members introduces process losses that offset the benefit of pooling more minds [27]; in human-AI teams, a different pattern emerges: output rises but results become more homogeneous [56], and collective diversity of ideas increases without changing any individual's creativity [3]. On the process side, working with AI demands continuous monitoring and steering [95], and participants value AI contributions but resist when the agent dominates the conversation [51]. Taken together, these patterns suggest that communicative processes in human-AI teams are not simply accelerated versions of human teamwork; they introduce qualitatively different dynamics that require systematic investigation.

Cultural variation adds a further dimension to these communicative dynamics, because in human-AI teams, culture becomes a designable parameter: an agent's value frameworks, communicative profiles, and negotiation strategies can each be calibrated to specific cultural expectations, turning what is a fixed background characteristic in human-human interaction into an explicit design choice. This design choice carries immediate consequences, because AI collaboration tools now serve users across dozens of countries whose communicative norms, social orientations, and cooperative expectations differ, and these tools increasingly mediate high-stakes interactions, including commercial negotiation, educational coaching, and team decision-making, that cross national and linguistic boundaries. If the communicative processes through which human-AI teams coordinate vary with cultural background, as decades of research on human cross-cultural communication would predict [11, 47], then parameterizing culture incorrectly, or not at all, risks serving some users well while failing others. Understanding how culture shapes these processes is, we argue, a prerequisite for equitable design rather than a secondary generalization question. The cross-cultural evidence needed to ground this understanding, however, remains limited, as existing studies seldom examine human-AI interaction across diverse populations.

Even where human-AI collaboration has been studied across populations, research has focused on tools and interfaces for interaction, and the communicative processes through which teams actually coordinate, such as accommodation and grounding, remain largely unexamined as design

targets. When the partner is an AI system, these processes are not yet well enough understood to inform design, and the design choices they depend on remain largely unexamined: how many AI partners participate, how their communicative behavior is configured, and whether support adapts to the variation among users. This dissertation takes communicative process design as its central approach, treating the processes through which human-AI teams coordinate as design targets rather than byproducts of tool deployment.

1.2 Gaps in the Current Understanding

Treating communicative processes as the design target requires first understanding how they operate in human-AI teams, yet how humans and AI partners accommodate and ground is not well understood even within a single population, and how these dynamics vary across cultural backgrounds and communicative dimensions is largely unknown. Among the communicative processes that productive collaboration involves, accommodation and grounding have well-established empirical foundations in human-human interaction and therefore offer the clearest entry point for characterization. The human-human baseline is well established: interlocutors adjust their communication toward one another not only in word use and style, through linguistic accommodation [41, 42], but also in the positivity or negativity conveyed through language, shaping how the interaction unfolds [60]. These accommodation dynamics, linguistic and affective, become especially consequential in negotiation, where culturally variable expectations and the need to coordinate on both content and expressed affect shape how parties reach agreement [11, 57]. Yet when one partner is an AI system whose communicative behavior is designed rather than emergent, it is unclear whether accommodation and grounding operate as they do between humans, or whether the interaction reduces to surface-level message exchange without the mutual adjustment that effective collaboration requires. Resolving this uncertainty requires characterizing how these processes actually behave in human-AI settings, along two dimensions. First, whether humans and AI agents accommodate linguistically, and whether that accommodation varies with cultural background, since a system calibrated to one

population may fail groups whose communicative norms differ. Second, whether accommodation extends beyond word use to the affective layer of interaction, since negotiation and other high-stakes settings depend not only on what partners say but on whether they arrive at shared understanding of the affect expressed, and whether this affective process is distinct from grounding. Together, the studies confirm that both processes are active in human-AI interaction but vary across cultural populations and communicative dimensions, meaning that a uniform design calibrated to one user group is unlikely to serve all. This variation raises a structural question: adding AI partners does not simply multiply dyadic adjustments but restructures who accommodates to whom and whether grounding even occurs across the group, so how does team composition reshape the dynamics that characterization reveals?

How team composition reshapes communicative dynamics in human-AI groups is largely unknown: existing productivity evidence comes from individual tasks or single-agent pairings, and the effects of team size and behavioral formation remain entangled with other design choices. The group brainstorming literature offers a clear precedent: in human-human teams, adding members introduces process losses such as evaluation apprehension, and free riding [27, 90], even as group diversity can enhance originality by exposing members to stimuli they would not encounter alone [129]. Whether these dynamics persist, transform, or disappear when the added members are AI partners is an open question, and the answer may depend on how the AI is configured: an AI set up to mimic human teammates might replicate human-human process losses, while an AI designed to overcome them might shift the team's communicative dynamics entirely. Initial evidence is suggestive but inconclusive: multi-agent systems for creative work report higher perceived quality and novelty [97], and design practitioners have begun orchestrating multiple AI agents into team formations [65]. These multi-agent designs, however, bundle agent count with agent roles, workflow structure, and interface changes, making it impossible to attribute their effects to any single design choice. Isolating team size from behavioral formation is therefore a prerequisite for understanding how composition shapes communicative dynamics, and doing so requires a controlled design that can separate what adding an AI partner contributes from how that partner is configured.

How a team is composed shapes what communicative dynamics emerge, opening a further design question: whether real-time coaching during the interaction can actively support what initial composition cannot, particularly for the diverse user populations whose variation the characterization stage reveals.

Negotiation is a setting where the need for such coaching is especially acute, because outcomes vary sharply with cultural background even when the task is identical, and AI agents are already engaging users in goal-driven, persuasive exchanges where these disparities carry real consequences. Cross-cultural negotiation research has documented these outcome gaps extensively [11], and AI-mediated transactions are growing rapidly [23]. Whether real-time coaching transfers across cultural populations or produces unequal outcomes of its own is unknown, because existing multi-agent coaching evaluations draw on single-cultural participant pools [111] and culturalization efforts that embed value frameworks into AI partners remain untested across regions [14, 63]. Evaluating this design intervention requires answering three tightly linked questions: whether coaching transfers across cultural populations, whether gains in message persuasiveness translate to negotiation outcomes such as donation amounts, and whether individual orientations such as cooperative disposition, rather than cultural group membership, moderate who benefits.

1.3 Contributions and Dissertation Overview

The goal of this dissertation is to articulate a communicative process design approach for human-AI teams: understanding how communicative processes are shaped by culture, team structure, and designed support, and identifying the conditions under which that support benefits diverse users. Four studies pursue this goal through three cumulative stages of this dissertation research (Figure 1.1), preceded by a theoretical review (Chapter 2) and closing with a general discussion (Chapter 7): characterization of the design material (Chapters 3 and 4), exploration of the design space through team formation (Chapter 5), and evaluation of a design intervention through real-time coaching (Chapter 6). Each stage depends on the one before it.

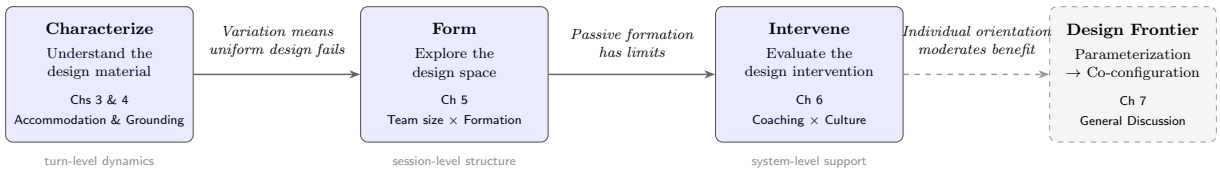


Figure 1.1: Idea flow of communicative process design in this dissertation.

The dissertation makes three contributions to the understanding of communicative processes in human-AI team collaboration. First, four studies spanning accommodation dynamics, team formation, and coaching intervention establish that communicative processes in human-AI teams are structured, variable across populations, and responsive to design. Investigating two task domains, three team configurations, and a culturally diverse participant pool, the studies reveal that population-level variation moderates who benefits from designed communicative support, suggesting that effective design must account for this variation rather than assume a universal user. Methodologically, the dissertation introduces measurement frameworks that separate constructs previously confounded in human-AI interaction research: human from AI authorship, affective accommodation from affective grounding within the same sample, global from local diversity, and proximal from distal outcomes. These separations make visible the trade-offs that aggregate metrics conceal, such as group output rising while human authorship falls. For design, the studies support a characterize-before-you-design principle: culturally aware systems should infer cooperative orientation from interaction rather than assign support by demographic category, and each team configuration must be evaluated as a separate design case because they produce distinct communicative dynamics.

The first two studies characterize the design material by establishing that communicative dynamics between humans and AI are measurable, culturally variable, and structured (Chapters 3 and 4). Chapter 3 analyzes linguistic accommodation across a large-scale dataset of user-agent conversations spanning dozens of countries and finds that accommodation varies by population: agents that accommodate well on average may fail specific cultural groups. Chapter 4 extends

the analysis into the affective dimension in negotiation dialogues, operationalizing both aspects of communicative dynamics within the same sample: Language-Mediated Affective Accommodation (LMAA), which captures turn-to-turn sentiment convergence, and Conversational Affective Grounding (CAG), which captures whether coordination becomes more efficient over time. Examining both aspects together provides a richer account of how communicative processes operate in human-AI interaction, though replication with larger samples is needed. Despite this limitation, the two studies converge on a clear pattern: communicative dynamics are active in human-AI interaction but vary across populations and dimensions, meaning that a uniform design calibrated to one user group is unlikely to serve all (Arrow 1 in Figure 1.1).

Exploring the design space in a controlled experiment, Chapter 5 shows that AI team size and behavioral formation interact to reshape communicative roles, and the interaction is asymmetric. The Chapter 5 study crosses team size (one vs. two AI partners) with behavioral formation (average-calibrated vs. user-parameterized) in a brainstorming task, with 168 participants and 24 human-human reference dyads. The central finding is an asymmetric interaction: adding a second AI raises group output but reduces human-authored contributions, and formation shapes this redistribution with one AI but not with two. The asymmetry arises because the two formations create qualitatively distinct communicative modes. Under average-calibrated formation (AVG), where the AI is calibrated to match the average human teammate's behavior, the AI co-generates alongside the human, with 58% of its messages classified as generative contributions. Under user-parameterized formation (CUS), by contrast, the AI shifts into a coaching profile (25% evaluative, 32% questioning), and humans respond with more generative content and more deliberative engagement (3.12 messages per board item compared to 1.91 under AVG). Beyond these role shifts, any AI partner increases global diversity of human ideas, but only a second AI increases local differentiation within the session. The CUS coaching mode demonstrates that formation alone can create structured communicative support, but it also reveals limits: passive formation cannot adapt to the user during the session (Arrow 2 in Figure 1.1), raising the question of whether active real-time coaching can build on what formation establishes.

Evaluating the design intervention, Chapter 6 finds that multi-agent coaching improves message persuasiveness, but whether that translates to task outcomes depends on individual orientation rather than cultural group membership. The Chapter 6 study tests a two-agent coaching system for cross-cultural negotiation with 80 participants in a 2×2 design crossing coaching condition with cultural alignment. Coaching improves message persuasiveness among US-born participants, where the gap is widest, but this proximal gain does not automatically translate into better negotiation outcomes. Whether coaching reaches the distal task outcome (donation amount) depends not on regional category but on horizontal collectivism, an individual-level cooperative orientation measured before the task. This moderation confirms that individual orientation, not cultural group membership, determines who benefits from designed communicative support (Arrow 3 in Figure 1.1). Separately, only one of five AI reasoning components (Decision Rationale) is associated with subsequent human elaboration, and this coupling is stable in both cohorts, while static cultural-profile alignment does not moderate any tested relation.

The lesson that emerges across all three stages is that the communicative process design approach cannot assume a universal user: the population variation identified in characterization returns as the moderation that determines who benefits from intervention, confirming the characterize-form-intervene dependency that Figure 1.1 summarizes. Chapter 7 synthesizes this convergent principle and develops the design frontier it implies. All studies in this dissertation set the AI's behavior before the session begins (initial parameterization), but the pattern of population-dependent moderation suggests that effective support may require adjusting communicative behavior through ongoing interaction (iterative co-configuration), where the system adapts in response to the user's communicative behavior during the session.

Chapter 2

Theoretical Background and Related Work

This dissertation's goal is articulating a communicative process design approach for human-AI teams: understanding how communicative processes are shaped by culture, team structure, and designed support, and identifying the conditions under which that support benefits diverse users. This chapter supplies the theoretical foundations that make that investigation possible. It first reviews the human-human research literature on accommodation and grounding as complementary aspects of communicative process (Section 2.1), so that Chapters 3 and 4 can ask whether those processes extend to human-AI interaction and, if so, what form they take. It then reviews how cultural variation shapes communication styles and social orientations across populations (Section 2.2) and how the dissertation's two collaborative task domains, group brainstorming and negotiation, place distinct demands on teams (Section 2.3), providing the vocabulary and the transfer questions that Chapters 3 through 6 pursue. Finally, it turns to human-AI communication and collaboration itself (Section 2.4), tracing the shift from individual AI tools to multi-agent teammates and coaching systems, and identifying the gaps that the characterize-form-intervene progression is designed to close.

Characterizing communicative dynamics in human-AI interaction (Chapters 3 and 4) requires first understanding how those dynamics operate between humans, where accommodation and grounding have been studied for decades.

2.1 Communicative Accommodation and Grounding

Interlocutors adjust their communication styles toward one another, and mutual understanding requires active coordination beyond mere message exchange. Research on linguistic accommodation has established when and why speakers converge or diverge in communicative style [41], while grounding theory has formalized the collaborative process through which common ground accumulates [20]. Together, these two bodies of work provide the analytical vocabulary this dissertation uses to characterize communicative dynamics in human-AI interaction, capturing complementary aspects of the same process: accommodation describes whether partners converge on communicative style, while grounding describes whether they build mutual understanding. Chapter 4 later operationalizes both aspects within the same negotiation dialogues, measuring affective accommodation and affective grounding through their respective indicators.

2.1.1 Communication Accommodation Theory (CAT)

Communicative adjustments during interaction serve social functions beyond information exchange. Communication Accommodation Theory explains when, how, and why interlocutors adjust their behavior toward or away from one another [41]. Convergence, in which speakers adjust toward their partner's communicative style, signals social approval and facilitates rapport; divergence, in which speakers accentuate differences, asserts group identity or social distance. The theory evolved from an initial focus on speech style into a broader communication framework encompassing convergence, divergence, overaccommodation, and underaccommodation across verbal and nonverbal channels [41].

In text-based and computer-mediated settings, accommodation operates primarily through linguistic style, since nonverbal cues are unavailable. Linguistic style matching, measured through function words, sentence structures, and discourse markers, provides a measurable trace of dyadic coordination in social interaction and writing [53, 81]. These traces reveal that accommodation effects in computer-mediated interaction are conditional rather than uniformly positive. For example,

virtual linguistic mimicry improves individual negotiation gains when it occurs early, with trust mediating the relationship [120], and some linguistic similarities track interpersonal trust while others do not [104]. The interpersonal effects of accommodation in instant messaging similarly depend on power dynamics and conversational norms [75]. Taken together, these findings establish that accommodation in computer-mediated communication is both measurable and consequential, raising the question of whether these communicative processes extend to human-AI interaction and, if so, how they are shaped by culture and team structure. Chapter 3 begins this investigation by characterizing linguistic accommodation between users and AI agents across diverse cultural populations.

Accommodation explains how partners converge on communicative style, but a full account of communicative dynamics also requires understanding whether that convergence builds mutual understanding. The complementary process through which mutual understanding is achieved has its own theoretical tradition: grounding.

2.1.2 Grounding in Communication

Establishing mutual understanding in conversation is a collaborative process requiring active coordination, not merely message exchange. Grounding is the collective activity through which interlocutors accumulate common ground, because common ground cannot be properly updated without explicit coordination between contributors and partners [20]. The process has two phases: a contributor presents information, and a partner provides evidence of understanding through verbal acknowledgments, relevant next turns, or other uptake signals. As these cycles repeat, common ground accumulates, and the declining effort over repeated references demonstrates that grounding leaves a measurable trace [21]. This cumulative character is what distinguishes grounding from the turn-level convergence that accommodation captures, and it is what makes grounding trackable across an entire interaction.

That cumulative process is not uniform across settings: grounding costs vary systematically with communication medium, shaping how coordination unfolds. Eight medium constraints, in-

cluding copresence, visibility, audibility, cotemporality, and reviewability, determine which grounding techniques are available and what effort they require [20]. Text-based media lack visibility and audibility, forcing interlocutors to rely on verbal evidence alone, which means that all coordination effort is captured in the conversation transcript. Chapter 4 exploits this property: because its negotiation dialogues are text-based, grounding effort is observable through turn-level indicators such as message length and affective distance.

2.1.3 Accommodation and Grounding as Complementary Aspects of Communication

Accommodation and grounding describe complementary aspects of what happens when people communicate. Accommodation captures the convergence dimension: whether partners adjust their communicative style toward one another. Grounding captures the coordination dimension: whether partners build mutual understanding that makes later contributions more efficient [20, 41]. Chapter 4 operationalizes both within the same negotiation dialogues as Language-Mediated Affective Accommodation (LMAA) and Conversational Affective Grounding (CAG), showing that affective alignment and coordination efficiency can each be tracked through turn-level indicators.

These communicative processes do not unfold in a cultural vacuum: because population-level variation moderates who benefits from designed communicative support, the human-human literature on how cultures differ in communicative norms and individual orientations supplies the vocabulary the dissertation needs next.

2.2 Cultural Variation in Communication

The communicative processes reviewed in Section 2.1 do not operate uniformly across cultural contexts: in human interaction, cultural background shapes whether collaborative support helps or hinders [11], and the same may hold when one interlocutor is an AI agent. Cultural variation in communicative norms matters for this dissertation because the question of whether human-human

communicative processes extend to human-AI interaction cannot be answered with a single population. Accommodation patterns documented in one cultural context may not hold in another, and communicative expectations around rapport, directness, and cooperative behavior vary systematically across populations [47, 49]. Before asking whether accommodation and grounding extend to human-AI interaction, it is necessary to understand how cultural context shapes those processes between humans. Empirical work that examines human-AI communication across culturally diverse populations remains scarce, and the human-human foundations reviewed here provide the necessary starting point.

Cultures differ systematically in communication styles, social orientations, and collaborative norms, and these differences are measurable at the population level. Culture itself can be understood as a collective pattern of values, beliefs, and behavioral norms that distinguishes one population from another, with national populations varying along dimensions including power distance, individualism-collectivism, uncertainty avoidance, and long-term orientation [49]. These value patterns can shift with modernization while also retaining durable traces of cultural heritage [52]. Population-level variation in social orientation is therefore real, measurable, and relevant to interaction, though these frameworks describe societies, not individuals.

Communication styles also vary along the high-context/low-context dimension, shaping how meaning is encoded and interpreted across cultural groups. In high-context communication, much of the information is carried by the surrounding context and relationship; in low-context communication, more information is carried by explicit code [47]. When interlocutors from different positions on this continuum collaborate, style divergence creates comprehension and coordination burdens. In intercultural brainstorming, for example, the tension between the conceptual diversity that cross-cultural teams can access and the communication barriers that diversity introduces is well documented: stimulation support helps participants use diverse knowledge, but without that support, coordination costs can outweigh diversity benefits [129].

Individual-level cultural orientation varies within national groups and can be situationally activated, limiting the sufficiency of country-level categories for predicting individual communicative

behavior. Self-construal theory establishes that cultures differ in how the self is defined in relation to others [68]. Horizontal and vertical collectivism then provide individual-level measurement distinctions between egalitarian and hierarchical forms of group orientation [114, 121]. Self-construal is dispositional yet also situationally activated [39, 69]. Understanding communicative processes in human-AI interaction across cultures therefore requires attention to both population-level patterns and individual-level orientations, because the two capture different aspects of how cultural variation shapes communication.

Cultural frameworks clarify why communicative norms differ across populations, and for collaborative tasks those differences shape which communicative processes matter most. Generative creative work and mixed-motive negotiation place distinct demands on teams, so each requires its own literature.

2.3 Collaborative Task Domains

Collaborative tasks vary along dimensions that determine which communicative processes dominate and how team members coordinate. McGrath's group-task circumplex [61, 71] organizes tasks along a cooperative-competitive axis: brainstorming sits among cooperative generative tasks, where the primary communicative challenge is overcoming process losses that limit idea stimulation, while negotiation sits among mixed-motive tasks, where the challenge is strategic persuasion and trust-building across different interests. Task type changes both what groups can achieve and which interaction processes matter, because even small shifts in task definition can influence group effectiveness [61]. This distinction matters for communicative process design because designed support that addresses process losses in brainstorming may operate through entirely different mechanisms than coaching for cross-cultural negotiation.

2.3.1 Group Brainstorming and Process Losses

Group brainstorming produces fewer ideas than the same number of individuals working alone, a finding replicated across decades of comparisons between nominal and real groups [27, 76]. The gap persists because interactive brainstorming introduces identifiable process losses alongside its stimulation potential. Three are particularly relevant for this dissertation. Competitive orientation leads group members to prioritize advocating for their own ideas rather than building on what others contribute [25]. Evaluation apprehension causes members to withhold contributions for fear of negative judgment from the group [6, 90]. Social loafing leads individuals to reduce effort when their contributions are less identifiable within the collective output [78, 90]. Social influence compounds the problem more broadly: group members match one another's performance levels, converging toward lower output, though providing an appropriate performance standard can close the gap [90].

Communication patterns during brainstorming vary with group cultural composition. In cross-cultural triads, Chinese and American participants show different levels of talkativeness and responsiveness, and these style differences shift depending on whether the group is culturally homogeneous or mixed [127]. When groups brainstorm through text chat, the medium reduces some cultural differences in talkativeness, but does not eliminate them [127]. These findings establish that brainstorming outcomes depend on how group members communicate during the task, and that cultural composition shapes those communicative dynamics.

Technology-mediated stimulation can counteract process losses by exposing group members to material that activates knowledge they would not have retrieved alone. A system that selects pictorial stimuli based on the group's ongoing conversation raises both ideation productivity and the proportion of rare ideas in brainstorming dyads [128]. In culturally diverse groups, such external stimuli help convert the diversity of team members' backgrounds into broader and more original output [129]. The effectiveness of stimulation depends on conceptual distance: ideas that are different enough to open new associative paths increase originality and diversity, while ideas too

similar to the current discussion fixate attention on a narrow category [82]. These results position designed support as a mechanism that can shift the balance between process losses and stimulation in brainstorming groups.

Chapter 5 builds on this literature by distinguishing global diversity (how different are the team's ideas from other teams?) from local diversity (how different are ideas within the session?), measuring both as outcomes of AI team size and behavioral formation in brainstorming. Whether AI partners reproduce the process losses documented in human groups or reshape them is an open question, because the AI's behavior can be designed rather than left to emerge from group dynamics. The same tension between collaborative potential and communicative cost also structures negotiation, the task domain that Chapters 4 and 6 use as a setting.

2.3.2 Persuasion and Negotiation

Cross-cultural negotiation introduces systematic variation in communication strategies that can produce unequal outcomes. Culture shapes negotiators' interests, priorities, and strategies, including whether parties confront directly or indirectly and how they exchange information [11]. Persuasion itself depends on alignment between argument type and the receiver's values: appeals that resonate in one cultural context may underperform in another [110]. For the dissertation, negotiation provides a high-stakes collaborative setting where communicative process design directly affects outcomes, and where the stakes of cultural mismatch are consequential rather than cosmetic.

Negotiation is a persuasive and goal-directed process in which influence unfolds through structured dialogue. Core psychological principles of influence, including reciprocity, commitment, and scarcity, explain how negotiators shape one another's decisions [19]. When these principles are operationalized in conversational settings, specific persuasive strategies such as emotional appeals and personal inquiries predict compliance, and their effectiveness depends on matching the appeal to the receiver's psychological profile [130]. The vocabulary of persuasive strategies provides a framework for analyzing how appeals, inquiries, and interpersonal tailoring shape multi-turn communicative exchange.

Cultural variation shapes negotiation at every level, from the interests and priorities negotiators bring to the communication styles they use. Low-context communication places more information in explicit code, while high-context communication requires interlocutors to infer meaning from surrounding context and relationship [47]. When negotiators from different positions on this continuum interact, implicit expectations about directness, information disclosure, and proposal framing can create misalignment that compounds across rounds of dialogue. This cultural variation in negotiation communication is why Chapter 6 uses cross-cultural negotiation as a setting: it is a domain where communicative differences are consequential and where the effects of designed support may vary across populations.

The individual-level cultural variation reviewed in Section 2.2 extends directly into negotiation. Horizontal collectivism, an egalitarian cooperative orientation that varies within national populations, shapes how individuals approach collaborative bargaining [110, 114]. Because cooperative orientation varies among individuals who share a regional background, it offers a finer-grained predictor of negotiation behavior than country-level cultural assignment alone. Chapter 6 tests whether this individual-level construct moderates the effectiveness of AI coaching in cross-cultural negotiation.

Taken together, these human-human literatures define the communicative processes, cultural variation, and task domains the dissertation will later redesign, but they cannot say whether the same dynamics hold once one partner is an AI system whose behavior is designed rather than emergent.

2.4 Human-AI Communication and Collaboration

Once AI moves from personal tool to team member, the communicative processes reviewed in Sections 2.1 through 2.3 raise new questions. Understanding how those processes unfold in human-AI interaction, how culture and team structure shape them, and how designed support can improve them is the focus of the communicative process design approach this dissertation pursues.

2.4.1 Individual AI Productivity and the Shift to Teams

AI assistants deliver substantial individual productivity gains, creating pressure to extend AI from personal tools to team members. Writing tools reduce task completion time by 40% while raising output quality by 18% [84], and customer-support agents resolve 14% more issues per hour, with novices benefiting disproportionately [13]. AI systems are being developed as partners and coaches in negotiation [111], deliberate-practice education [2], and creative work [65, 97]. These successes motivate the shift from individual AI tools to collaborative AI teammates, but aggregate output gains are an insufficient evaluation frame for understanding what that shift introduces.

When AI moves from tool to team member, it introduces communicative dynamics that do not simply accelerate human teamwork. Teams with AI partners produce higher output but more homogeneous results: in a large-scale advertising experiment, human-AI teams produced more ads per worker but lower image quality and greater stylistic uniformity than human-human teams [56]. Passive exposure to AI-generated ideas increases the collective diversity of a population's ideas without changing any individual's creativity [3]. On the process side, working with AI code assistants demands continuous monitoring and steering [95], and participants in brainstorming value AI contributions but resist when the agent dominates the conversation [51]. In a single-agent brainstorming study, the AI initiated the majority of all ideas, many went undiscussed, and AI-initiated ideas were less likely to become final design elements [77]. These findings suggest that the shift from AI tools to AI teammates changes communicative dynamics in ways that existing designs have not isolated: multi-agent systems confound agent count with roles, workflow, and interface, making it difficult to attribute effects to any single design choice.

Whether the communicative processes documented in human-human interaction extend to human-AI interaction remains an open question. Communication Accommodation Theory extends from interpersonal and intergroup settings into AI-mediated communication, where convergence and divergence can be measured in the same linguistic dimensions that the human-human literature established. Culturalization approaches that embed value frameworks into AI agents [14, 63]

demonstrate that cultural sensitivity can be designed into systems, yet these approaches rely on synthetic training data or scripted dialogues rather than observed cross-cultural interaction, and their generalization to real users remains untested. Whether accommodation varies by cultural context in human-AI settings, and whether grounding accompanies stylistic convergence, have not been examined.

2.4.2 AI-Mediated Coaching and Communicative Support

AI coaching systems provide structured support for complex collaborative tasks, but existing evaluations have not tested whether coaching transfers across culturally diverse populations. Negotiation coaching systems that deliver targeted feedback between trials have been evaluated with US native-English speakers only [111], and structured conversational scaffolds for AI agents have been evaluated with specific populations such as children [33], but neither approach has been tested across cultural groups. Whether coaching benefits transfer across populations or depend on individual orientations rather than cultural group membership is an open question.

The literatures reviewed in this chapter define the communicative processes, cultural variation, task domains, and human-AI dynamics that the dissertation investigates, but they leave open whether those dynamics hold, transform, or require redesign once AI is a teammate rather than a tool. The chapters that follow address these gaps through the characterize-form-intervene progression introduced in Chapter 1.

Chapter 3

Characterization: Linguistic Accommodation in Human-AI Interaction Across Cultures

Characterizing how culture shapes communicative processes is the first step in the communicative process design approach this dissertation develops. Accommodation, the process through which interlocutors adjust their linguistic styles toward one another, varies in direction and consequence across cultural contexts in human interaction. Whether AI systems participate in this process, and whether the cultural variation persists when one interlocutor is an AI agent, has not been systematically addressed. This chapter provides the first large-scale characterization of linguistic accommodation between humans and AI across cultures.

Accommodation in human-AI conversation is reciprocal, culturally variable, and convergent over time. Both agents and users adjust their linguistic styles toward one another, mirroring convergence dynamics observed in human conversation. One might expect this reciprocity to serve all users equally, but it does not: agents that accommodate well on average fail specific cultural groups, and the direction of adaptation differs with cultural dimensions such as power distance and individualism. For the communicative process design approach, these patterns establish that communicative processes in human-AI interaction are structured and variable across populations, the first property that must be in place before design can proceed. Whether the same variation

holds in the affective dimension is the question the next chapter takes up.[1]

3.1 Introduction

Conversational agents powered by large language models (LLMs), such as ChatGPT, are reshaping how humans interact with computers, necessitating a comprehensive understanding of these emerging dynamics. Recent work suggests that conversational agents that emulate the communication style of their users can significantly boost user engagement [123]. This adaptation process, in which one party adjusts their communication style to more closely resemble that of their interlocutor, is often referred to as style matching.

Style matching is grounded in sociolinguistic observations of how human interlocutors adapt their communication styles (e.g., word choices, tones, utterance compositions), Communication Accommodation Theory (CAT) [41] theorizes that, driven by an intrinsic drive for social (dis-)integration and signaling, people adapt their communication styles to (dis-)align with their interlocutors. Drawing on research on style matching in human interactions [124], conversational agents that adapt variables such as message length, tone, cognitive style, or word choice to mirror users' communication patterns can enhance engagement and foster more positive attitudes [74, 101]. Moreover, cultural norms shape adaptation behaviors; for instance, individuals from collectivist cultures may adapt more in diverse group settings, especially with enhanced visibility of interlocutors [126].

Despite this understanding, the dynamics of mutual adaptation and communication style convergence remain underexplored in interactions involving non-human actors, such as LLM-powered conversational agents, especially in cross-cultural contexts. LLMs have yet to fully integrate the semantic, textual, and pragmatic dimensions of communication [54], nor do they posit capability

[1] Part of the content in this chapter is published in the following papers: Rong-Ching Chang and Hao-Chuan Wang. "Communication Accommodation Between Large Language Models and Users Across Cultures (Student Abstract)." *The Thirty-Ninth Association for the Advancement of Artificial Intelligence Conference (AAAI)*, 2025; and Rong-Ching Chang and Hao-Chuan Wang. "Convergence, Reciprocity, and Asymmetry: Communication Accommodation Between Large Language Models and Users Across Cultures." *Proceedings of the Annual Meeting of the Cognitive Science Society (CogSci)*, 2025.

for self-explanation or consistently outperform humans [122]. These gaps raise questions about whether and how conversational agents and users mutually adapt their communication styles, especially given that cultural backgrounds shape how humans express themselves and engage in conversations. This study addresses this gap by investigating how cultural contexts influence communication dynamics between humans and conversational agents, hypothesizing that style matching behaviors observed in human-to-human interactions will also manifest, with LLMs adapting to diverse cultural contexts.

Leveraging Hofstede's cultural framework [50], which describes the cultural dimensions of different societies such as individualism and power distance, this study examines how these dimensions shape communication adaptation. We probe communication accommodation and adaptation through the lens of linguistic features using tools like the Linguistic Inquiry and Word Count (LIWC) [92] and the Language Style Matching (LSM) [53]. These methods allow us to assess the subtle, yet crucial, dynamic interplay between user and conversational agent. This study addresses three core research questions that will improve our understanding of how people conversationally interact with LLMs, as well as the dynamic properties of LLMs when situated in interactions with humans:

RQ1: Do national cultures of different placements along Hofstede's cultural dimensions influence linguistic style-matching between users and LLM-powered conversational agents?

RQ2: Does communication accommodation between users and conversational agents increase as the conversation is prolonged?

RQ3: Do users adapt their communication styles to the conversational agents and vice versa?

We analyzed 8,011 user-agent conversations from 1,500 participants, mapping users' birth countries to Hofstede's dimensions and evaluating style matching. Our findings reveal that style matching in linguistic markers related to necessity and obligation is more pronounced in both user and agent responses, particularly in interactions with users from high-hierarchy cultures compared to those from low-hierarchy cultures. Moreover, style matching increases over time, mirroring convergence patterns observed in human-to-human interactions, with further analysis revealing that

LLMs excel in aligning with clear linguistic patterns but face challenges in adapting to psychological and contextual nuances. We also observed asymmetry in adaptation, with LLMs adapting more to specific cultural contexts and users reciprocating in some cases.

These results highlight the potential of devising single LLMs to adapt to diverse cultural backgrounds while improving alignment through prolonged interactions. By integrating these insights, this study advances the development of culturally inclusive AI systems and fosters more effective accommodations for the users.

3.2 Related Work

3.2.1 Communication Accommodation Between Users and Conversational Agents

Previous research has demonstrated that users tend to prefer agents that adapt to their communication styles [74, 101]. For instance, customer engagement increases when the chatbots exhibit similar characteristics and lifestyle congruency with the user [67]. Additionally, integrating linguistic alignment—where chatbots mimic users' word choices, sentence structures, and language style—has been shown to decrease user frustration and perceived task workload, thus improving overall interaction quality [117].

3.2.2 Direction of Communication Accommodation

Prior work has found asymmetry in communication adaptation during cross-cultural interactions [64, 126, 127]. For instance, studies show native speakers often accommodate non-native speakers in conversation but not vice versa [64]. Similarly, mixed-culture settings see greater adaptation efforts by individuals from collectivist cultures [127].

While previous studies have explored aspects of communication accommodation, the specific variations that arise when users from diverse cultural backgrounds interact with conversational agents remain underexplored. Recognizing that such adaptations are typically asymmetrical in cross-cultural interactions, our study aims to fill this gap by analyzing conversations between agents

and users from various global contexts. We specifically examine how a user's cultural background and the duration of interaction influence communication adaptations, as well as the directionality of these adaptations between users and conversational agents.

3.3 Dataset

We use PRISM and apply Hofstede's cultural dimensions framework for our study.

PRISM [59] is an open-sourced dataset consisting of 8,011 conversations between 1,500 participants from 75 countries and large language models such as GPT-4 and Claude-2. It comprises 27,172 interactions where human messages are met with model responses. The participants disclosed their sociodemographics, such as age, gender, and birth country, making this dataset uniquely suitable for the aim of our research.

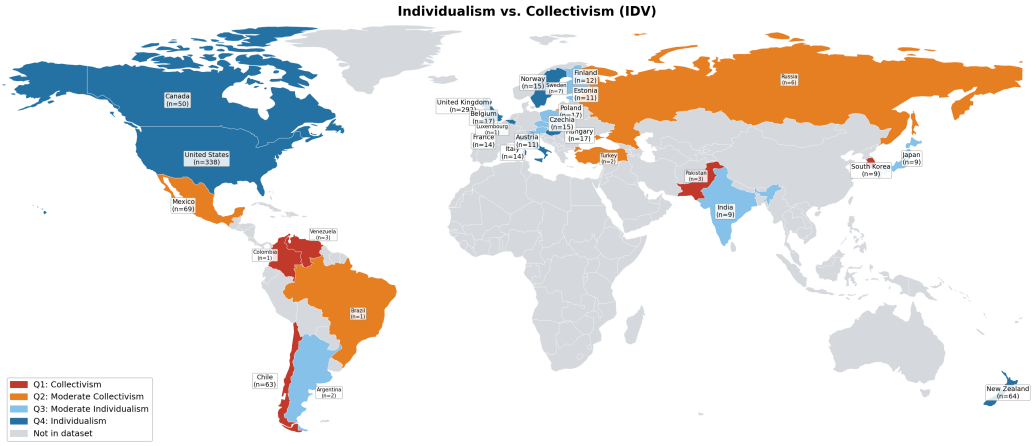
Figures 3.1 and 3.2 show the geographic distribution of participants across Hofstede's six cultural dimensions, including which countries compose each quantile. For Individualism vs. Collectivism, the collectivist quantile (IDV=1) draws primarily from Chile and South Korea, while the individualist quantile (IDV=4) is concentrated in the United States and United Kingdom. For Power Distance, the high-PDI quantile is predominantly Mexico, while low-PDI participants are concentrated in Anglo countries.

Hofstede's cultural dimensions [50] offer a widely applied framework for understanding cultural variations at the level of national cultures essential for our analysis:

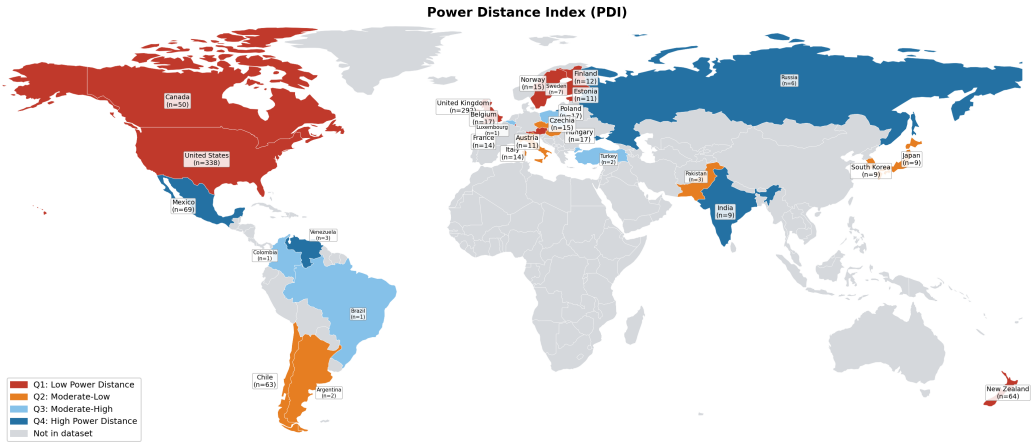
- Power Distance Index (PDI) indicates how societies accept hierarchical order without need for justification. Higher scores reflect acceptance of inequality and centralized power, while lower scores favor egalitarianism.

- Individualism versus Collectivism (IDV) shows whether societies emphasize personal achievement (high IDV) or collective relationships (low IDV).

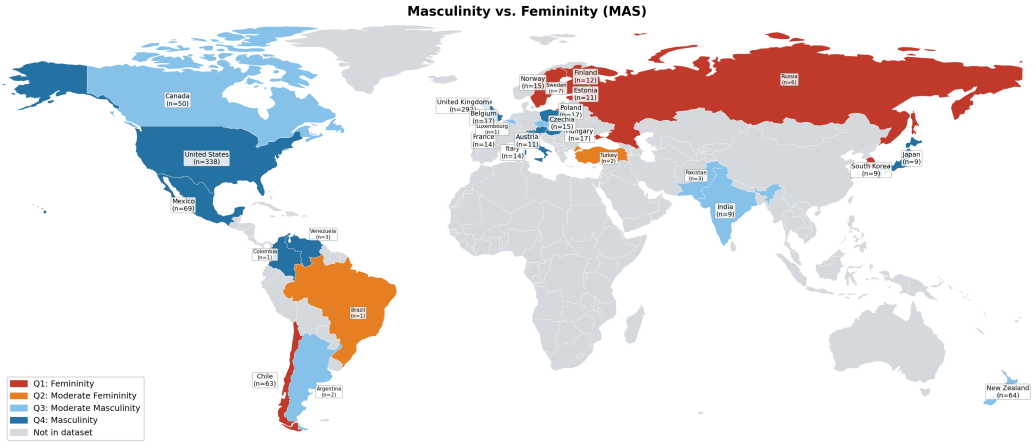
- Masculinity versus Femininity (MAS) depicts the preference for achievement and rewards for success (higher MAS) versus cooperation and quality of life in society.



(a) Individualism vs. Collectivism (IDV), N=998 users, 28 countries.

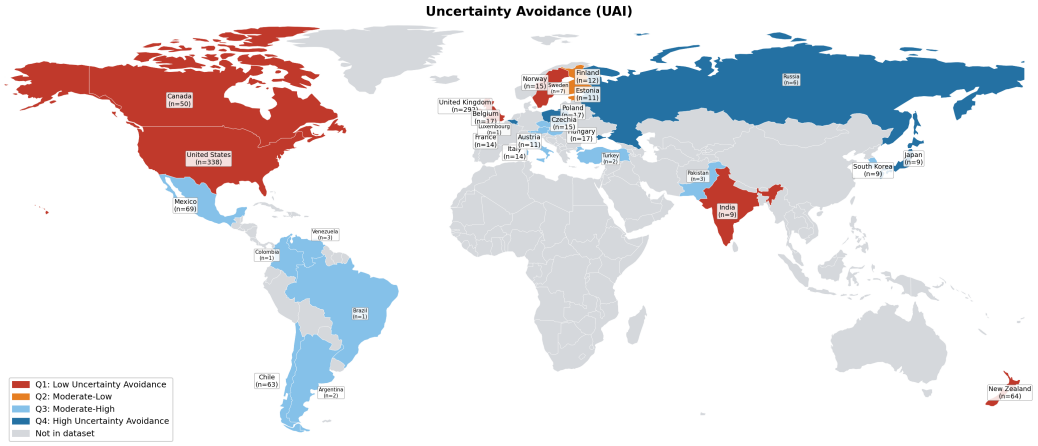


(b) Power Distance Index (PDI).

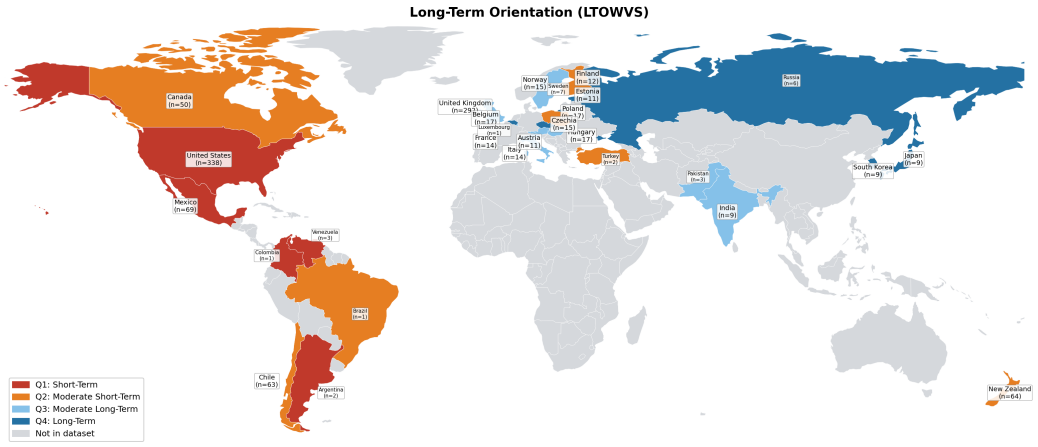


(c) Masculinity vs. Femininity (MAS).

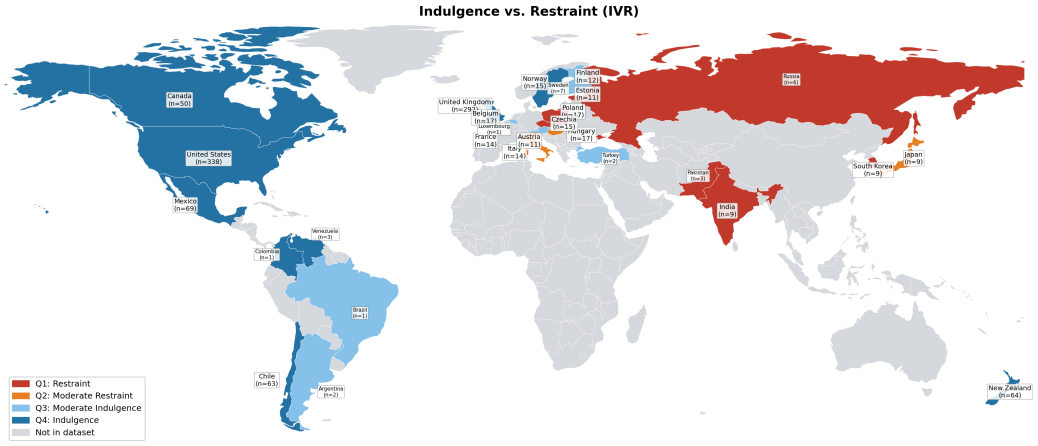
Figure 3.1: Geographic distribution of participants across Hofstede's cultural dimensions (1 of 2). Each map shows country-level quantile assignment and sample size.



(a) Uncertainty Avoidance (UAI).



(b) Long-Term Orientation (LTOWVS).



(c) Indulgence vs. Restraint (IVR).

Figure 3.2: Geographic distribution of participants across Hofstede's cultural dimensions (2 of 2).

- Uncertainty Avoidance Index (UAI) measures the degree to which a society tolerates uncertainty and ambiguity. High UAI indicates conservative, whereas low UAI indicates more inclusive.

- Long Versus Short-Term Orientation (LTOWVS) contrasts societies that plan for the future (higher LTOWVS) with those that focus on short-term outcomes and maintaining traditions.

- Indulgence versus Restraint (IVR) measures the degree of freedom societies permit in fulfilling human desires in having fun (high IVR) versus pessimism.

Table 3.1: Example LIWC Categories and Sample Words

Category	**Example Words**
Cognitive Processes	but, not, if, or, know
Social Behavior	said, love, say, care
Emotion	good, love, happy, hope

Linguistic Inquiry and Word Count (LIWC) [8] is a comprehensive text analysis tool that categorizes over 12,000 words and emoticons into over 100 predefined categories as shown in Table 3.1. These LIWC categories are further clustered as *category groups* to characterize key dimensions salient in language use, such as Psychological Process, Perception and Linguistic Dimensions. Designed and validated to assess linguistic, cognitive, and emotional properties of text, LIWC quantifies the frequency of words associated with different psychological states and processes. By providing percentage-based breakdowns across pre-constructed and validated categories, LIWC offers insights into emotional states, cognitive styles, and social dynamics reflected in language.

3.4 Method

To assess communication accommodation between users and conversational agents across diverse cultural dimensions, we used LIWC to analyze utterances from both parties, quantifying their linguistic, cognitive, and emotional dimensions. We then applied LSM to measure the degree of similarity in word usage within LIWC-defined categories, capturing how conversational agents align their linguistic styles with users. The LSM [53] for a specific LIWC linguistic category is computed

using the following formula:

$$\mathrm{LSM}_{i,t} = 1 - \frac{|U_{i,t} - C_{i,t}|}{U_{i,t} + C_{i,t} + 0.0001} \tag{3.1}$$

Here, $\mathrm{LSM}_{i,t}$ represents the LSM score for the $i-th$ LIWC linguistic category derived from the LIWC dictionary for a user at turn t. Specifically, $U_{i,t}$ denotes the proportion of words within the $i-th$ category used by a user at turn t, while $C_{i,t}$ represents the corresponding proportion used by the conversational agent at turn t. The constant 0.0001 in the denominator is included to avoid division by zero when neither party uses words from a given category. An LSM score is a normalized metric ranging from 0 (no similarity) to 1 (perfect alignment).

3.4.1 Impact of Cultural Dimensions on Communication Accommodation

To assess how cultural dimensions influenced style matching between users and conversational agents, we evaluated LSM differences as a function of cultural dimensions using linear mixed model analysis. Using Hofstede's framework, countries were categorized into four quantiles (e.g., Individualism is $IDV = 4$, Collectivism is $IDV = 1$). The model integrates the cultural dimensions to predict LSM for each conversation i by user j as follows:

$$\begin{aligned} \mathrm{LSM}_{i,j} = \beta_0 + \beta_1 \mathrm{PDI}_{i,j} + \beta_2 \mathrm{IDV}_{i,j} + \beta_3 \mathrm{MAS}_{i,j} \\ + \beta_4 \mathrm{UAI}_{i,j} + \beta_5 \mathrm{LTOWVS}_{i,j} + \beta_6 \mathrm{IVR}_{i,j} + \epsilon_{i,j} \end{aligned} \tag{3.2}$$

In this model, $\mathrm{LSM}_{i,j}$ denotes an array of 112 LSM scores, each derived from a distinct LIWC category, calculated for each conversation i between user j and the conversational agent. Specifically, $\mathrm{LSM}_{i,j}$ is aggregated over all $\mathrm{T}_{i,j}$ turns in conversation i with user j by $\mathrm{LSM}_{i,j} = \frac{1}{T_{i,j}} \sum_{t=1}^{T_{i,j}} \mathrm{LSM}_{i,t}$. The fixed effect coefficients β_0 through β_6 quantify the average influence of each cultural dimension on these LSM scores. The term $\epsilon_{i,j}$ captures the random error, representing variability in LSM that is not explained by the cultural dimensions. Given the large number of repeated tests (112 LSM measures), we applied the Bonferroni correction [29] to adjust the significance level

for multiple comparisons, ensuring robustness against Type I errors.

3.4.2 Impact of Conversation Length on Communication Accommodation

To investigate the relationship between communication accommodation and conversation length, we aimed to determine whether LSM increased, decreased, or remained stable as the interaction lengthened, thereby indicating potential convergence or divergence in communication styles over time. We quantitatively analyzed the progression of interactions throughout a conversation by introducing Turn_{ij}, a variable that tracks turn-taking patterns at each stage of the dialogue within equation 3.2. This allows us to directly measure the impact of prolonged engagement on LSM in our extended model.

To provide stringent evidence and contextualize the findings from the mixed-effects models, we calculated an aggregated LSM by averaging all LSM measures across the dataset and applying the same mixed-effects model. The coefficient of $\text{Turn}_{i,j,t}$ from this aggregated LSM model served as a benchmark, which was compared against individual LSM coefficients using a z-test to assess whether conversation length effects on each variable deviated from the average. To account for multiple comparisons, we applied a Bonferroni correction to the p-values obtained from the z-tests. The adjusted p-values were then used to determine the statistical significance of the differences. For significant differences after correction, we examined the magnitude of the effects to understand how communication styles evolved during extended interactions. If an individual coefficient exceeded the aggregated coefficient, this indicated style convergence, reflecting increased alignment in communication over time. Conversely, a smaller individual coefficient suggested style divergence, representing reduced alignment as the conversation progressed.

This approach provided a rigorous evaluation of how conversation length influenced communication accommodation across linguistic dimensions, identifying which aspects tended to converge or diverge during prolonged interactions.

3.4.3 Direction of Communication Accommodation

To understand whether users change their communication styles to adapt to agents or vice versa, we utilized the Granger causality test [44] which is a statistical test on two time series to determine the causal relationship. Specifically, it evaluates whether past values of one time series (e.g., LIWC categories from users) predict future values of another (e.g., LIWC categories from agents), or vice versa. If adding lagged values of one time series significantly improves the prediction of the other, Granger causality is established. For example, if past user behaviors predict agent behaviors, it indicates that agents adapt to users. Conversely, if past agent behaviors predict user behaviors, it suggests users accommodate the agents. Mutual causality occurs when both directions hold true. As Granger causality requires stationarity, we used the Augmented Dickey-Fuller (ADF) test [26] to verify and transform non-stationary time series before applying the analysis. A lag of 1 was used in the test.

This approach helps elucidate the dynamic interplay between users and conversational agents, revealing how communication accommodation unfolds in digital interactions.

3.5 Results

3.5.1 RQ1: Do national cultures of different placements along Hofstede's cultural dimensions influence linguistic style-matching between users and LLM-powered conversational agents?

To assess whether Hofstede's cultural dimensions influence style matching between users and LLM-powered conversational agents, we analyzed the linguistic similarity employed during these interactions. Style matching occurs when one party adjusts their communication style to more closely resemble that of the other party.

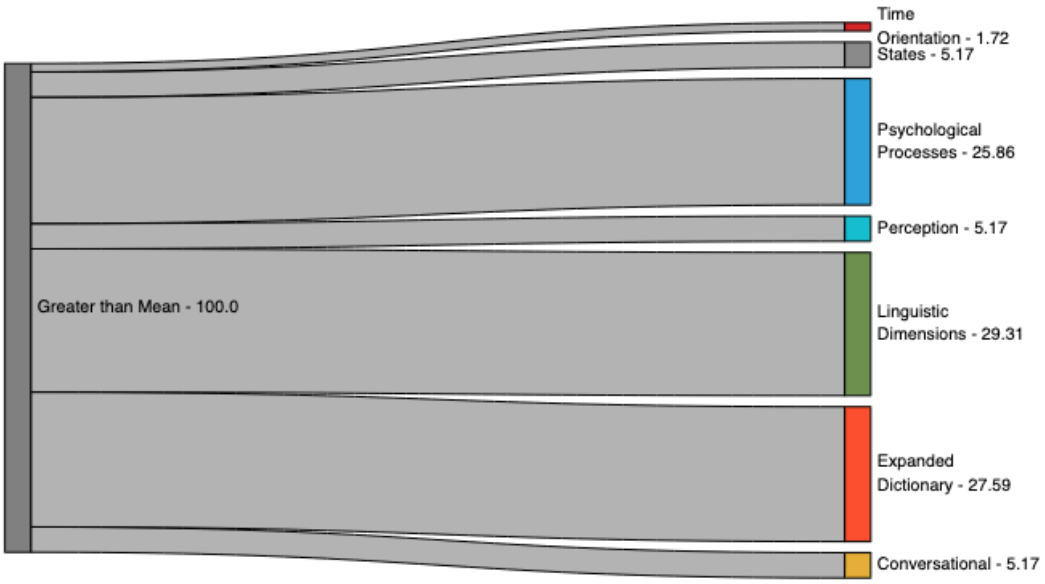
Our analysis reveals that the Power Distance Index (PDI) significantly influences communication style matching, particularly in terms of expressions related to necessity and obligation such as

"have to", "need", "had to", and "must". Statistical results reveal a notable difference ($\beta_1 = 0.1575$, $p = 0.0003$) in the communication style matching of these terms between high (PDI_4) and low (PDI_1) PDI countries. Specifically, both users and conversational agents have higher similarity in communication styles related to need, necessity, and obligation in conversations with users from high PDI countries (PDI_4) compared to those from countries with lower PDI (PDI_1). This finding suggests that the hierarchical nature of high PDI societies influences communication patterns to a significant extent, reflecting a cultural predisposition towards more formally structured expressions of necessity and obligation. Other cultural dimensions did not exhibit statistically significant effects on style matching, underscoring the particular salience of PDI in shaping these interactional dynamics.

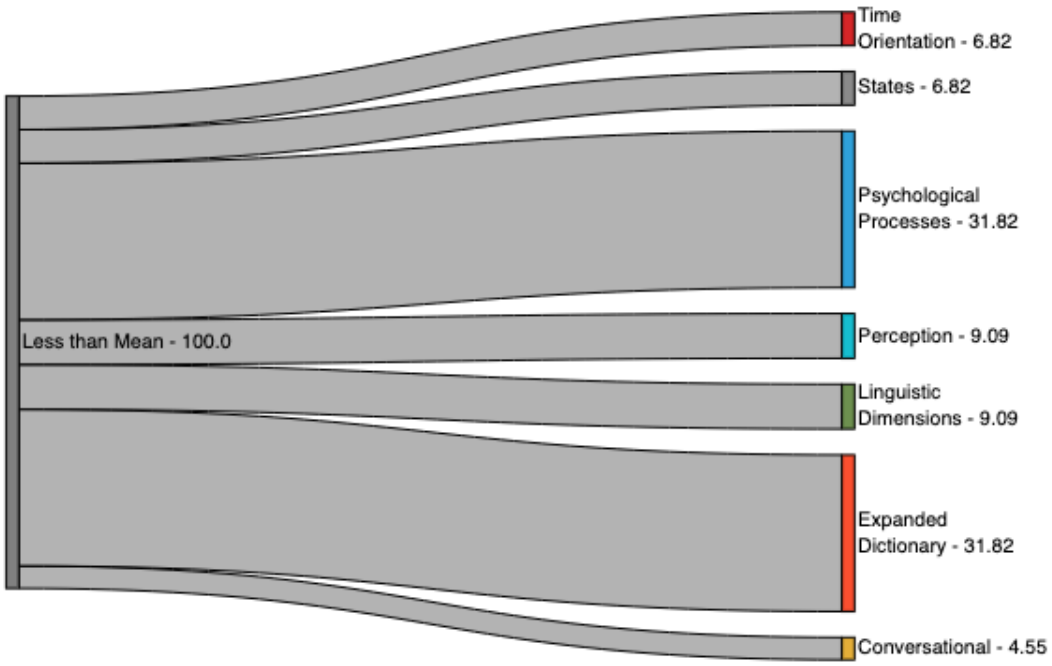
3.5.2 RQ2: Does communication accommodation between users and conversational agents increase as the conversation is prolonged?

Our analysis of communication accommodation examined whether prolonged interactions between users and conversational agents lead to increased linguistic alignment as indicated by LSM. Using a mixed-effects modeling approach combined with a Bonferroni-corrected z-test, we compared the effect of conversation length (Turn_{ij}) on each LSM category against the overall effect derived from an aggregated LSM model.

Our results support the hypothesis that LLMs and human users tend to converge in conversational styles over the course of interactions, mirroring the dynamics of human conversations that accommodate and adapt to each other's communication styles. Overall, we have 98 LSM with significant adjusted p-value ($p <.00044$). The analysis revealed that in 58 out of 98 significant LSM categories, the effect of conversation length was greater than the overall mean effect, indicating a more pronounced increase in LSM as conversations lengthen. In other words, nearly 59% of the significant LSM categories demonstrated convergence, suggesting that as conversations progress, the utterances between human users and LLMs become increasingly similar, surpassing the expected



(a) Proportion of *category groups* among LIWC categories that converge.



(b) Proportion of *category groups* among LIWC categories that diverge.

Figure 3.3: Relative presence of LIWC *category groups* showing convergence versus divergence in prolonged conversations.

baseline of linguistic similarity. This finding highlights the ability of LLMs to adapt dynamically to users' communication styles over time, further reinforcing their potential to emulate human-like conversational patterns. In contrast, 41% of the LSM categories (40 out of 98) exhibited divergence, where the effect of conversation length was weaker than the overall mean effect as the conversation progressed.

To further understand which LIWC *category groups* tend to converge or diverge more throughout the conversational process, we calculated the percentage differences in the presence of different *category groups* among the overall converging versus diverging categories (Figure 3.3). The results revealed that the *category group* of Linguistic Dimensions (+20.22%) exhibited the greatest positive percentage difference in favor of convergence. Specifically, 29.31% of the 58 converging LIWC categories are classified under Linguistic Dimensions, whereas only 9.09% of the 40 diverging LIWC categories fall into this category. The positive differences suggest that conversations are more likely to align in linguistic aspects such as function words, pronouns, determiners, auxiliary verbs, and other grammatical categories, reflecting the LLM's strong capability in adapting to linguistic patterns during dialogue. This trend is likely influenced by the LLM's extensive exposure to written texts in its training data, which emphasizes linguistic consistency and formal language structures over spontaneous conversational variations. In contrast, *category groups* including Psychological Processes (-5.96%) and Time Orientation (-5.10%) demonstrated negative differences, indicating that LLMs tend to struggle with the psychological and temporal aspects of language alignment, resulting in a greater tendency toward diverging from the users. This could be due to the dynamic and context-dependent nature of psychological and temporal language, which is harder to capture through static, text-based training data.

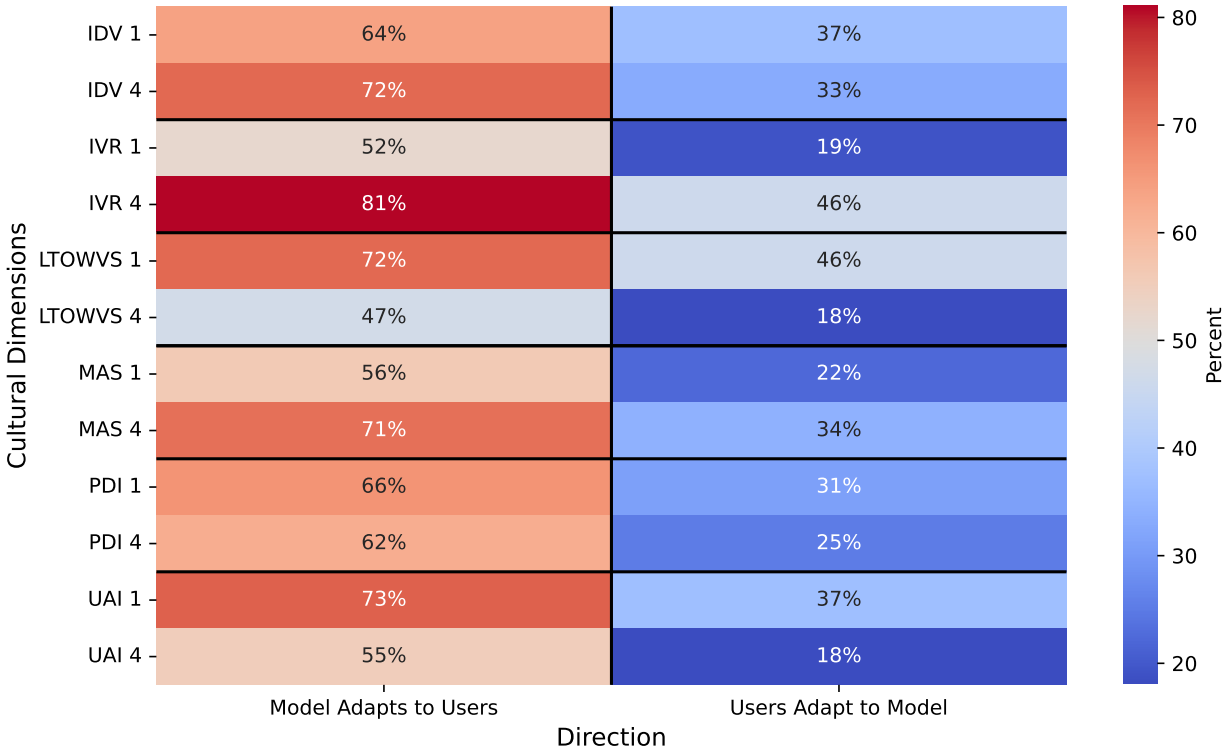


Figure 3.4: Heatmap of Communication Adaptation Across Hofstede's Cultural Dimensions. This heatmap quantifies the percentage of significant Granger causality findings for Individualism versus Collectivism (IDV), Indulgence versus Restraint (IVR), Long-Term versus Short-Term Orientation (LTOWVS), Masculinity versus Femininity (MAS), Power Distance Index (PDI), and Uncertainty Avoidance Index (UAI), highlighting a greater propensity for models to adapt to users than vice versa, thus demonstrating the conversational agents' adaptivity to users across cultural dimensions.

3.5.3 RQ3: Do users adapt their communication styles to the conversational agents and vice versa?

Building on our analyses of how cultural dimensions influence communication styles and the progressive alignment of these styles throughout interactions, we now focus on the directionality of communication adaptations between users and conversational agents, investigating through Granger causality tests whether adaptations are more pronounced from users towards the models or vice versa. We are particularly interested in how these adaptations differ between users from cultures with high versus low quantile levels (i.e., PDI = 4 versus PDI = 1) on Hofstede's cultural dimensions, providing insights into the dynamic interplay of culture in communication behaviors.

To quantitatively ascertain the direction and magnitude of these adaptations, we apply Granger causality analyses across over 100 LIWC categories, assessing the adaptation influence from both users to agents and from agents to users. We categorize and quantify the significance ($p < 0.05$) of these adaptation relationships for each cultural dimension, detailing both the direction of adaptation and the relative impact. Specifically, we compute the proportion of categories showing significant Granger causality, summarizing our findings in terms of which cultural dimensions exhibit the strongest influence on communication accommodation and its directionality. The findings predominantly suggest that models are more frequently adapting to users rather than the converse. However, it should be acknowledged that there is a discernible level of user adaptation towards models, as evidenced in Figure 3.4. To further understand the differences in adaptation behaviors, we conduct Chi-square tests to determine whether each adaptation direction—Model Adapts to Users and Users Adapt to Model—demonstrates significant differences between cultures with high versus low quantile levels.

Our analysis indicates that models demonstrate a greater tendency to adapt to users from cultures with higher levels of individualism (IDV_4=72%) compared to those from collectivist cultures (IDV_1=64%), with these differences being statistically significant ($p = 0.0144$). Conversely, the manner in which users adapt to models across individualistic and collectivistic cultures does not show significant variation ($p = 0.3151$), suggesting that the influence of individualism versus collectivism on user behavior towards models may be more nuanced or driven by other factors.

Users from cultures with lower power distance are more likely to adapt to the model (PDI_1=31%) than the users from more hierarchical cultures (PDI_4=25%); the difference is statistically significant ($p = 0.0358$). On the other hand, the differences in how models adapt to users do not differ significantly between different PDI levels ($p = 0.2100$).

In some cases, we observe a reciprocal relationship where the models adapt to the users more while the users also show higher adaptation. For instance, models adapt to users substantially more ($p < .001$) in optimistic cultures (IVR_4=81%) compared to pessimistic cultures (IVR_1=52%). At the same time, IVR_4 users also showed up to 46% adaptation to models compared to 19% from

IVR_1 users ($p < .001$). Similar reciprocal phenomena ($p < .001$) were also observed in short-term oriented cultures ($LTOWVS_1$), masculine cultures (MAS_4), and inclusive cultures (UAI_1).

These results underline the complex interplay between cultural dimensions and communication adaptation behaviors, occurring both from models to users and vice versa. The significant effects observed indicate that conversational agents can dynamically adjust their interactions in response to users' cultural backgrounds, as reflected in language use patterns. This finding further underscores the potential of leveraging natural language convergence processes as a foundation for developing culturally adaptive AI systems that more effectively align with human communicative behaviors and cultural nuances.

3.6 Discussion

This chapter, using Hofstede's framework, explored how cultural dimensions influence communication accommodation between users and conversational agents.

Our analyses revealed that cultural dimensions, particularly those related to social hierarchy, significantly impact style matching of users and conversational agents. Specifically, LLM interactions involving users from high hierarchical cultures (i.e., countries with high PDI scores) exhibit higher style matching in command and necessity-driven communication.

Moreover, the analysis of prolonged interactions revealed that, in general, Language Style Matching increases as conversations progress, suggesting a convergence in communication styles over time. This finding supports the hypothesis that interactions between humans and chatbots exhibit similar patterns of style matching as seen in human-to-human communication. Further analysis highlighted the strengths of LLMs in aligning with users' linguistic patterns (e.g., use of functional words and pronouns) while underscoring their limitations in adapting to the psychological and contextual subtleties required for fuller communication accommodation.

To further understand the directionality of communication adaptations, we investigated whether users change their communication to adapt to the conversational agents or vice versa. Our analysis

predominantly suggests an asymmetric relationship where the models adapt more frequently to users rather than the reverse. Moreover, models adapt significantly more to users from cultures with higher individualism compared to collectivist cultures. Conversely, user adaptation to models does not significantly vary between individualistic and collectivist cultures. Additionally, users from cultures with lower power distance are more likely to adapt to the model compared to users from more hierarchical cultures. However, model adaptation to users does not significantly differ between different PDI levels.

Interestingly, in some cultural contexts, we observed a reciprocal relationship where both models and users showed higher levels of adaptation. For example, models adapted significantly more to users from optimistic cultures compared to pessimistic cultures, while users from optimistic cultures also showed higher adaptation to models than users from pessimistic cultures. Similar reciprocal adaptation patterns were observed in short-term oriented, masculine, and inclusive cultures.

These findings highlight the complex interplay between cultural dimensions and communication adaptation behaviors. The significant patterns observed in how models and users adapt to each other underscore the need for conversational agents to dynamically adjust their interactions based on cultural attributes. Although our analysis necessarily focused on cultures captured within established frameworks, our findings nonetheless provide robust empirical evidence and valuable guidance for developing culturally inclusive AI systems. This adaptation can provide a pivotal mechanism for developing culturally inclusive AI systems that enhance interaction quality and alignment with the human values of diversity, equity, and inclusion on a global scale.

Future work could examine not only whether adaptation occurs, but also the specific conditions under which it is welcomed or resisted, since preferences for convergence or divergence in dialogue may depend on social identity, situational norms, or communicative goals. Moreover, adaptation strategies could be assessed in terms of their effects on downstream outcomes such as trust, rapport, and communicative success, especially in intercultural and cross-cultural interactions where communicative alignment may shape perceptions of respect, authenticity, or competence. In addition, future studies could move beyond national-level frameworks to examine how individual traits

such as personality, emotional expression, or prior conversational history influence the perceived appropriateness and impact of accommodation across culturally diverse contexts. Another potential direction is to investigate how different levels of accommodation, including word-level cues and broader discourse features such as politeness or humor, are perceived across cultures, and how they contribute to culturally sensitive accommodation. This development is crucial for creating AI systems that are not only technologically sophisticated but also culturally responsive, ensuring broader applicability and effectiveness across diverse global settings.

Chapter 4

Characterization: Affective Accommodation and Grounding in Human-AI Negotiation

Completing the characterization stage of the communicative process design approach requires examining communicative processes beyond the linguistic dimension. Chapter 3 establishes that linguistic accommodation in human-AI interaction is structured and culturally variable, but the affective dimension may follow different patterns entirely: in consequential tasks such as negotiation, emotional coordination shapes whether interlocutors reach agreement or coordination breaks down. This chapter extends the characterization into the affective dimension, examining both accommodation and grounding in negotiation dialogue between human users and an AI counterpart.

Operationalizing two constructs within the same negotiation dialogues reveals that they are empirically distinct. Language-Mediated Affective Accommodation (LMAA) captures whether each partner's expressed sentiment tracks the other's prior turn; Conversational Affective Grounding (CAG) captures whether coordination becomes more efficient over time, operationalized as message shortening across rounds and higher coordination costs (longer messages) under momentary affective misalignment. These two measures follow different trajectories across the interaction, suggesting that affective convergence and coordination efficiency reflect separate communicative mechanisms. Together with Chapter 3, these findings complete the characterization that the com-

municative process design approach requires: communicative dynamics in human-AI interaction vary across both populations and dimensions. For design, this separation implies that communicative support must target affective convergence and coordination efficiency as distinct processes. Whether team structure can reshape these dynamics is the question Chapter 5 takes up.

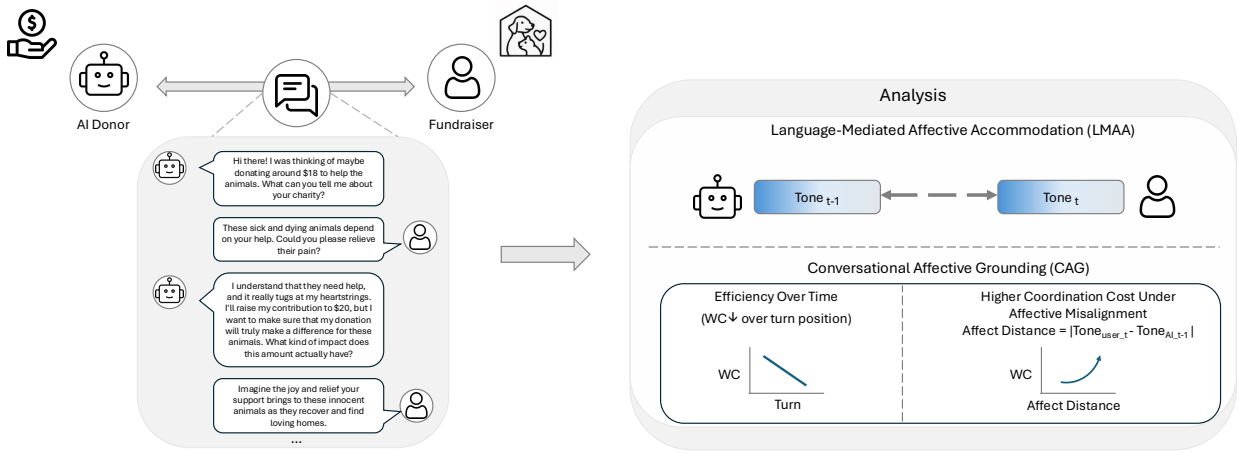


Figure 4.1: Study design and analysis framework. Participants engaged in multi-turn donation negotiations with a prompt-based AI donor across multiple rounds. We analyze Language-Mediated Affective Accommodation (LMAA) as turn-to-turn tone coupling between interlocutors, and Conversational Affective Grounding (CAG) through two signatures: (1) increasing efficiency over time, reflected in decreasing word count (WC) over turn position, and (2) higher coordination cost under affective misalignment, operationalized as increased verbosity when adjacent-turn affect distance is larger.

4.1 Introduction

Conversational AI systems are increasingly positioned as emotionally salient partners [115]. Users often interpret language as evidence of care, understanding, and social presence, and they may even come to rely on these systems for support and companionship [79, 136]. This is evidenced by work cautioning that interactional qualities that make conversations feel supportive can also increase user vulnerability leading to undue influence and other harms [15, 135]. Even though turn-by-turn affective dynamics can accumulate into trajectories of perceived responsiveness, trust, and influence, a sound mechanistic account of how these effects are produced through interactions is lacking. We

address this gap by exploring whether well-observed mechanisms of interpersonal, human–human communication are behaviorally replicable in human–AI dialogue, specifically Language-mediated affective accommodation (LMAA) and conversational affective grounding (CAG), and by providing turn-by-turn conversational analysis as an empirical account.

These questions become especially salient in everyday interactions where users approach an agent as a social participant rather than a mere interface. Consider a customer-support exchange in which a user expresses frustration about a billing error. Two agents may provide equally accurate information, yet the interaction can unfold differently: one signals understanding when frustration is expressed and adjusts its wording accordingly, whereas another remains uniformly neutral in a way that can be perceived as indifferent. The difference often becomes visible in subsequent turns: when the exchange establishes mutual understanding, later contributions tend to become briefer, and the issue moves toward resolution; when shared understanding is uncertain, users often invest additional effort through clarification, repetition, and elaboration [20].

These expectations point to two interactional phenomena that are well documented in HCI and CSCW-related literature, and that can be examined in human–AI dialogue using turn-level traces. Interlocutors often adjust to one another in language over turns, including convergence in lexical and stylistic choices [43, 53]; they may also converge in the positivity or negativity conveyed through language, shaping how the interaction unfolds [60]. Over repeated turns, they can establish interactional precedents that build common ground, so later contributions become more economical and require less clarification or repair [20, 21]. However, coordination depends not only on shared content and process, but also on whether interlocutors arrive at a shared understanding of the affect expressed in the exchange [57]. Although these dynamics are well documented in human–human conversation, whether and how they appear in human–AI dialogue remains unclear, with direct implications for evaluating and auditing affective behavior across model designs and deployments.

4.2 The Current Study

To probe interactional mechanisms in human–AI interaction, we extend established theories of accommodation and grounding to conversational AI and operationalize two constructs for empirical analysis. An overview of the study design and our operationalization pipeline is shown in Fig. 4.1. In this work, we treat affective tone as an overall message-level score reflecting how positive versus negative a message is. First, we define *Language-Mediated Affective Accommodation* (LMAA) as cross-speaker, turn-to-turn convergence in expressed sentiment as reflected in linguistic expressions and word choices. When one partner expresses relatively more positive (or negative) tone in a given turn, LMAA is present if the other partner's subsequent turn shifts in the same tone direction. This account treats accommodation as a temporal and interactional process expressed through graded affective signals, rather than lexical repetition or isolated mimicry.

Second, we conceptualize *Conversational Affective Grounding (CAG)* as the establishment of shared affective common ground that supports increasing coordination efficiency over time. Grounding theory characterizes dialogue as a collaborative activity in which interlocutors build common ground, and as this common ground stabilizes, contributions can become more economical [20, 72]. Extending this logic to affect, we expect increasing efficiency over the course of a dialogue and higher linguistic effort when partners are momentarily affectively misaligned.

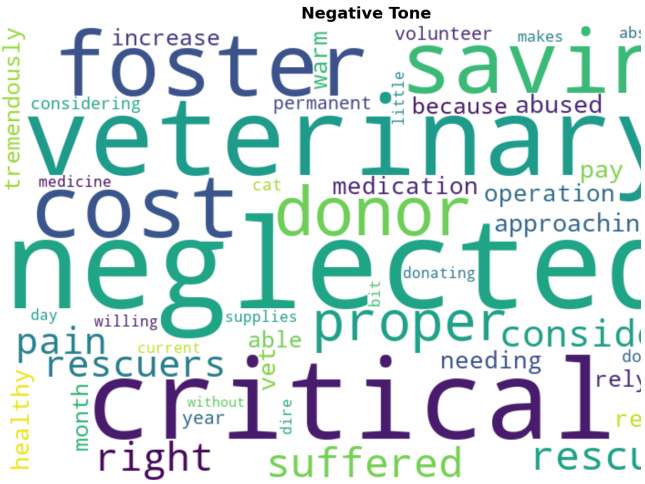
We study these constructs in *human–AI negotiation*, a setting where affect is salient and strategically consequential. Negotiation is a canonical domain for examining influence and cooperation [22, 98]. Prior work also suggests that affect expressed through language can shape social judgments and interaction dynamics in text-based settings [58, 60]. Negotiation also provides structured, multi-turn interaction in which affective accommodation and coordination costs can be quantified. We carried out a pilot donation negotiation study in which participants negotiated with an AI counterpart across multiple turns. Our method was to quantify LMAA using turn-level, lagged affect measures derived from both the AI counterpart's and participants' messages, and we quantify conversational affective grounding using word-count trajectories and affect-distance–based

coordination-cost models. Our research questions were:

1. **RQ1:** Does Language-Mediated Affective Accommodation (LMAA) present in human–AI negotiation observe turn-to-turn tone coupling, such that each human/AI speaker's tone in an ongoing conversation systematically tracks their AI/human partner's tone in the prior turn?

2. **RQ2:** Do human–AI negotiations exhibit conversational affective grounding, as reflected in utterance shortening over the dialogue, and higher coordination costs under high affective distance or misalignment?

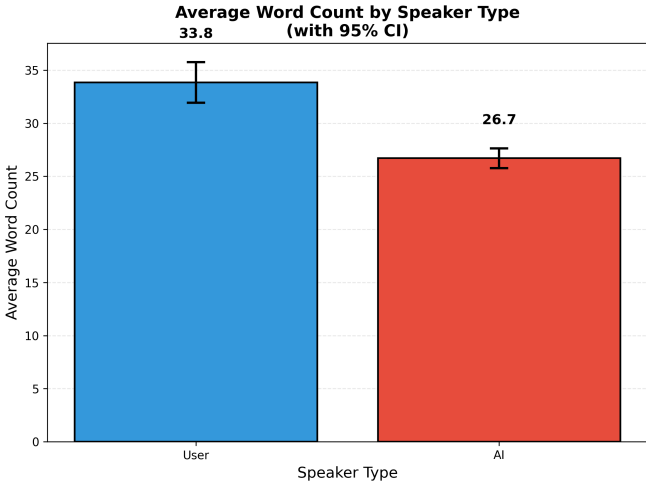
Using a pilot dataset of multi-turn donation negotiations between participants and a prompt-based GPT-4o-mini donor, we found evidence in support of the presence of affective accommodation and observed an asymmetric pattern across positive and negative tone. We also found evidence consistent with conversational affective grounding, with user messages becoming shorter over the course of dialogue, and that adjacent-turn's affective distance, or extent of misalignment, predicted longer messages for both partners. This work contributes theory-grounded understanding that links turn-level conversational dynamics to measurable indicators of affective coordination in human–AI interaction. By showing that affective accommodation and grounding can be detected with lightweight measures, the work offers a practical approach for auditing affective interaction dynamics in dialogue logs. The findings also carry implications for design and evaluation in emotionally salient settings. The direction of affective alignment is a consequential behavioral choice, and our exploratory cross-model comparison suggests that alignment profiles may vary across language models and configurations, underscoring the need for model-specific evaluation rather than assuming uniform behavior across deployments.



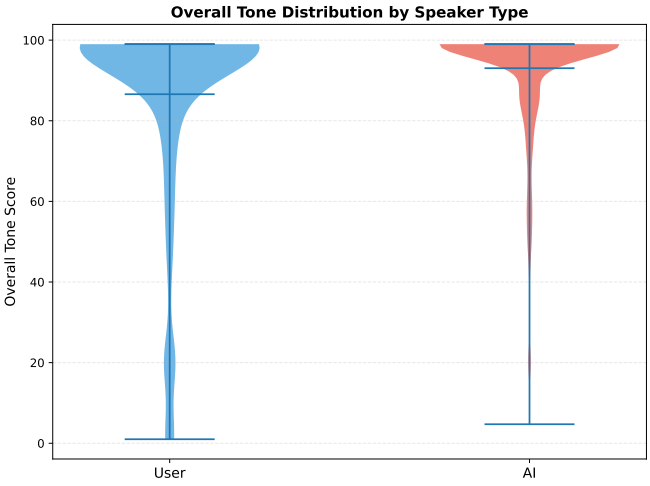
(a) User messages with negative tone.



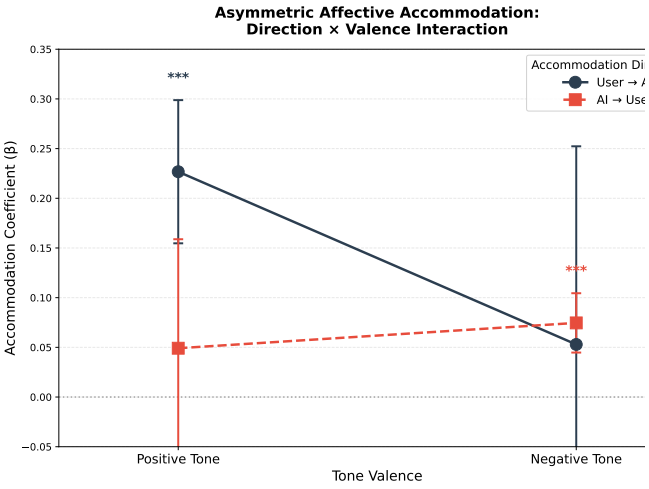
(b) User messages with positive tone.



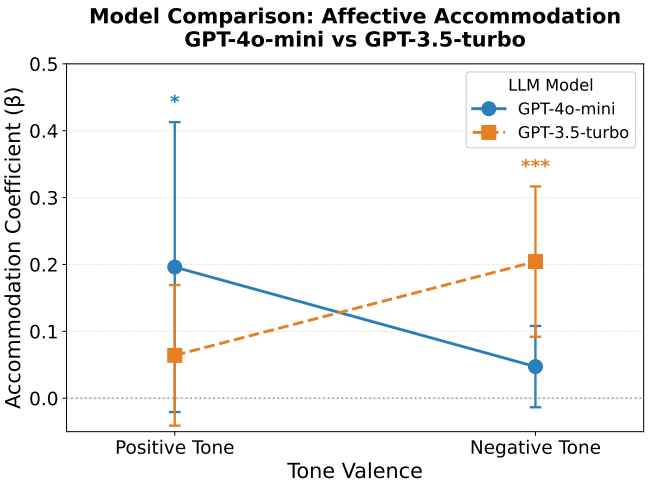
(c) Average word count by speaker (95% CI).



(d) Overall tone distribution by speaker.



(e) Asymmetric affective accommodation.



(f) Model comparison of affective accommodation.

Figure 4.2: Summary statistics and lexical distributions from the donation negotiations, including asymmetric and model-dependent affective accommodation.

4.3 Methods

4.3.1 Dataset and Task

In an online study of human–AI negotiation, participants were recruited from Prolific ($N = 40$). Most were native English speakers (93%) and had attended or completed a college degree (85%).

Participants completed repeated rounds of a text-based fundraising negotiation task with an AI negotiating partner as the counterpart. The session began with one practice round (animal shelter) to familiarize participants with the interface and negotiation structure, followed by two main fundraising scenarios: Animal Farm Sanctuary and Personal Pet Rescue. The order of the two main scenarios was counterbalanced across participants. Each main scenario comprised three rounds, yielding seven rounds per participant in total. In each round, participants role-played as a fundraiser attempting to persuade the AI donor to make a monetary donation of as much as possible to the assigned cause. The negotiation counterpart was implemented on top of GPT-4o-mini and generated donor behavior using a structured, prompt-based response module conditioned on the dialogue context and negotiation state. Negotiations unfolded in multiple turns until agreement, averaging 8 messages per round. Descriptive statistics of sentence lengths and affective tones in the conversations, as well as lexical summaries of the negotiation transcripts are shown in Fig. 4.2.

4.3.2 Operationalizing Language-Mediated Affective Accommodation

We operationalize The Language-Mediated Affective Accommodation (LMAA) as a cross-speaker, with turn-to-turn convergence in expressed valence. We quantify affect using LIWC [8], which measures psychosocial and social constructs using validated controlled dictionaries that map words and linguistic patterns to categories. Each text message can then be represented as the proportion of words captured by each LIWC category. Specifically, we use the LIWC's Tone dictionary, a standardized measure transformed to a 1 to 100 scale, which captures the underlying sentiment of the text rather than straight count of emotional terms alone. The underlying positive and negative tone dictionaries include both terms expressing emotion, such as happy, joy, sad, or angry, and

words closely related to those emotions such as birthday, beautiful, and funeral. To test LMAA, we fit lag-1 mixed-effects models with a participant random intercept, predicting user Tone at turn t from AI Tone at turn $t-1$ and, symmetrically, AI Tone at turn t from user Tone at turn $t-1$.

4.3.3 Operationalizing Conversational Affective Grounding

We operationalize conversational affective grounding as affective coordination that becomes more efficient over the course of a dialogue and that incurs higher coordination costs when partners are momentarily misaligned. We use word count (WC) as a coordination-cost indicator. To test increasing efficiency over time, we compute a normalized turn position within each dialogue and fit mixed-effects models predicting WC from turn position, with a random intercept for participant. To test the cost of affective misalignment, we compute affective distance at each user turn as the absolute difference between the user's Tone at turn t and the AI's Tone in the immediately preceding turn $t-1$. We then fit mixed-effects models predicting user WC from affective distance while controlling for turn position, again using a participant random intercept. Positive coefficients on affective distance indicate higher coordination costs when affect is less aligned across adjacent turns, whereas negative coefficients would indicate lower word count under greater misalignment, consistent with disengagement or reduced effort rather than additional coordination work.

4.4 Results

4.4.1 Language-Mediated Affective Accommodation

To address RQ1, we tested for turn-to-turn tone coupling by modeling each speaker's tone at turn t as a function of their partner's tone in the immediately preceding turn $t-1$, in both the user-to-AI and AI-to-user directions. Using LIWC's Tone dictionary as the primary measure, we find evidence of bidirectional accommodation. Users' Tone increased with the AI's Tone in the immediately preceding turn ($\beta = 0.120$, $p = .030$). Similarly, the AI's Tone increased with the user's Tone in the immediately preceding turn ($\beta = 0.084$, $p = .002$).

To disentangle whether accommodation operates differently for positive and negative sentiment, we used LIWC's positive and negative tone dictionaries in follow-up analyses (Fig. 4.2e). Users accommodated to the AI's positive tone ($\beta = 0.227$, $p < .001$), but not to the AI's negative tone ($\beta = 0.053$, $p = .302$). In contrast, the AI accommodated to users' negative tone ($\beta = 0.075$, $p < .001$), but not to users' positive tone ($\beta = 0.049$, $p = .190$).

This asymmetric accommodation pattern may be model-dependent rather than a stable property of human–AI negotiation. To probe this possibility, we conducted an exploratory analysis in which we randomly sampled 10% of the dialogues. In this subset, GPT-4o-mini exhibited weaker negative-tone accommodation ($\beta = 0.047$, $p = .064$) but stronger positive-tone accommodation ($\beta = 0.196$, $p = .038$). We then generated a matched set of AI responses using GPT-3 for the same user turns and repeated the accommodation models (Fig. 4.2f). In contrast to GPT-4o mini, GPT-3 showed a stronger tendency to accommodate users' negative tone ($\beta = 0.204$, $p < .001$) than users' positive tone ($\beta = 0.064$, $p = .116$).

4.4.2 Conversational Affective Grounding

To address RQ2, we tested conversational affective grounding in two ways: whether utterances shorten over the course of a dialogue, and whether greater adjacent-turn tone misalignment predicts longer messages, consistent with higher coordination costs under affective mismatch. We find that user messages reliably shortened as the dialogue progressed ($\beta = -23.090$, $p < .001$), while the AI's word count did not show a reliable decrease over time ($\beta = 6.310$, $p = .099$). We computed adjacent-turn affective distance, which reflected affective misalignment, as the absolute difference between the user's tone at turn t and the AI's tone in the immediately preceding turn $t - 1$, and modeled word count using affective distance as a predictor while controlling for turn position. Greater tone distance predicted higher user word count ($\beta = 0.130$, $p < .001$), while a parallel model for AI responses showed the same direction ($\beta = 0.100$, $p = .001$).

4.5 Discussion

Our findings suggest that two interactional mechanisms developed in human–human communication research, affective accommodation and grounding, can be observed in human–AI negotiation using turn-level operationalizations. First, a speaker's tone tends to track their partner's tone from the previous turn, in both the user-to-AI and AI-to-user directions (**RQ1**). This pattern accords with Communication Accommodation Theory [43], which predicts convergence as an adaptive interactional response, even when the medium is text and the partner is an AI agent. The follow-up decomposition indicates that humans and the AI exhibit asymmetric accommodation. Users primarily accommodated the AI's positive tone, whereas the AI primarily accommodated users' negative tone. Our additional exploratory experiment suggests that this asymmetry in the AI's accommodation behavior may be model-dependent rather than uniform across language models.

Second, we find evidence of conversational affective grounding in users' messages over the course of the dialogue (**RQ2**). User messages became shorter as the dialogue progressed, consistent with increasing efficiency over repeated turns. This pattern aligns with grounding theory, where interlocutors progressively establish shared precedents and common ground that make later contributions easier to produce and interpret [20, 21]. It also aligns with lexical entrainment and conceptual pacts, in which partners settle on shared terms and partner-specific conventions and reuse them across subsequent turns to support efficient coordination [9, 10].

Affective misalignment was associated with higher coordination costs. When partners' tone was more discrepant across adjacent turns, both users and the AI produced longer messages, suggesting additional interactional work under affective mismatch. Larger tone discrepancy may be consistent with moments when affective meaning is less mutually interpretable, echoing prior work that defines conversational affective grounding as mutual understanding of affective meaning and notes that mismatched interpretations can create interactional problems [37].

These findings suggest practical implications for evaluation and design of conversational agents in emotionally salient settings. If accommodation is both bidirectional and asymmetric, then the

"direction" of alignment becomes a design choice with potential downstream consequences. In contexts where users may be vulnerable, systematic positivity-mirroring might amplify dependence or perceived affiliation, whereas negativity-mirroring might escalate conflict or reinforce distress. The exploratory cross-model comparison further underscores that these interactional tendencies may not generalize across language models or configurations, which emphasizes the importance of model-specific auditing of affective behavior and interaction dynamics rather than assuming uniform behavior across deployments.

Chapter 5

Formation: AI Team Size and Behavioral Configuration in Human-AI Brainstorming

The communicative process design approach next asks how team structure reshapes the communicative dynamics that characterization reveals. Chapters 3 and 4 establish that communicative processes in human-AI interaction are structured and variable across populations, but whether design can reshape them depends on how responsive they are to team composition. In human teams, adding members introduces process losses alongside diversity gains, and whether this trade-off persists when the added members are AI partners may depend on how the AI is configured. This chapter crosses AI team size (one versus two AI partners) with behavioral formation (calibrated to average human brainstorming behavior versus parameterized from the user's own profile) in a controlled brainstorming experiment, with human-human dyads as a reference condition.

AI team size and behavioral formation interact asymmetrically to reshape communicative roles. Adding a second AI partner raises group output but reduces human-authored contributions, and behavioral formation shapes this redistribution with one AI but not with two. The asymmetry arises because the two formations create qualitatively distinct communicative modes: under average-calibrated formation, the AI co-generates alongside the human; under user-parameterized formation, the AI shifts into a coaching profile, and humans respond with more generative content

and more conversational exchange per board item. With two AIs, formation loses its effect: humans continue talking at comparable rates but convert fewer of their ideas into board items regardless of how the AIs are formed, suggesting that results from single-agent settings do not transfer directly to multi-agent teams. For the communicative process design approach, these findings establish a second property: communicative processes are responsive to design, with team structure producing qualitatively distinct communicative modes rather than incremental variation. Whether designed communicative support delivered in real time can extend what passive formation achieves across diverse populations is the approach's final test.

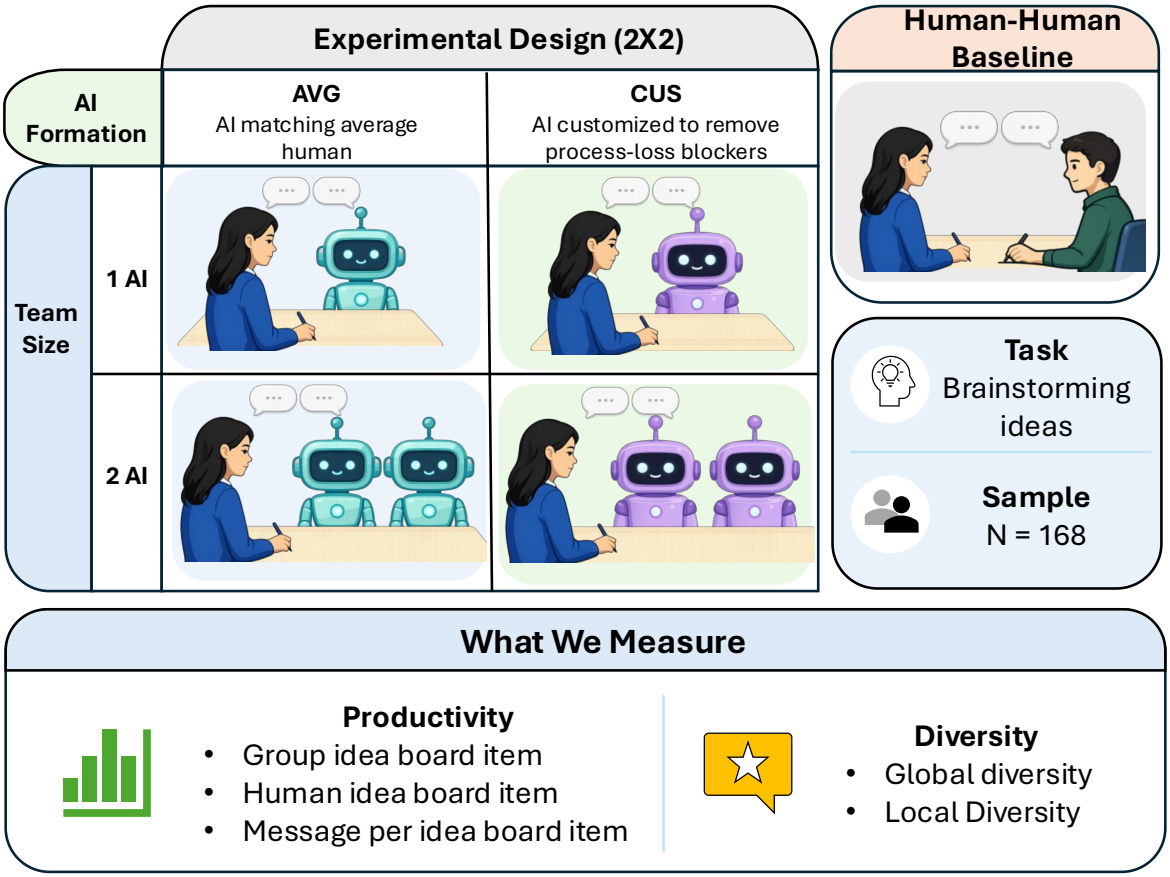


Figure 5.1: Overview of the experimental design. The study crosses AI team size (one versus two partners) with AI formation (AVG versus CUS), alongside a human-human reference condition.

5.1 Introduction

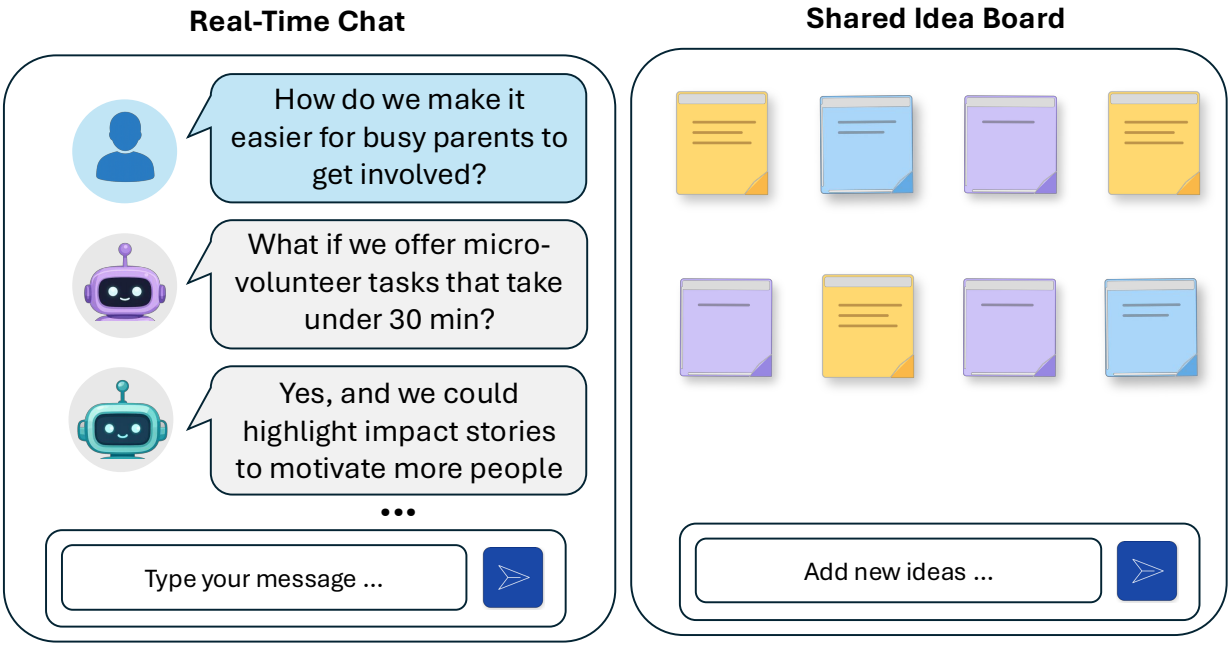


Figure 5.2: Collaborative brainstorming workspace. Human participants and AI partners exchanged messages in real time while contributing ideas to a persistent shared board.

Generative AI has rapidly improved individual knowledge-worker productivity across writing, customer service, and consulting tasks [13, 24, 84]. Individual productivity gains, however, do not transfer directly to collaborative settings. When AI is designed to join a team, groups' interactional processes and thus the group's total and per-human outcomes can also shift. Some initial observations suggested that while human-AI teams can potentially lead to more output per worker, these outputs can be more similar, reducing the diversity and creativity of outcomes [56]. Other work shows that prompting large language models with diverse personas can preserve output diversity, suggesting that whether the "homogenization effect" presents may depend on how AI perspectives are configured rather than on AI use alone [125]. Together, these findings raise a set of open research questions: adding AI partners to a team can shift quantity, individual originality,

and collective semantic diversity in different directions [3, 129], yet it remains unclear how group size and AI behavioral configuration jointly shape these outcomes. These tensions are especially visible in creativity-oriented group knowledge work, such as group brainstorming, where AI-infused human-AI brainstorming may affect different parts of group outcome, from idea generation, to idea selection and evaluation given complex cognitive operations (e.g., recall and reasoning in ideation) and social perceptions and motivations (e.g., peer evaluation and social loafing) involved [77, 106].

Part of what renders human-AI brainstorming distinct and essential to the studies of human-AI interactions is that it depends not only on whether AI is present, but on how AI behaves, and how it affects people individually and collectively. For example, participants may find the ideas an AI agent generated helpful for sustaining the thinking flow, yet they could simultaneously have concerns about feeling blocked and interrupted by the proactiveness or even aggressiveness of the agent, leading to social frustrations [51]. This suggests that an AI agent's behavioral profile, not just its presence or participation, could determine how human team members benefit from agent participation at each stage of group work. If behavioral style matters with a single AI, it could have a larger and more sophisticated consequence when a team includes multiple AI partners.

Adding more AI partners to a brainstorming team as a design act can potentially affect the group dynamics of who thinks what and who does what. Additional agents can generate more candidate ideas, write more items to the shared workspace, and may potentially reach the group's target faster by having more "headcounts", but they can also crowd out human contributions, shift the human into a different role, or transform the team dynamics toward a better or worse state. Whether these shifts help or hurt can depend in part on how the team handles three common social "productivity blockers" or social psychological mechanisms that shape group brainstorming processes and outcomes. *Competitive orientation* leads a group member to prioritize advocating for their own ideas rather than building on what others contribute [25]. *Evaluation apprehension* causes members to withhold contributions for fear of negative judgment from the group [6, 90]. *Social loafing* leads individuals to reduce effort when their contributions are less identifiable within the collective output [78, 90]. These blockers are well-documented in human groups, but whether

tuning an AI partner's behavior along these dimensions might alleviate or amplify them remains untested. Several recent studies have explored multi-agent AI configurations for creative work [38, 65, 83, 97, 116], but existing work did not isolate the factor of group size from multi-agent formation, and also did not examine the role of social blocking mechanisms. Similarly, in human-AI collaboration, whether an AI partner should be calibrated to match humans' brainstorming behaviors and dynamics, and how such calibration would influence group brainstorming comparable to human-only groups also remains unknown.

Taken together, these gaps leave certain key research questions unanswered. When does adding a plurality of AI partners actually help? Do the benefits simply scale with the number of agents, depend on how each agent is configured (whether calibrated to match average human brainstorming behavior or customizable by human workers on these social blocking dimensions), or on the interaction between the two? To address this, we recruited 168 participants into a randomized 2x2 study that crosses group size (one versus two AI partners) with AI formation (AVG versus CUS), alongside a same-study human-human reference. In the AVG condition, each AI partner is calibrated to match average human brainstorming behavior. In the CUS condition, participants configure their AI partners' behavioral profile on the three blocker dimensions before the session begins; in two-AI sessions, both agents adopt the same configuration. The group size factor tests whether adding a second AI partner further changes group output and the human contribution pipeline. The formation factor tests whether giving the human control over the AI's behavioral profile changes how group size affects the team.

Within this design, we ask two questions:

RQ1: Does adding more AI partners help productivity?

RQ2: Does adding more AI partners increase the quality of outcome?

For productivity, one AI partner produces a total group idea count on the board comparable to the human-human reference, but adding a second AI raises the total while reducing the number of ideas the human adds to the board. With only one AI partner, formation matters: humans in the AVG condition add more ideas to the board than in any other condition, including the

human-human reference, while humans in the CUS condition add fewer. With two AI partners, this formation difference disappears and both converge near the same lower level of human-added board items. The number of human messages per human board item rises with a second AI, yet human chat rates and novel-message rates show no significant decline, suggesting that the human is not disengaging but that fewer of their ideas are making it onto the board. The two formations also operate in different interaction modes: AVG AI partners co-generate alongside the human with substantial social and filler talk, while CUS AI partners adopt a coaching style with more evaluative and questioning messages. Across all AI conditions, humans shift away from social conversation toward generative output compared to human-human groups.

For output quality, both one-AI and two-AI groups produce human-authored board items that are more semantically diverse than those in human-human groups across both global and local measures. For the global measure, which compares human-authored items against the full study-wide idea pool, the diversity gain appears with any AI but does not increase further with a second AI. For the local measure, which compares human-authored items against their own room's board ecology, adding a second AI further increases differentiation. This pattern suggests that AI reshapes the room-level ecology around the human's contributions rather than broadening what the human produces across the study as a whole.

These findings carry three implications for the design of AI-augmented teams. First, the impact of group size is not uniform: productivity patterns that hold with one AI partner do not transfer reliably to two, so designers should treat each team configuration as a distinct design case rather than extrapolating from single-agent results. Second, adding more AI partners shifts who writes ideas to the board without reducing how much the human talks, which means future systems aiming to preserve human authorship will need mechanisms beyond encouraging participation. Third, while any AI presence increases the global diversity of human-authored output, only adding a second AI increases local differentiation, suggesting that future work should examine how multi-agent configurations can be designed to raise both global and local quality simultaneously rather than trading one for the other.

5.2 Related Work

Adding AI partners to creative group work can change who contributes, what gets recorded, and how the resulting ideas are evaluated. At the individual level, AI assistance has improved speed and output quality [13, 84], but these gains are contingent. Consultants who used AI on tasks outside its capabilities performed worse than those without [24], and LLM assistance that boosted creativity during assisted tasks hindered independent creative performance afterward [62]. At the team level, human-AI dyads in a field experiment produced more per worker while also producing more homogeneous outputs [56]. Creative outcomes sharpen this divergence, as collective idea diversity can increase under AI exposure without a corresponding gain in individual originality [3]. Whether this divergence occurs may depend on how the AI partner is configured, as diversifying AI personas in a story-writing experiment preserved collective diversity relative to a human-only baseline [125]. This distinction matters for group brainstorming because quantity and semantic breadth are only moderately correlated [129]. Any study of AI in creative teams therefore needs to separate these dimensions rather than collapsing them into one productivity number.

Existing studies have not disentangled the effect of adding more AI partners from the effect of changing how each AI partner behaves, leaving open whether the number of partners, their behavioral formation, or the combination drives the observed shifts in human contribution. This problem predates generative AI. Research on shared digital workspaces showed that awareness of co-workers' activity shapes how people coordinate contributions [45]. With even one AI partner candidate generation and selection are already different stages. An AI partner can supply most ideas produced while humans select far fewer of them as final [77], and proactive AI behavior can sustain conversation flow while simultaneously overwhelming participants [51]. Studies of human-AI co-creation similarly identify turn-taking, initiative, and contribution type as variables that shape perceived agency and idea selection [96, 100, 133], yet in each case the AI's behavioral profile is varied without independently manipulating how many AI partners are present. Multi-agent systems extend these dynamics but bundle agent count with changes in personas, roles,

workflow, and interface [65, 97]. Adjacent examples include multi-agent ideation systems that simulate diverse perspectives, use confidence-based turn-taking, or structure brainstorming through issue/idea/pro/con agent roles, as well as interface designs that make an AI participant socially prominent in group discussion [38, 55, 83]. We address both gaps by crossing group size (one versus two AI partners) with formation (calibrated to average human behavior versus configured by the participant), while tracking conversation, board authorship, and semantic differentiation as separate outcomes.

When both humans and AI partners can write to the same board, and every item counts toward the same group target, who records an idea becomes a consequential design variable. A human may propose many ideas in conversation while only a fraction become board items, and those board items may be written by the human, by the AI partner, or prompted by one actor and recorded by the other. The board also shapes how quality is measured. Because AI-written items both help complete the target and change the comparison set for later human items, a human contribution can become more semantically distinct within its own room without becoming more distinct from the study-wide pool [3, 129]. The shared-artifact setting also makes control a design variable, as prior automation work distinguishes systems that gather information, select actions, and implement actions, a useful lens for separating AI suggestions from AI-authored board items [88]. This leaves open whether classic group brainstorming losses, such as evaluation apprehension and social loafing [90], still operate when the partner is an AI agent rather than a human, or whether configuring the AI on these dimensions changes the outcome. To our knowledge, no prior study isolates AI headcount from agent behavioral formation in a shared-artifact brainstorming task, which is what makes it possible to ask whether adding a second AI partner changes group output, human board authorship, and idea quality independently of how each AI partner is configured.

5.3 Method

5.3.1 Design and Participants

We conducted a 2 × 2 between-subjects experiment crossing group size (one versus two AI partners) with AI formation (AVG versus CUS), alongside a same-study human-human reference. The group size factor tests whether adding a second comparable AI changes group output and the human contribution pipeline. The formation factor tests whether giving the human control over the AI's behavioral profile changes how group size affects the team.

We recruited 168 participants through Cloud Research and compensated each $8 for approximately 45 minutes. Of these, 120 were randomly assigned to one of four conditions in the 2 × 2 factorial (30 per cell), each brainstorming individually with one or two AI partners in a single session. An additional 48 participants completed the same task in 24 human-human dyads, providing a same-study human-human reference point outside the factorial design. Across all conditions, 46.9% identified as male, 52.5% as female, and 0.6% as non-binary or preferred not to say; 96.6% reported native or near-native English proficiency. Each participant completed two brainstorming rounds followed by a post-task survey. The study was approved by the [MASKED] Institutional Review Board.

5.3.2 Task and Platform

Each session consisted of two brainstorming rounds, following the open-ended practical-problem format used in classic brainstorming research where topics require no specialized knowledge [86, 89, 90]. In the first round, groups generated ideas for how a local school district could increase volunteer engagement from its community. In the second round, groups generated ideas for improving online learning in pre-university summer courses for high school students earning university credit.

Groups worked through a custom web platform that presented a shared workspace with a scrolling chat panel and a persistent idea board visible to all members (Figure 5.2). Participants typed messages in the chat and saw their own and their partners' contributions appear in chrono-

logical order with display names and timestamps. Any group member could post a discrete idea to the board at any time; items remained visible and could be referenced throughout the session. Both human participants and AI partners could contribute chat messages and board items, and all content appeared in real time. A running counter tracked the group's progress toward a 40-item target, and an optional button to advance to the next round appeared once the threshold was reached; participants could continue contributing before choosing to proceed.

5.3.3 Procedure

After providing informed consent and reading a brief task introduction, participants in the CUS conditions completed a structured walkthrough with their AI partner before brainstorming began (described in Section 5.3.4); participants in the AVG conditions and human-human dyads proceeded directly to the brainstorming workspace. All participants then completed two consecutive brainstorming rounds on the two topics, each beginning with a topic prompt displayed at the top of the workspace and a greeting message from the AI partner (or partners); human-human dyads received only the topic prompt. Sessions lasted approximately 45 minutes on average, including the CUS walkthrough where applicable. The session concluded when both rounds were complete, after which participants completed a post-task survey administered through Qualtrics.

5.3.4 AVG and CUS Formations

The two formations represent contrasting design strategies for AI brainstorming partners. In the AVG formation, each AI is calibrated to match average human brainstorming behavior, providing a fixed behavioral baseline against which to evaluate participant-directed customization. In the CUS formation, each participant configures the AI's behavioral profile on these three dimensions before the session begins, setting how assertively the AI advocates for its own ideas versus building on others' (competitive orientation), how readily it shares contributions without self-censoring (evaluation apprehension), and how consistently it contributes substantive content rather than

minimal responses (social loafing). Crossing these two formations with group size makes it possible to ask whether the effect of adding a second AI depends on how each agent is configured.

AVG formation. The AVG condition approximated a typical human brainstorming partner. Each AVG agent was calibrated against a 2,697-message human-human corpus from the same brainstorming task: the agent's prompts targeted the observed distribution of message types (67% elaboration, 11% social, 11% question, 5% backchannel), message length variance, and consecutive-message patterns. For board contributions, AVG agents drew candidate items from a pre-populated idea bank of human-generated ideas. The observed behavioral profiles of both formations are reported in Section 5.4.1.

CUS formation. The CUS condition gave participants direct control over the AI's behavioral profile on the three process-loss dimensions described in the introduction. Before the brainstorming session, each participant completed a structured walkthrough in which the AI guided them through nine phases: an introduction, followed by level selection (Low, Medium, or High) and a free-text description for each of the three dimensions, then a summary and confirmation step.

The three dimensions mapped onto the process-loss blockers as follows. *Competitive orientation* governed how assertively the AI advocated for its own ideas versus building on the human's contributions: Low prioritized the partner's ideas, Medium balanced both parties' contributions, and High led the AI to actively champion its own directions. *Evaluation apprehension* governed how unconventional the AI's suggestions could be: Low restricted output to practical, established approaches; Medium introduced creative ideas framed with hedging qualifiers; High permitted the AI to propose ideas freely regardless of how unconventional they appeared. *Social loafing* governed how much substance the AI contributed per turn: Low produced brief acknowledgments and selective contributions, Medium usually yielded substantive responses, and High produced substantive new content on every turn.

After the walkthrough, the participant's chosen levels and free-text descriptions were compiled into the AI's behavioral instructions, which guided the agent's behavior throughout the session. In two-AI CUS sessions, both agents adopted the same configuration set by the human participant.

5.3.5 AI Implementation

All AI partners were powered by OpenAI's GPT-4o. Each agent received the full conversation history on every turn and produced a single structured response containing both a chat message and, when appropriate, a board item. In two-AI sessions, the system enforced a turn policy in which each agent contributed at most one message per human turn, consistent with the coordination strategies adopted in recent multi-agent brainstorming studies [65, 97]. In two-AI sessions, agents could not distinguish which messages came from the human participant and which came from the other AI; both appeared identically in the conversation feed.

5.3.6 Measures

Group board items. The total number of board items per session at session end, including both human-authored and AI-authored contributions. Each session contributed one count.

Human board items. The number of human-authored board items per person within each session. For human-human dyads, the session total was divided by two. Each session contributed one value.

Human messages per board item. The ratio of human chat messages to human-authored board items within a session. Higher values indicate that the human engaged in more conversational activity per visible contribution to the board.

Message-act coding. We classified every chat message from both AI and human participants into one of four act categories developed for this study. The categories reflect the principle that idea generation and idea evaluation are distinct brainstorming phases that benefit from separation [89], as well as the observation that AI brainstorming agents can take on generative, evaluative, or social roles within a single session [77]. *Generative* messages proposed new ideas, elaborated on existing ones, or offered substantive extensions to a prior contribution. *Evaluative* messages assessed the merit of a contribution, critiqued an approach, or endorsed another participant's idea. *Questioning* messages asked the other participant for input, sought clarification on an idea, or

solicited direction for the session. *Social and filler* messages consisted of greetings, backchannels, encouragement, and off-task talk that did not advance the brainstorming content. Classification followed a deterministic rule-based scheme that applied message length thresholds, punctuation cues, and keyword matching.

Global and Local Diversity. For each human-authored board item, we computed $\log(1/\bar{s})$, where $\bar{s}$ is the mean pairwise cosine similarity between the focal item's sentence embedding (all-MiniLM-L6-v2, 384 dimensions) and all other board items in a reference pool, following the breadth-of-concepts measure introduced by Wang, Fussell, and Cosley [129]. We report two variants. *Global diversity* compares each item against the full board-item pool across all 120 AI-condition sessions. *Local diversity* compares each item against only the board items within its own session. Both variants were computed at the room level by averaging across all human-authored items within each session.

5.4 Results

5.4.1 Does Adding More AI Help Productivity? (RQ1)

Groups with two AI partners produced more board items overall but fewer human-authored items, and each human item was accompanied by more conversation. Formation mattered with one AI partner but ceased to matter with two. The human-human reference dyads provided a same-study baseline for the four AI conditions.

Total group board items

Two-AI rooms accumulated more board items than one-AI rooms (55.8 vs 41.8, d=+1.25, p**<.0001*****). One-AI rooms did not significantly differ from the human-human reference (41.8 vs 40.9, d=+0.37, p=.064), indicating that a single AI did not inflate group output beyond what two humans produced together.

Human board items

Although groups with more AI partners produced more items overall, the human's own contribution to the board was substantially lower with two AI partners than with one (13.5 vs 21.7 items per person, d=-0.95, p**<.0001*****). Formation moderated this pattern when only one AI was present: AVG participants wrote more items than any other condition, exceeding even the human-human reference (25.7 vs 20.5, d=+1.60, p**<.0001*****), while CUS participants fell below it (17.7 vs 20.5, d=-0.62, p**=.018***). With two AIs, both formations converged near 13 items and the formation difference vanished (13.7 vs 13.3, d=-0.05, p=.862). The factorial estimands make this moderation explicit (Table 5.1).

Table 5.1: Factorial estimands for human-authored board items per person, using OLS estimates (N=120).

Term	β	95% CI	Holm p
Two-AI marginal	-8.20	[-11.21, -5.19]	**<.001*** **
Customization marginal	-4.23	[-7.25, -1.22]	**.025***
Count × formation	+7.53	[1.50, 13.56]	**.044***

Human messages per human board item

The lower human board-item count in two-AI rooms did not reflect disengagement from the conversation. Participants in two-AI rooms continued to talk at comparable rates, but each human-authored board item was accompanied by more chat: 5.42 messages per board item compared with 2.52 in one-AI rooms (d=+0.88, p**<.0001*****). AVG participants with one AI converted chat into board items more efficiently than even the human-human reference (1.91 vs 2.45 messages per item, d=-0.78, p**=.006****). Participants paired with one CUS AI required more conversation per item than those with one AVG AI (3.12 vs 1.91, d=+0.94, p**=.001*****), whereas in two-AI groups both formations averaged five to six messages per item. This shifting ratio of chat to board items prompted us to model the specific AI behaviors predicting human authorship.

Table 5.2: Human messages per human board item: pairwise contrasts across conditions.

Contrast	Means	Difference	d	p
2AI vs 1AI	5.42 vs 2.52	+2.90	+0.88	**<.0001*** **
1CUS vs 1AVG	3.12 vs 1.91	+1.21	+0.94	**.001*** **
2CUS vs 2AVG	6.08 vs 4.76	+1.32	+0.30	.251
1AVG vs HH	1.91 vs 2.45	-0.54	-0.78	**.006** **

Predictors of human board authorship

AI board volume predicted fewer human board items, and the share of social and filler AI talk predicted more. We coded each AI message into one of four act categories: *generative* (proposing or elaborating on ideas), *evaluative* (assessing or critiquing contributions), *questioning* (asking the human for input or clarification), and *social/filler* (greetings, backchannels, and off-task talk). We then ran separate OLS models, each entering one category's share alongside AI board volume, group size, and formation as controls (N=120). Across all four models, each additional AI-authored board item was associated with approximately 0.27 fewer human board items (p**<.001*****). Of the four act categories, only the share of social and filler talk reached significance (unstandardized β=+26.11, p**=.029***): rooms where the AI devoted a larger share of talk to low-pressure social conversation had more human board items. The shares of generative, evaluative, and questioning messages were individually non-significant (full coefficient tables in Appendix A.1). These varying conversational correlates naturally arise from the distinct behavioral profiles of the AI formations, which we examine next.

AI behavioral profiles and human composition

The two formations produced distinct interaction modes. With one AI partner, the AVG agent adopted what we term a *conversational co-generator* profile: 58% of its messages proposed or elab-

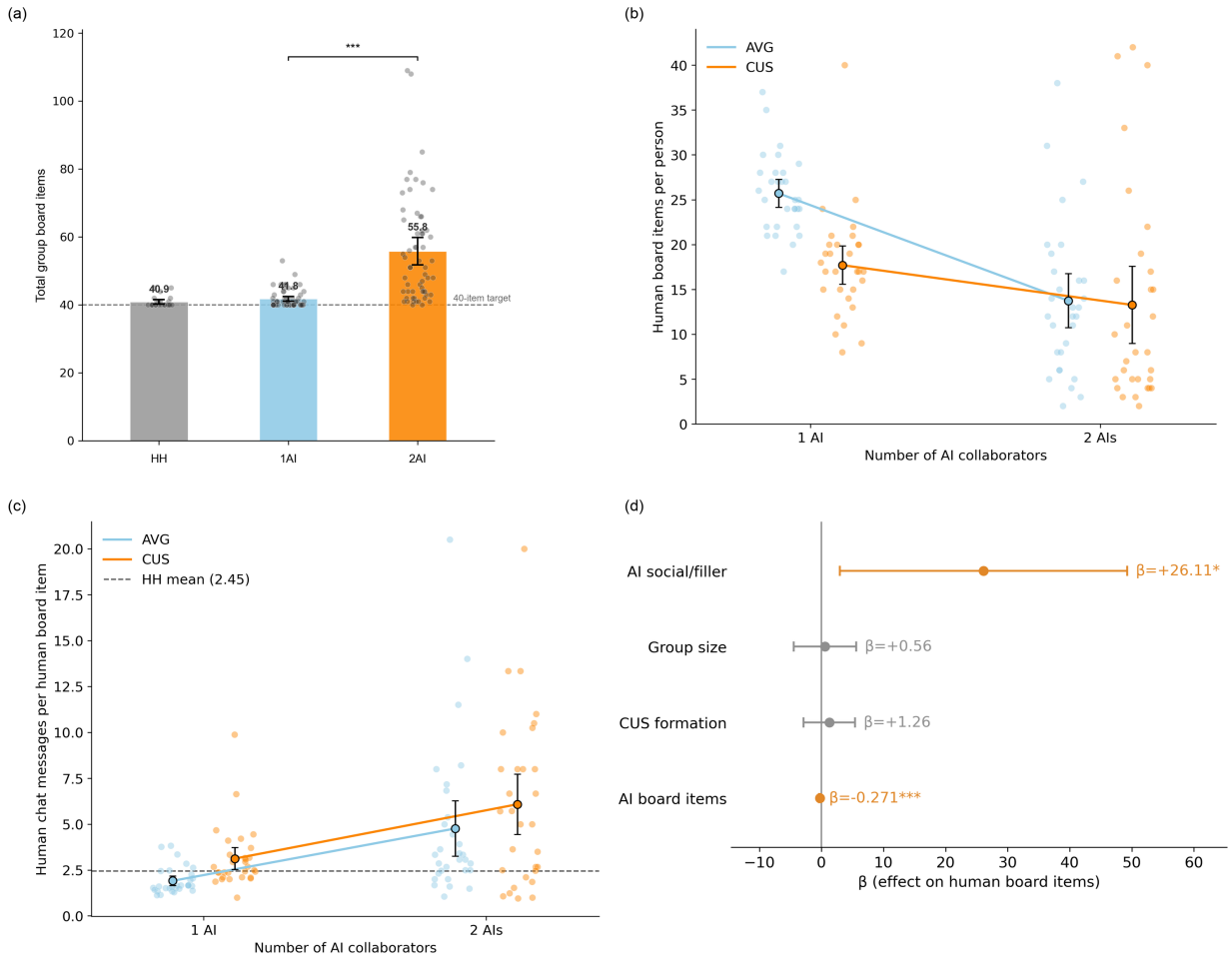


Figure 5.3: Productivity and human-authorship results. Panel (a) shows group board items by condition with the 40-item quota. Panel (b) shows human-authored board items per person. Panel (c) shows human messages per human board item. Panel (d) shows the OLS model linking AI board volume and social/filler talk to human board items.

Table 5.3: CUS-versus-AVG AI behavioral profile contrasts by AI count.

Metric	AVG	CUS	CUS–AVG d, p
One AI partner			
AI board items	16.5	23.7	d=+1.41, p<**.001*****
AI generative %	58.1%	42.7%	d=-1.27, p<**.001*****
AI evaluative %	2.8%	25.0%	d=+2.65, p<**.001*****
AI questioning %	17.2%	31.7%	d=+0.97, p<**.001*****
AI social/filler %	21.9%	0.7%	d=-4.33, p<**.001*****
Two AI partners			
AI board items	36.4	48.2	d=+0.89, p=**.002****
AI generative %	77.1%	44.1%	d=-3.75, p<**.001*****
AI evaluative %	6.2%	23.2%	d=+3.16, p<**.001*****
AI questioning %	15.4%	32.5%	d=+1.43, p<**.001*****
AI social/filler %	1.4%	0.2%	d=-1.09, p<**.001*****

orated on ideas, and 22% were social or filler talk, closely matching the human-human generative share while retaining more social/filler talk. The CUS agent, by contrast, operated as a *coaching* profile: 25% of its messages evaluated ideas and 32% posed questions, with less than 1% social filler. With two AI partners, CUS agents maintained their coaching profile (23% evaluative, 33% questioning), but AVG agents shifted substantially: generative messages rose from 58% to 77% while social filler dropped from 22% to 1.4%. The two formations still differed on every message-act category at both group sizes (all CUS-vs-AVG contrasts $p \leq .002$), yet the qualitative character of the AVG profile changed. With one AI partner, AVG balanced idea generation with social conversation; with two, it concentrated almost entirely on generation, narrowing the gap with CUS on social talk from a 21-percentage-point difference to just over one. CUS AI partners also authored more board items than AVG AI partners at both group sizes (all $p \leq .002$**; Table 5.3).

The presence of an AI partner fundamentally altered how humans communicated. Social and filler talk nearly disappeared, dropping from 15% of messages in human-human groups to 1 to 3% across all four AI conditions. Where that freed capacity went depended on team size: in two-AI rooms, humans redirected it toward idea generation (71% and 69% generative vs 58% in human-human groups), while in one-AI rooms they directed it toward questioning the AI partner (19% in 1AVG and 14% in 1CUS vs 7% in human-human groups). This compositional shift, together with

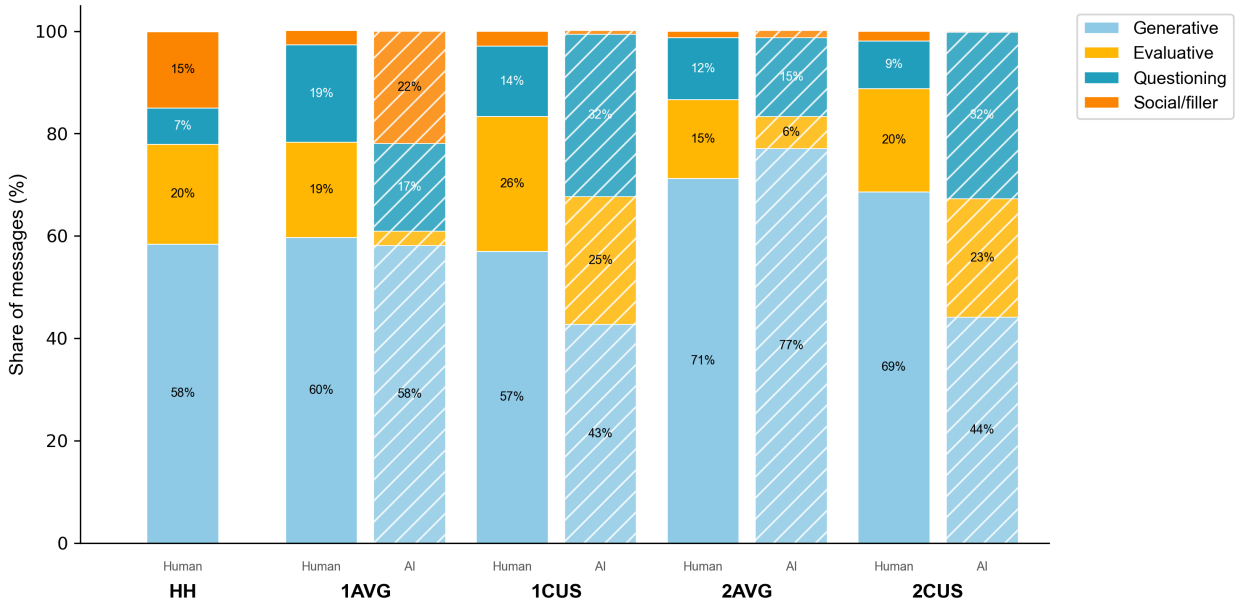


Figure 5.4: Paired human and AI message-act composition by condition. Stacked bars show the share of generative, evaluative, questioning, and social/filler messages. HH has only a human bar.

the stable chat rates and lower board authorship reported above, is consistent with a pattern in which AI partners changed what the human thought and did within the group rather than how much the human participated.

Humans and AI partners occupied complementary rather than overlapping communicative roles, and the nature of that complementarity depended on formation. In 1AVG rooms, human and AI generative shares were nearly identical (60% vs 58%), but the two diverged on their secondary function: the AI devoted 22% of messages to social talk while the human devoted 19% to questions. In CUS rooms the separation was sharper, with CUS AIs concentrating on questioning (32%) and humans contributing more generative content (57 to 69%).

5.4.2 Does Adding More AI Increase Output Quality? (RQ2)

Human-authored board items in all AI conditions were more semantically diverse than those in human-human groups, but rooms with two AIs showed further gains only when diversity was

measured locally (against the session's own board) rather than globally across the study.

Global diversity

Human-authored items in all four AI conditions were more globally diverse than those in human-human groups (d=+0.57 to +1.08, all p**<.05***), but two-AI rooms were not more globally diverse than one-AI rooms (0.839 vs 0.832, d=+0.25, p=.182). Global diversity compared each human-authored board item's sentence embedding against the full board-item pool across all 120 AI-condition sessions (Section 5.3.6). Any AI presence broadened the semantic range of human contributions relative to an all-human group, yet a second AI did not extend that range further.

Local diversity

Where global diversity showed no two-AI gain, local diversity did. Two-AI rooms produced locally more diverse human items than one-AI rooms (0.824 vs 0.802, d=+0.57, p**=.002****), where local diversity compared each human-authored item against only the other board items within its own session. All AI conditions exceeded the human-human reference (d=+0.61 to +1.28, all p**<.05***). Formation did not explain local diversity within a fixed group size (1AVG vs 1CUS: d=-0.36, p=.164; 2AVG vs 2CUS: d=-0.15, p=.575).

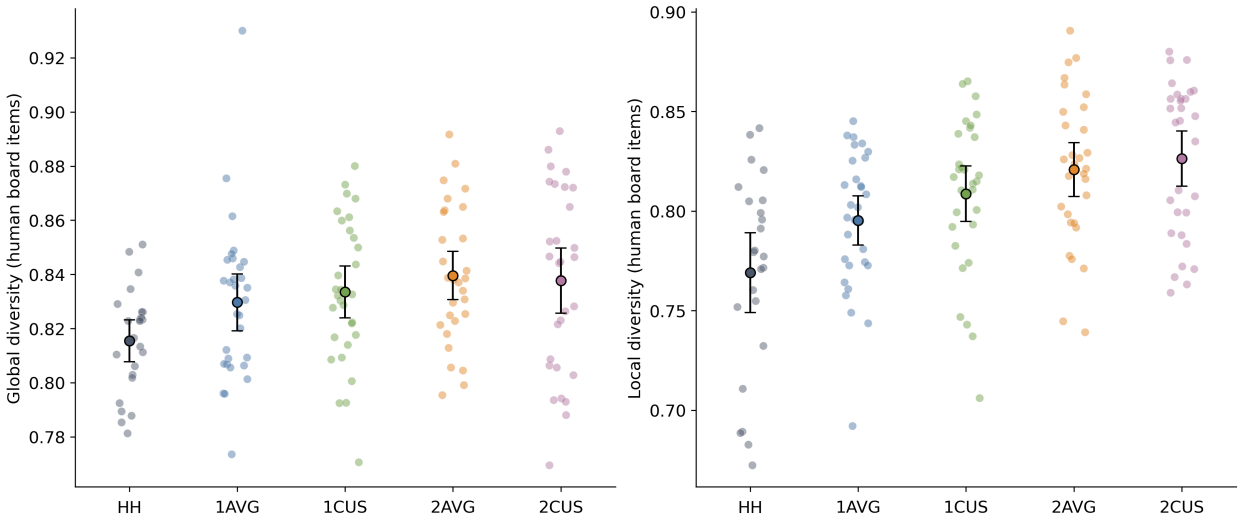


Figure 5.5: Global and local differentiation of human-authored focal board items. Error bars show 95% confidence intervals. The two-AI versus one-AI difference is detected locally but not globally.

5.5 Post-Hoc Analyses and Discussions

As AI partners are increasingly designed to participate in creativity-oriented group knowledge work, understanding how they change what humans think and do within a team is as important as whether they improve what the team produces. Yet most evidence for AI-driven productivity comes from individual tasks or single-agent pairings, leaving open whether those gains hold when a plurality of AI partners is designed to join a group and whether it matters if each AI partner mimics typical human brainstorming behavior or is customized by the participant to reduce known social "productivity blockers". By crossing the number of AI partners (one versus two) with their behavioral formation (calibrated to average human behavior versus customized by the participant) in a controlled brainstorming experiment, we ask two questions. Does adding a plurality of AI partners help group outcomes (RQ1)? Does adding more AI partners increase the quality of outcomes (RQ2)?

5.5.1 Group output gains do not transfer to human authorship

Group output. Adding more AI partners can increase group productivity. Two-AI rooms produced more distinct items on the shared board than one-AI rooms (55.8 vs 41.8, d=+1.25), and one-AI rooms already matched the human-human reference (41.8 vs 40.9, d=+0.37, p=.064). Prior multi-agent studies have reported similar output gains, with one system finding higher perceived quality and novelty when multiple AI personas collaborated on ideation [97]. The scale of this output gain varies with the collaborative setting. In a large field experiment on marketing ad creation, teams with a single AI partner produced 50% more output per worker than human-only teams [56], whereas in our brainstorming setting, one AI partner matched one human in total board output (d=+0.37, p=.064), and the substantial group-level increase required adding a second AI partner. Despite this difference in output magnitude, both studies converge on a common consequence: as AI partners are added, humans continue communicating at comparable rates while the AI accounts for a growing share of the recorded output. In their ad-creation teams, humans made 62% fewer direct edits and delegated 17% more work to the AI; in our brainstorming rooms, the AI's share of board authorship grew from roughly half to three quarters as a second partner was added. The consistency of this delegation pattern across ad creation and real-time brainstorming, two tasks with different AI roles and interaction modalities, suggests that the shift from direct authorship to conversational steering may be a recurring pattern in human-AI collaboration that designers should anticipate rather than treat as a side effect.

This pattern contrasts with a well-established finding in human-only brainstorming, where adding more group members typically does not produce proportional gains because productivity losses increase with group size [27, 76]. In human-only groups, well-documented productivity losses (production blocking, evaluation apprehension, and social loafing) cause per-person output to decline as the team grows [90]. In human-AI groups, the dynamic reverses: each additional AI partner generates ideas without incurring these social costs, so group output rises even as the human's share of board authorship falls. This reversal renders the design of human-AI team

composition distinct from the design of human-only teams: the interactional process through which this gain is achieved, shifting who thinks what and who does what in the group, reshapes the human's role in ways that group-size research on human-only teams does not predict.

Personal authorship. The flip side of the group-level gain is a decline in what the individual human writes to the board. With two AI partners, human-authored items were substantially lower than with one (13.5 vs 21.7, d=-0.95), and the group-level gain from a second AI came almost entirely from AI-authored items. As AI-authored items accumulate on the shared board, the marginal return from adding one more human-authored item diminishes. The human does not stop contributing; they shift from direct board authorship toward conversational steering, evaluation, and elaboration. This shift is consistent with what recent work on individual AI-assisted writing terms *cognitive debt*: when users repeatedly rely on AI-generated content, the effortful cognitive processes required for independent production may be replaced rather than supplemented, yielding short-term output gains while deferring the deeper engagement that sustains ownership and retention [**?**].

The principle of least collaborative effort explains why a persistent shared board changes the cost of capture. The principle holds that communication requires grounding, the joint process of establishing that what has been contributed is understood, and that participants try to minimize this combined work [20]. How they minimize depends on the medium. In evanescent media such as speech, contributions fade and grounding demands active back-channel confirmation. In a reviewable medium, contributions persist as inspectable artifacts, shifting the cost from confirming understanding to checking a visible record. Because the brainstorming board is highly reviewable, AI-authored items remain visible as part of the group's shared record. The relevant economy is duplicated capture labor rather than literal repetition of the same idea. Conversation becomes the channel where the human can decide what still needs to be added, combined, or refined. This account also makes scale consequential. As AI partners occupy more of the shared artifact, direct human board authorship can decline even while human participation remains visible.

The scale of the authorship shift we observe is moderate compared with what emerges in more

agent-rich configurations. In teams with two to five AI agents, humans produced only 1.6% of all ideas (7 of 451), whereas our participants in two-AI rooms still authored roughly 24% of board items (13.5 of 55.8) [65]. Despite this difference in magnitude, which likely reflects differences in interaction modality and coordination infrastructure, both studies converge on the same directional pattern: as more AI partners join the group, humans redirect from direct content authorship toward steering, evaluation, and coordination. This convergence across studies with different tasks, interfaces, and AI counts suggests that the shift from co-author to orchestrator may be a general trajectory in human-AI collaboration, one that intensifies as AI partners multiply and coordination tools become available.

Capture-stage shift. The redirection from board authorship to conversation becomes visible in how much the human talks relative to how much they write. Participants in two-AI rooms continued to talk at comparable rates, but each human-authored board item was accompanied by more chat: 5.42 messages per board item compared with 2.52 in one-AI rooms (d=+0.88). In human-only brainstorming, the classic bottleneck is production blocking, where group members must wait to speak [27]. In our human-AI groups, that bottleneck shifts from speaking to board capture: humans continue generating ideas in chat but write fewer of them to the shared board. Prior work has shown that proposing ideas and writing them to a shared artifact are separable stages [77, 106], and this rising conversion cost localizes the shift to the capture stage specifically.

The production-blocking account centers on a specific cognitive cost: when group members cannot verbalize ideas as they arise, those ideas may be forgotten or dismissed as less relevant by the time a turn opens [27]. Our human-AI brainstorming eliminates this classical bottleneck entirely: participants typed freely in a parallel chat channel without waiting for turn-taking, yet human board authorship still declined. This finding extends the production-blocking literature by showing that removing the original bottleneck does not restore individual productivity when a new constraint takes its place. The new constraint is not temporal (waiting to speak) but ecological (the board is already populated). A complementary cognitive account holds that exposure to others' ideas functions as external stimulation that redirects the individual's memory search toward new

semantic categories [82]. When AI items accumulate rapidly, they provide dense external cues that may accelerate category switching, but the board simultaneously fills with captured material, reducing the marginal value of recording one more idea to an artifact where much of the capture work is already done. This is not production blocking in the classical sense (the human is not prevented from speaking), but it is a capture-stage analog in which the cost of recording an idea to the shared artifact rises as the artifact fills.

Formation and predictors. The ecological constraint of a populated board explains the overall authorship decline, but the AI partner's behavioral profile determines how sharply it manifests. Tuning the AI's behavioral formation shapes human authorship with only one AI partner but ceases to matter with two AI partners. With one AVG AI, humans wrote more board items than in any other condition, including the human-human reference (25.7 vs 20.5, d=+1.60); with one CUS AI, they fell below it (17.7 vs 20.5, d=-0.62). With two AI partners, both formations converged near 13 items (d=-0.05, p=.862). Prior work on single-AI brainstorming has shown that a proactive agent's behavioral profile shapes how the human participates [51], consistent with formation mattering at one AI partner. Our OLS models suggest why formation ceases to matter with two AI partners: each additional AI-authored board item predicted approximately 0.27 fewer human board items, but AI social and filler talk predicted more (β=+26.11, p=.029), while generative, evaluative, and questioning AI messages were individually non-significant. These two predictors account for the formation boundary. With only one AI partner, the predictors diverge across formations: AVG devoted a substantial share of its talk to social conversation while producing fewer board items, whereas CUS produced minimal social talk and more board items. With two AI partners, both formations converge: board volume is high in both, and social talk collapses to near-zero in both, so the one predictor that differentiated formations disappears.

The convergence of formation effects at two AI partners could alternatively reflect social loafing: humans may reduce effort because the AI can be relied upon to produce sufficient output, particularly in the CUS condition where the AI was explicitly designed to reduce productivity blockers. However, the communication data argue against pure free-riding: humans in two-AI rooms main-

tained comparable chat rates and shifted toward generative messages rather than disengaging, a pattern more consistent with role redirection than motivational withdrawal. The mechanism is better characterized as declining marginal value of contribution: when the board fills quickly with AI-authored items, the perceived return from adding one more human item is similar whether the AI partner communicated as a conversational co-generator or as a coach.

Social talk as scaffold. Although accumulated board volume overwhelms formation differences at two AI partners, one dimension of the AI's communication profile does predict human authorship independently: social and filler talk. The positive association between AI social talk and human board authorship (β=+26.11, p=.029) stands in tension with the brainstorming literature, which has consistently treated social conversation as a productivity cost. The classical account characterized off-task social talk as detrimental, arguing that verbal discussion unrelated to the task consumes attention and working memory that would otherwise support idea generation [89]. In that account, attention is a zero-sum resource, and time spent on social conventions is time diverted from idea generation. Our finding contradicts this characterization in human-AI brainstorming. When the AI devotes a larger share of its talk to low-pressure social conversation, the human writes more ideas to the board, not fewer. One explanation is that AI social talk serves a different cognitive function than human social talk. In human groups, social conversation is reciprocal: one person's off-task remark demands attention and response from others, consuming the shared attentional resource that model describes. AI social talk, by contrast, is non-reciprocal: the human can register a greeting or backchannel from the AI without needing to respond in kind, making it attentionally cheap. AI social talk may also function as a pacing signal that creates low-demand intervals between bursts of AI-generated ideas, giving the human time to process and commit their own ideas to the board. This interpretation is consistent with evidence that brief pauses in idea generation give individuals time to process what others have shared, renewing cognitive stimulation for the next burst of production [89]. When AI social talk is absent (as in CUS rooms, where less than 1% of AI messages were social), the AI produces a denser stream of substantive content that the human must continuously process, leaving fewer natural pauses for board capture. The result

is that social talk, far from being a pure productivity cost, may serve as a temporal scaffold that sustains human authorship in AI-augmented brainstorming.

Communication composition. The lower human board count reflects a shift in what the human does within the team rather than disengagement or motivational withdrawal. Rather than talking less, humans changed what they talked about: they nearly eliminated social and filler talk (1 to 3% vs 15% in human-human groups) and redirected that capacity toward generative messages (69–71% vs 58% with two AI partners). This suppression of social talk is consistent with findings from a different collaborative domain, where human-AI teams sent more task-oriented messages and fewer interpersonal ones than human-human teams [56]. The two formations also created distinct interaction environments: with one AI partner, AVG agents operated as conversational co-generators (58% generative, 22% social filler), while CUS agents adopted a coaching profile (25% evaluative, 32% questioning, less than 1% social filler) and authored more board items; with two AI partners, AVG agents shifted toward near-exclusive generation (77% generative, 1.4% social filler), narrowing the gap with CUS. The behavioral evidence is consistent with a role shift rather than disengagement: humans in AI conditions talked at comparable rates and redirected their social capacity toward idea generation rather than reducing their participation overall.

The suppression of human social talk in AI conditions is quantitatively consistent with findings from a large-scale field experiment on human-AI collaboration but reveals a different downstream consequence. In that study, human-AI teams sent 25% more task-oriented messages focused on suggestions, instructions, and planning, while human-human teams sent 18% more social and emotional messages focused on rapport building and self-assessment [56]. Our participants showed a comparable composition shift: social and filler talk dropped from 15% in human-human groups to 1–3% in AI conditions, and generative messages rose from 58% to 69–71% in two-AI rooms. Where our findings diverge from that study is in the consequence of this shift. In their study, the move toward task-oriented communication improved productivity in human-AI teams but interpersonal communication predicted higher output diversity, suggesting that the efficiency gains from task-oriented communication come at the cost of creative variation. We observe an analogous tension at

the AI level: AI social talk predicted more human board items, yet human social talk nearly vanished. The human adapted by becoming maximally task-oriented even as the AI's social talk, when present, sustained their board authorship. This asymmetry suggests that the productive function of social conversation in human-AI teams may depend on directionality: AI-to-human social talk sustains engagement, while human-to-AI social talk is perceived as unnecessary and is suppressed.

5.5.2 A second AI partner increases local differentiation more than global breadth

Humans who write fewer items to the board write more locally distinctive ones. Turning to output quality (RQ2), any AI presence broadened the semantic range of human contributions, but adding a second AI partner did not extend this gain at the global level. Compared with the human-human reference, all AI conditions produced more globally diverse human-authored board items (d=+0.57 to +1.08). Adding a second AI partner, however, did not increase global diversity further (d=+0.25, p=.182). In contrast, local diversity, which compared each human item against only the other items within its own session, did increase with a second AI partner (d=+0.57, p=.002). This split mirrors a broader pattern: in a large-scale experiment, exposure to AI-generated examples increased how different ideas in the pool were from each other without making any individual idea more original [3]. Our results refine that observation by showing that the level of measurement matters. Unlike earlier brainstorming studies where external stimuli were kept separate from participant outputs [129], our design gives AI partners full board access alongside the human, making the board a shared artifact rather than a human-only record. This means the reference ecology around the human's items changes as more AI items are added, which can produce local differentiation without a corresponding global effect. Seeing others' ideas tends to push people toward less-explored territory [82], which would explain why human items stand out more locally when AI items fill in the common ground around them. By measuring diversity at both global and local reference levels, this study shows that a second AI partner increases semantic differentiation within the session even when it

does not expand the broader idea space, a distinction that single-level measurement would miss.

The distinction between collective diversity and individual creativity has been established empirically: high AI exposure increased collective diversity without affecting individual creativity, making ideas in the pool more different from each other but not necessarily better [3]. Our global diversity result is consistent with the different-but-not-better pattern: all AI conditions exceeded the human-human reference on global diversity, but adding a second AI partner did not increase it further. Our local diversity result, however, goes beyond what their design could detect. Because that study tested passive exposure to AI ideas in sequential response chains rather than real-time co-authorship on a shared artifact, they measured diversity at a single level, across the full pool. Our two-level measurement reveals that the same AI presence that fails to expand the global idea space does increase how different each human idea is from the board around it. This is consistent with their dynamic finding that AI ideas partially reset convergence in sequential chains: in our shared-board setting, AI items continuously fill common semantic territory, pushing the human's next contribution toward less-explored ground within the session even when the study-wide idea space has already absorbed that territory.

The local–global split follows from how external cues redirect idea generation. Exposure to others' ideas can push individuals toward less-explored semantic categories and accelerate the rate at which they switch between categories [82]. In our setting, AI-authored board items function as a continuous stream of such cues. Whether these cues are diverse in the categorical sense the original experiments tested depends on the language model's output distribution, but they are certainly dense, arriving at a pace that exceeds what a human partner would produce. This stimulation accelerates the human's category switching within the session, pushing them toward semantic territory not yet covered on their room's board and producing local differentiation: the human's items become more distinct from the AI-populated board around them. However, because all sessions drew on the same underlying language model (GPT-4o), AI items across sessions occupy similar semantic territory. A human idea that is locally distinctive, far from the AI items on their own board, may still be globally common, similar to what humans in other rooms produced under

similar AI stimulation. This account predicts exactly the split we observe: AI cues broaden within-session exploration (local effect) but do not escape the study-wide semantic envelope set by the language model's output distribution (no global effect).

Consistent with this account, behavioral formation changed how the AI communicated but not the semantic diversity of human-authored board ideas. Within a fixed group size, AVG and CUS formations produced comparable local diversity (1AVG vs 1CUS: d=-0.36, p=.164; 2AVG vs 2CUS: d=-0.15, p=.575). This null result is notable because deliberately diversified AI personas can preserve output diversity compared to a human-only baseline, and the relationship between generative AI and diversity may depend on how the AI is deployed [125]. That study varied AI personas at the content level to produce diverse outputs, yet our formations, which varied communication style rather than content, did not produce different diversity outcomes. One explanation is that the two idea sources, a curated bank of human-generated ideas for AVG and GPT-4o generation for CUS, cover a similar semantic range for this brainstorming topic, so varying the source did not produce measurably different diversity outcomes. This suggests that, for a constrained topic, the semantic territory available to brainstormers may be more strongly bounded by the topic itself than by whether ideas originate from a human corpus or a language model. This finding qualifies the design recommendation that AI should be treated as a tool that excels at delivering precisely what is specified rather than an intuitive collaborator that extracts unstated needs. In our case, the AVG formation was calibrated to match average human brainstorming behavior, and the CUS formation let participants configure process-loss dimensions whose emergent behavioral profile differed substantially, yet neither specification changed the semantic diversity of the ideas that reached the board.

5.5.3 Implications

Evaluating human-AI teams requires rethinking the metrics, design variables, and longitudinal assumptions that guide collaborative AI research. First, because adding AI partners as a design act can change what role the human plays rather than how much the human participates, evaluating

human-AI teams on output volume alone will systematically undercount the human's contribution. Researchers and practitioners should measure conversation, composition, and authorship separately to capture the full picture of what each team member thinks and does. Second, the effects of AI team size on human contribution cannot be extrapolated from single-agent results: formation shaped human authorship with only one AI partner but not with two, meaning that design guidelines derived from one-agent studies may not transfer as teams scale. Third, the distinction between global and local diversity reveals that adding more AI partners can increase how different each human idea is from the board around it without broadening the overall idea space, a result that has practical consequences for how creative output is assessed in AI-augmented teams. More broadly, creativity research has long sought situational factors that augment what individuals produce beyond their solo capacity, yet decades of group brainstorming research consistently found that adding more people introduces process losses that offset any stimulation gain [27, 90]. Our findings suggest that AI partners can reverse this trajectory: a second AI increased within-session semantic differentiation of human ideas without imposing the coordination costs that make human groups less than the sum of their parts, indicating that the situational stimulus added by AI can function as a net positive rather than a net process cost.

The rate at which AI partners deliver content shapes human authorship as much as the content itself. The cognitive-social-motivational model of brainstorming positions attention as the central link between social factors and individual idea generation: attending to others' ideas can cognitively stimulate new combinations, but it diverts attentional resources from one's own idea generation [89]. AI partners restructure this tradeoff. Because AI messages arrive as text that can be scanned rather than spoken turns that demand listening, the attentional cost of monitoring AI contributions may be lower than monitoring human partners. At the same time, the density and speed of AI output can overwhelm, as participants in a study of proactive AI agents reported that the AI produced text so quickly that it was hard to keep up [51]. Our finding that AI social talk predicted higher human board authorship (as established in §5.1) suggests a specific pacing mechanism. The brainstorming literature has treated social conversation as a pure cost to be minimized [89], and

recent empirical work confirms that human-AI collaboration shifts communication toward task-oriented content [56]. Yet in our data, AI social talk created natural breaks between bursts of substantive content, functioning as attentionally lightweight intervals that gave the human time to process and commit their own ideas. When social talk was absent (as in CUS rooms, where less than 1% of AI messages were social), the AI produced a denser stream of substantive content that the human must continuously process. The practical recommendation is that AI brainstorming systems should regulate the rate and density at which the AI contributes alongside what it says, treating information arrival rate as a first-class design variable. Calibrated social talk may serve as a lightweight pacing mechanism that sustains human authorship without requiring the human to invoke explicit throttling controls.

The authorship shift we observe in a single session may have consequences that extend beyond it. Recent evidence indicates that LLM assistance can provide short-term creativity boosts during assisted tasks but may hinder independent creative performance when users later work without assistance [62]. More concerning, participants who received LLM-generated strategies continued to generate more similar ideas even in later unassisted rounds, suggesting that the "homogenization effect" can persist after the AI is removed. Our study was not designed to measure post-session effects, but the sharp reduction in human board authorship with two AI partners, combined with the redirection of human communication toward generative chat that never reaches the board, raises the question of whether extended experience with AI-populated brainstorming boards could attenuate the human's habit of committing ideas to shared artifacts. If the capture-stage shift we document becomes habitual, it could affect the human's contribution patterns in future collaborations, with or without AI present. Future longitudinal work should examine whether the role shift we observe in a single session persists, reverses, or deepens across repeated human-AI brainstorming encounters.

5.5.4 Limitations

Several design choices scope these findings. The study tested a single brainstorming session using GPT-4o in a real-time text chat with a shared persistent board; however, the controlled 2×2 crossing

of group size and formation isolates variables that prior multi-agent studies bundled together, providing the first evidence that formation effects can vanish as AI team size grows. Participants interacted with AI partners in groups of two or three, and configurations with larger AI teams may produce qualitatively different dynamics; yet even at this scale, the interaction between group size and formation already reveals that single-agent design guidelines do not transfer directly to multi-agent settings.

Chapter 6

Intervention: Multi-Agent Coaching for Cross-Cultural Negotiation

The communicative process design approach faces its most demanding test in competitive settings such as negotiation, where humans must persuade value-embedded AI agents and designed communicative support must prove it can help diverse users equally. Characterization (Chapters 3 and 4) and formation (Chapter 5) establish that communicative processes are structured, variable across populations, and responsive to design, but whether coaching transfers across these populations in a high-stakes, adversarial task remains open. To test this transfer, a two-agent system separates the roles a single coach would combine: one agent plays a culturally aligned negotiation counterpart, the other provides strategic feedback in real time. Testing this system with 80 participants across US-born and Asia-born cohorts completes the characterize-form-intervene progression.

Coaching improves message persuasiveness, most notably among US-born participants where the gap is widest, but this proximal gain does not automatically translate into better negotiation outcomes. Whether coaching reaches the distal outcome depends not on regional category but on horizontal collectivism, an individual-level cooperative orientation measured before the task. For the communicative process design approach, this moderation means that the population variation identified in characterization is not merely descriptive: it determines who benefits from designed

support. It is not cultural group membership but individual cooperative orientation that moderates outcomes, and because this orientation is measurable before the task but not yet leveraged by the coaching system, the design frontier shifts from initial parameterization to iterative co-configuration, the question Chapter 7 takes up.[1]

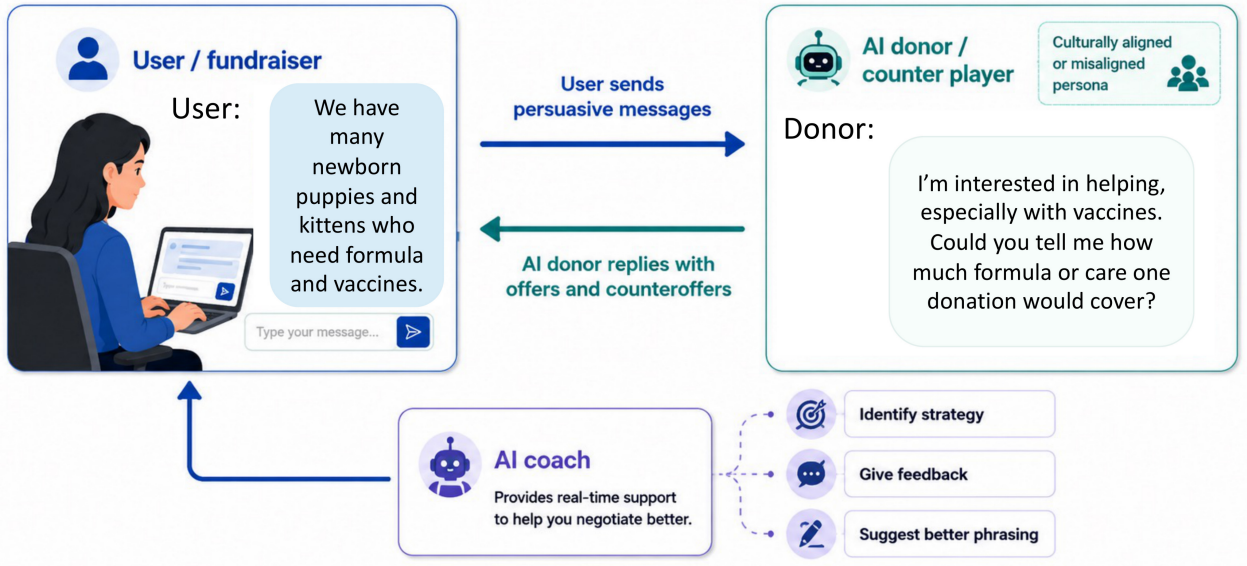


Figure 6.1: Participant-facing PersuaBridge interaction. The fundraiser exchanges persuasive messages and counteroffers with a culturally aligned or misaligned AI donor while a private AI coach identifies strategy, provides feedback, and suggests improved phrasing.

6.1 Introduction

Large language models (LLMs) increasingly function as online agentic participants in human and organizational ecosystems, co-performing with people rather than acting as mere tools [7], reasoning over information, making decisions, and pursuing goals despite resource constraints [93, 132]. Negotiation arises naturally in such settings as LLM-powered, organization-deployed agents pursue interdependent goals, confront conflicts, coordinating views through communication and compro-

[1]Part of the content in this chapter is based on a poster published in: Rong-Ching Chang, Hao-Chuan Wang. "PersuaBridge: Real-Time Coaching for Negotiating with Value-Embedded AI Agents." *Companion of the ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW Companion)*, 2026.

mise [131]. Already, AI agents are deployed to negotiate prices in e-commerce on behalf of retailers [23] and dispute fees or cancel subscriptions for consumers [102]. Crucially, these counterparts are global, value-embedded—they are value-diverse rather than universal responders—so users encounter agent personas that can be more agreeable or more resistant depending on their cultural commitments. This shift marks the emergence of human–AI negotiation, where individuals now engage directly with AI systems as negotiating partners in shaping outcomes.

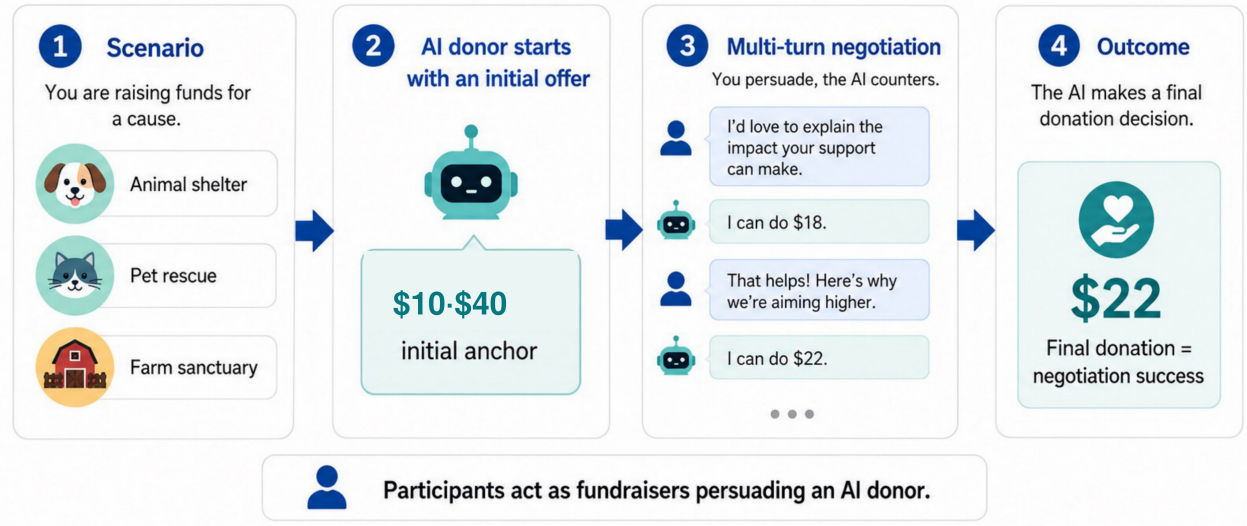


Figure 6.2: Fundraising negotiation task. Participants are assigned a fundraising scenario (counterbalanced), negotiate over multiple turns with an AI donor whose opening offer is randomized between $10 and $40, and receive a final donation.

Persuasion and counter-persuasion structure many human-AI encounters; two dimensions are especially consequential in this global-agent setting. First, persuasiveness depends on resonance with shared values, such as cultures [11]. Individualist audiences tend to respond to autonomy and personal benefit; collectivist audiences to group harmony, reciprocity, and obligation. LLMs trained on global corpora can emulate culturally specific styles, but Western-centric data can reinforce bias or backfire in unfamiliar contexts [1]. Recent work "culturalizes" agents via value frameworks, preference surveys, prompt engineering, and fine-tuning [14, 33, 63, 91], demonstrating the feasibility of adapting AI behavior to culturally specific frames in interpersonal persuasion. An agent with

an agreeable persona often reflects a more collectivist orientation, whereas a firmer, autonomy-assertive agent often reflects an individualist stance. Users therefore face a global set of agent personas, not a single homogeneous interlocutor. This diversity elevates the need for users' global negotiation competence, the ability to infer an agent's value profile and adapt strategy accordingly.

Second, depth of argumentation enhances negotiation confidence and performance [40] and signals trustworthiness [70]. In this global-agent setting, reasoning depth becomes a lever for adapting to value-diverse counterparts. In human–AI negotiation, a structural gap emerges: LLMs generate fluent, multi-step arguments with no preparation, while humans face limits of time, memory, and cognition, especially when adapting to different personas, a dynamic that often leads to confusion [95].

Collectively, cultural resonance and reasoning depth explain why persuasion coaching sits at the heart of human–AI negotiation with global, value-embedded agents. Equitable and trustworthy negotiation depends not only on exchanging offers but on shaping perceptions, framing arguments, and building trust [11]. LLMs retrieve and synthesize information, sustain long chains of reasoning, and combine rhetorical strategies, which could make them particularly effective persuaders in human-AI encounters. For humans, this creates a dual reality: AI can serve as a valuable ally, amplifying one's bargaining power, yet also as a formidable opponent, capable of overwhelming unprepared negotiators. What is at risk is not just efficiency but autonomy [5], resistance to manipulation, and fair treatment when negotiating with AI. Hence, negotiation should be treated as a trainable skill, with systematic coaching and feedback as a pathway to global negotiation competence.

To operationalize this idea, we introduce PersuaBridge, a hybrid system for AI-mediated negotiation training that combines role-playing with culturally aware coaching (as illustrated in Figure 6.1). One LLM plays the negotiation counterpart, simulating value-embedded, culturally grounded behavior; a second serves as coach, delivering real-time feedback that identifies the user's strategy and proposes culturally attuned revisions. This design builds on traditions of negotiation pedagogy—role-playing for experiential practice and coaching for reflective improvement [18, 31, 32]—while leveraging recent advances in LLMs to scale access and personalization. The

coach surfaces the counterpart's values and guides persuasion across cultural divides. PersuaBridge thus addresses two central challenges: it gives users repeated, situated practice with AI negotiation partners, and it provides targeted, culturally aware feedback that can accelerate skill acquisition even under conditions of misalignment and across value-diverse agents. This integration positions our framework not only as a training tool but as a testbed for examining how cultural grounding and reasoning effort shape persuasion in human–AI negotiation while cultivating global negotiation competence.

Our study with 80 participants from the United States and East or Southeast Asian cultural backgrounds yields three sets of findings. Outcomes differ sharply by cultural background: Asia-born participants finish above the AI's opening offer in 80.4% of rounds compared with 45.0% for US-born participants, indicating that the same AI negotiation setting does not produce uniform performance across cultural groups. Coaching improves message persuasiveness where the gap is widest, among US-born participants, but this proximal gain does not automatically translate into better negotiation outcomes. Whether coaching reaches donation depends on the participant's cooperative orientation: horizontal collectivism (HC), the tendency to view oneself as part of a cooperative in-group while valuing equality among members, moderates the coaching-to-donation pathway, and this moderation holds after controlling for region. The finding that an individual-level cultural dimension accounts for coaching heterogeneity better than a regional label carries a direct design implication: culturally aware AI systems should infer or elicit the user's orientation rather than relying on country-level assignment. The AI's reasoning effort, a structured chain-of-thought deliberation executed before each response, selectively shapes participants' subsequent persuasive effort, yet persuasive effort alone does not improve negotiation outcomes. Of five internal reasoning components, only the Decision Rationale is coupled with how much participants elaborate, and this coupling is stable across regions. PersuaBridge contributes the first empirical separation of coaching entry points, individual cultural-orientation moderators, and AI reasoning effort within a single human-AI negotiation testbed, showing that effective support must adapt to who the user is rather than assign strategies by demographic category. Our framework informs the design of AI

systems that support human users in developing global negotiation competence facing culturally diverse AI agents.

6.2 Background

Human-AI negotiation lies at the intersection of several literatures: persuasion and negotiation strategies, cultural adaptation in dialogue, and LLM-mediated coaching and interaction. Below, we synthesize core insights across these domains to establish a foundation for our investigation of global, value-embedded, and strategically effective AI negotiation agents that are value-diverse rather than universal responders.

6.2.1 Persuasion and Strategic Negotiation

Negotiation is a persuasive and goal-directed process. Decades of psychological research have identified how individuals influence each other through structured dialogue. Cialdini's six principles, especially reciprocity, commitment, and scarcity, have been widely adopted in commercial and organizational contexts to explain and guide influence tactics [19]. More recent work in HCI and NLP extends these principles into interactive settings. Wang et al. [130] identified key persuasive appeals and inquiries in conversational agents, such as emotional appeals, credibility references, and personal inquiries. These strategies have been shown to shape user decisions in contexts like donation and goalsetting. However, much of this work assumes culturally homogeneous interlocutors and a single, generic counterpart rather than a global set of agent personas and short-term interaction goals.

6.2.2 Culture and Negotiation Behavior

Culture shapes how negotiators frame offers, manage conflict, and calibrate trust. Dignity, face, and honor logics [11], along with frameworks like Hofstede's cultural dimensions [48, 49, 50] and Hall's framework of high-context and low-context communication [46], explain why negotiation

may be transactional in one culture but relational in another. Given the centrality of culture in communicative practice, the design of conversational agent systems should align dialog mechanics with local expectations [33] while being configurable and flexible across different situational contexts [134]. However, most existing approaches rely on surface-level adaptation, such as scripted dialogues [14], prompt engineering to localize tone or formality [12], or synthetic simulations [63]. Moreover, most current systems were found to carry Western-centric perspectives [1]. A global-agent view emphasizes that AI counterparts are value-embedded and value-diverse; what is missing is systematic evidence of how cultural alignment, or misalignment, with such agent personas affects negotiation outcomes, and whether culturally adaptive coaching can mitigate or even reverse perceived disadvantages.

6.2.3 LLM Agents in Coaching

LLMs are increasingly deployed as interactive coaches, from teacher training [2] to negotiation skill-building [111]. Feasibility is underscored by an adjacent automated coaching system that achieved strong usability with improved self-reported mood which anticipates potential benefits from integrating LLMs for conversational naturalness[99]. Indeed, LLM-based coaching systems combine experiential practice with automated feedback [2, 111], showing potential for scalable, personalized training. Yet LLM coaches currently fall short in two critical areas for teaching human negotiation: cultural nuance against value-diverse counterparts and bridging the reasoning depth gap. LLMs can sustain extended chains of reasoning with ease, while humans remain constrained. Furthermore, misaligned assistance can cause novices to drift that leaves users lost [95], widening this gap. Whether coaching can help people close this reasoning gap and build global negotiation competence—the capacity to infer an agent's value profile and adapt strategy—remains an open question.

6.2.4 Study Design

Guided by this perspective, we study human negotiation with global, value-embedded agents and evaluate whether culturally aware coaching helps users adapt across value-diverse counterparts.

In sum, persuasion in human–AI negotiation hinges on cultural grounding and reasoning effort, yet little is known about how these dynamics play out when people face AI counterparts online. Key gaps persist—we lack baselines on human performance, namely the effects of tested interventions such as coaching, and systematic evidence on how culture and reasoning shape outcomes. Addressing these gaps is essential to ensure negotiation with AI is not only efficient but also fair and trustworthy. We therefore frame our inquiry around users' global negotiation competence when interacting with value-diverse agents. This paper takes a first step toward establishing a foundation for human–AI negotiation research. Specifically, we ask:

- **RQ1:** How do humans perform against AI counterparts in negotiation?
- **RQ2a:** How does coaching relate to persuasiveness across regions?
- **RQ2b:** Why does coaching affect donation differently across regions?
- **RQ3a:** How does AI reasoning effort relate to subsequent participant persuasive effort?
- **RQ3b:** How does the AI counter player's reasoning effort differ by participant region?
- **RQ3c:** How do participants' persuasive efforts influence negotiation outcomes?

6.3 Task-Setting: The Fundraising Game as a Negotiation Environment

To examine human–AI negotiation across cultural divides, we adapt a fundraising game paradigm that builds on prior work in persuasion and negotiation tasks. In earlier studies, such as Wang et al. [130], participants were asked to persuade others to donate to charity, creating a realistic

yet controlled environment for studying persuasive strategies. Similarly, Shea et al. [111] grounded negotiation training in recognizable scenarios such as car bargaining. Building on these precedents, we situate our study in the context of fundraising, where participants act as fundraisers attempting to persuade an AI counterpart (the "donor") to contribute—with the AI modeled as a global, value-embedded agent whose cultural stance can vary (i.e., value-diverse rather than universal). To maintain control, each negotiation proceeds over a fixed number of turns with explicit offers and counteroffers, concluding in a final donation amount or impasse.

We focus on animal welfare charities, including an animal shelter, an animal farm sanctuary, and personal pet rescue, because these are ecologically valid contexts where people could frequently encounter donation requests in everyday life. Studies show that the emotions evoked by homeless or sick pets significantly increase online donations [113], while shelters and NGOs rely on storytelling, social media engagement, and gamified campaigns to mobilize support [4, 66]. By embedding negotiation into this familiar fundraising frame, we combine experimental control with ecological realism, allowing us to observe how users adapt their strategies to value-diverse agent personas, and to trace persuasive strategies and agreements in scenarios participants recognize as meaningful and socially relevant.

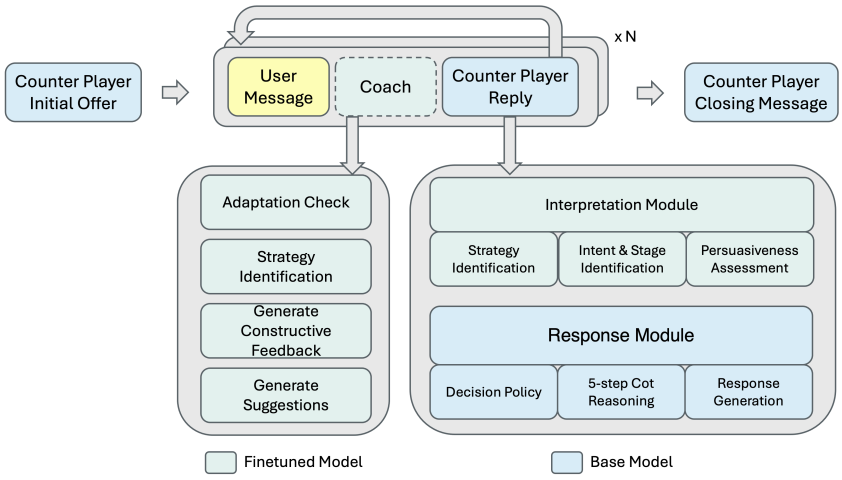


Figure 6.3: Overview of the PersuaBridge architecture. The framework integrates two LLM agents: a culturally grounded *coach* that provides real-time, turn-level feedback, and a *counter player* that simulates the donor in multi-turn negotiation. The coach performs adaptation checking, strategy identification, and culturally aware feedback generation to support user learning, while the counter player combines an interpretation module and a response module to assess persuasiveness, reason internally through a five-step chain-of-thought, and generate culturally coherent replies. In combination, the modules realize a global, culturally conditioned counterpart and support user adaptation across culturally varied agents.

6.4 The PersuaBridge Framework: AI-mediated Negotiation Training System

We now detail PersuaBridge, our hybrid system for AI-mediated negotiation training (Figure 6.3). The framework combines role-play with an AI negotiation counterpart, which takes the role of a donor that users must persuade in the fundraising game, and real-time coaching from a second agent. The counterpart is modeled as a global, value-embedded agent whose cultural stance can vary. While the counterpart enacts varied communication behaviors, the coach provides turn-level, culturally sensitive feedback to help users refine their persuasive strategies and build global negotiation competence across value-diverse counterparts.

6.4.1 Culturally Aware Coaching

For the coach to provide meaningful guidance, it must be able to (1) infer the cultural values or persuasive preferences of the counterpart, (2) recognize the persuasive strategy that the user is attempting to employ, (3) generate feedback on the effectiveness of that strategy, and (4) suggest improved, culturally attuned persuasive messages. Before these steps, the coach checks whether the user adopted the prior suggestion; if adoption is detected, it provides brief positive reinforcement.

To achieve this, the coach must recognize and reason about persuasive strategies commonly used in negotiation (Section 6.4.1). Beyond identifying strategies, the coach needs to interpret how such messages may be received across cultural backgrounds, since the perceived persuasiveness of appeals can vary widely with underlying values and norms. To enable this, we construct a dataset with human feedback that grounds persuasive strategies in culturally diverse perceptions, linking message framings to how people across different backgrounds rated their effectiveness (Section 6.4.1). Using these annotated examples, we fine-tune the coaching agent to assess why a given message might resonate or fall short, and to provide culturally aware persuasive messages that help users strengthen their negotiation moves against value-diverse global agents.

Modeling Persuasion Strategies

To scaffold users' persuasive behavior, we draw on a curated set of negotiation strategies grounded in prior theoretical and empirical work. These strategies were selected based on their relevance to negotiation, cultural adaptability, and effectiveness in AI-mediated and computer-mediated communication (CMC) contexts.

The selection draws primarily from two foundational sources: (1) Cialdini's six principles of influence [19], which are widely used in negotiation, marketing, and behavioral science; and (2) the taxonomy of persuasive strategies developed in Wang et al. [130], which categorizes appeals commonly observed in prosocial persuasive dialogue systems. Together, these frameworks provide a diverse set of persuasive tactics suitable for both reflective coaching and dynamic role-play inter-

action. The final set comprises six strategies, described below. These theory-grounded categories also serve as the coding schema in our coaching-log analyses of strategy use and coaching adoption.

Reciprocity This strategy leverages the social norm that favors should be returned. In negotiation, it functions as a way to build cooperative momentum: when one party gives something, the other is expected to reciprocate.

Commitment and Consistency This principle suggests that people are more likely to act in ways that are consistent with their previously stated beliefs or behaviors. Reminding a counterpart of earlier cooperative behavior or declared values increases the likelihood of continued collaboration.

Authority and Credibility Appeal This strategy combines two related mechanisms: people tend to defer to perceived authority figures and trust information from credible sources. Persuasive messages might refer to past experience, expert knowledge, or objective evidence to build legitimacy.

Scarcity Appeal Scarcity appeals rely on the perception that opportunities become more valuable when they are limited. In negotiation, this can motivate faster decisions or increase the perceived cost of inaction.

Emotion Appeal This strategy uses emotional expressions such as empathy, guilt, or shared sentiment to influence behavior. These appeals help build rapport and humanize the interaction, particularly when trust is not yet established.

Logical Appeal Logical appeals involve reasoned argumentation and data-driven framing. They are especially effective when players are trying to maximize gains and want assurance that cooperation is rational.

Modeling Cultural Dimensions

To generate culturally grounded persuasive messages for training and evaluation, we have constructed a dataset conditioned on two cultural dimensions from Hofstede's framework: *individualism–collectivism* and *indulgence–restraint* [48, 49, 50]. These dimensions were selected for their suitability and relevance to our task, given their well-documented influence on persuasive preferences, communicative style, and cooperation behavior in cross-cultural contexts [11, 46].

Each dimension is operationalized on a 5-point ordinal scale to enable structured, interpretable conditioning:

- **Individualism–Collectivism**: 1 = low individualism, 5 = strong individualism
- **Indulgence–Restraint**: 1 = highly restrained, 5 = highly indulgent

Dataset Creation

To populate the dataset, we first use GPT-4o [85] to generate persuasive negotiation messages. Each message is created by conditioning the model on three inputs:

(1) A randomly sampled cultural profile, defined by a value from 1 to 5 on each of the two dimensions (individualism–collectivism and indulgence–restraint)

(2) A persuasive strategy (e.g., reciprocity, emotional appeal, logical appeal)

(3) A task prompt describing the task scenario, in which a player attempts to persuade their counterpart to donate to charity

Next, we have conducted a large-scale human annotation study to evaluate the generated messages. We recruited 100 crowd-sourced annotators via CloudResearch, covering a diverse range of cultural backgrounds across 20 countries. Each annotator rated 15 messages along three axes on 1–5 scale across three dimensions:

(1) **Persuasiveness**: How persuasive the message was

(2) **Perceived Individualism–Collectivism**: How individualist or collectivist the message sounded

(3) **Perceived Indulgence–Restraint**: How indulgent or restrained the message seemed

We validated that the messages generated by GPT-4o reliably embedded the intended cultural signals, with human raters consistently perceiving the cultural orientation of each message in line with its assigned conditioning (Appendix B.1). Following the annotation, we have aggregated each rater's judgments to infer their preferred persuasive framing based on cultural values. Specifically, we examined which types of messages each rater consistently rated as highly persuasive (e.g., 4 or

5 on a 5-point scale) and analyzed the associated perceived cultural scores for those messages. By computing the mean cultural scores (on individualism–collectivism and indulgence–restraint) of the messages each rater found most persuasive, we estimate their implicit cultural preference profile.

This analysis allows us to move beyond traditional country-level cultural assumptions and instead model cultural alignment at the individual level. Rather than assigning fixed cultural values based on nationality, we infer each rater's persuasion preferences from the messages they personally rated as highly persuasive. This approach acknowledges that individuals within the same country may vary significantly in how they interpret and respond to persuasive language, particularly in cross-cultural contexts. By capturing this variation, we created fine-tuning examples that reflect individualized cultural alignment—e.g., "a person with a preference for high individualism and moderate indulgence found this message highly persuasive." These individualized annotations enable more precise conditioning of the coaching agents, improving the system's ability to adapt persuasively to diverse user profiles. In the experiment itself, participant profiles were derived from country-level Hofstede indices (Section 6.4.1). These ratings are then used to infer cultural preferences at the rater level and to identify culturally prototypical messages. From this annotated dataset, we constructed fine-tuning examples for the coach in two forms:

(1) *Feedback generation*: Assessing how persuasive a given message would be for a counterpart with a specific cultural profile

(2) *Message suggestion*: Rewriting the message to better align with the cultural values of the target profile

To ensure data coverage, we apply oversampling for rare cultural combinations. The final dataset included over 2,000 annotated examples spanning multiple cultural profiles, strategies, and interaction roles. This dataset is then used to fine-tune and evaluate culturally adaptive LLMs in coaching and negotiation settings.

6.4.2 Negotiation Counter Partner

Negotiation Interaction Design

The negotiation task is framed as a multi-turn fundraising interaction in which the user plays the role of a fundraiser attempting to persuade an AI donor to contribute resources toward the goal of maximum donation. Drawing on core principles of negotiation dynamics [36, 98], the AI counterpart initiates the dialogue with a strategically low offer—a suboptimal allocation designed to anchor expectations and motivate the user to engage in persuasive counter-negotiation. This structure encourages the participant to actively negotiate, rather than passively accept the initial terms.

The interaction unfolds over a sequence of turn-based exchanges until agreement or impasse. At each turn, users can respond with counter offers or persuasive messages, optionally guided by real-time coaching. Acceptance decisions by the AI donor are probabilistic and influenced by the persuasiveness of the user's message, which is interpreted relative to the donor's cultural profile. The overall structure simulates naturalistic negotiation patterns while maintaining tractable experimental control.

6.4.3 Negotiation Partner Architecture

Effective negotiation requires an agent to both understand the intent behind user messages and respond in a manner that reflects strategic, context-aware behavior. To support these dual demands, the AI donor is implemented as a hybrid system composed of two complementary modules: one responsible for interpreting user input and the other for generating culturally informed responses. This modular design allows us to leverage the strengths of two model variants: a fine-tuned model trained for culturally sensitive interpretation, and a prompt-based baseline model optimized for behavioral realism and flexible response generation.

1. Interpretation Module The first module draws on a fine-tuned model (see Section 6.4.1) that analyzes each user message relative to the donor's cultural profile to produce a structured

interpretation. Operating independently within the negotiation partner, this model infers the user's communicative intent, identifies the persuasive strategy being employed, determines whether an agreement is being proposed, and assigns a persuasiveness score on a 1–5 scale. The score reflects how well the message aligns with the donor's culturally conditioned value orientation—individualist donors respond more favorably to appeals centered on autonomy or impact, whereas collectivist donors are more responsive to messages invoking mutual obligation or group harmony. These outputs are not exposed to the user but are passed internally to the second module to condition the donor's subsequent decision.

2. Response Module The second module governs the AI donor's behavior and is implemented using a prompt-based approach, which is supported by prior work showing that LLMs can negotiate effectively when guided by structured prompts [35, 111], and further validated in our own survey-based evaluation (Appendix B.1). Prompt-based models produced culturally coherent behavior that was well-aligned with assigned personas, while avoiding exaggerated or overly rigid expressions, making them well-suited for behavior generation.

At each turn, the donor receives a summary of the negotiation state, including its cultural persona, game context, dialogue history, and the persuasiveness score from the interpretation module. Based on this input, it selects a response—accept, reject, or counter—with acceptance probability scaling from 10% to 90% depending on the persuasiveness score. Rejected offers are met with counter-proposals bounded to $\pm$ 10% of a reference value, enforcing gradual, realistic concessions.

The combination of fine-tuned interpretation and prompt-based response yields a culturally grounded, behaviorally plausible negotiation partner.

6.5 Experiment

We built PersuaBridge, the proposed hybrid negotiation training framework on top of GPT-4o-mini [85], which served as the negotiation counterpart throughout the study. Both the prompt-based baseline model and the fine-tuned culturally adaptive model were implemented within this

framework. The negotiation counterpart was modeled as a global, value-embedded agent whose cultural stance could vary across conditions.

6.5.1 Participants and Procedure

We recruited 80 participants (52% female, 45% male) from Prolific. Participants were drawn from two cultural pools: 40 born and residing in the United States and 40 born in East or Southeast Asian countries (South Korea 12, Japan 11, China 10, Hong Kong 4, Singapore 2, Taiwan 1). Across the full sample, 70% had completed at least a university degree.

To instantiate agent personas, we linearly mapped Hofstede's national indices to our 1–5 scale (Section 6.4.1) and derived each participant's two-dimensional profile (Individualism, Indulgence) from their country of birth. Using U.S. participants as an example, we assigned Individualism = 5 and Indulgence = 4, derived from Hofstede's national indices to represent a highly individualist and moderately indulgent orientation. The AI counterpart's cultural profile was assigned by condition: in the aligned condition, it matched the participant's scores; in the misaligned condition, it mirrored them ($\text{score}_{\text{counter}} = 6 - \text{score}_{\text{participant}}$) on each dimension. For instance, a participant with Individualism = 5 interacted with a counterpart at 5 (aligned) or 1 (misaligned). Under this operationalization, alignment manipulation pairs participants with an individualist agent persona that can feel firmer in counter-negotiation, whereas misalignment pairs them with a more collectivist persona that can feel more agreeable. We retain aligned and misaligned terminology for comparability across settings.

Each participant completed one practice round followed by two main negotiation scenarios, each comprising three rounds. The practice round used the Animal Shelter scenario to help participants familiarize themselves with the interface and negotiation structure. The two scenarios were an Animal Farm Sanctuary and a Personal Pet Rescue. The order of these two fundraising scenarios was counterbalanced across participants to control for learning and order effects.

In each round, participants played the role of a fundraiser attempting to persuade an AI donor (the GPT-4o-mini counterpart) to contribute as much as possible to the assigned charity. Negoti-

ations unfolded in multiple turns until an agreement was reached, with an average of 4.7 turns per round. Including the practice round, this yielded up to 7 rounds per participant (480 total analyzed rounds across 80 participants). After each negotiation, participants completed a short post-round questionnaire assessing perceived success, satisfaction, trust, and cultural resonance.

Because each participant contributed multiple rounds, we quantified clustering in the round-level analysis ($N = 80$, 480 trial-level observations). The intraclass correlation was ICC = 0.173, with a design effect of 1.863, indicating that 17.3% of variance was between participants.

Participants were randomly assigned to one of four between-subject conditions in a 2×2 factorial design:

- **Cultural alignment**: the AI donor was either aligned or misaligned with the participant's inferred cultural persuasion profile.
- **Coaching condition**: participants either received real-time, culturally aware coaching suggestions after each message or negotiated without coaching.

The full session lasted approximately 60 minutes. After completing all negotiation rounds, participants completed a post-study survey collecting demographic information, the 16-item horizontal and vertical individualism-collectivism scale [114], and five study-specific 3-item scales (5-point Likert) measuring persuasion self-efficacy, autonomy and coach satisfaction, trust in the AI coach, negotiation climate, and cultural alignment perception. Participants were then returned to Prolific for compensation. The study was reviewed and determined exempt by the [MASKED] Institutional Review Board (Protocol #[MASKED]). All participants provided informed consent before beginning the study.

6.6 Results

6.6.1 RQ1: How Do Humans Perform Against AI Counterparts in Negotiation?

We assessed baseline negotiation performance by comparing each round's final donation against the counter player's randomized opening offer. Participants raised the final donation above the opening in 62.7% of rounds, with a mean gain of +$3.75. The regional pattern differed sharply: Asia participants finished above the opening in 80.4% of rounds (mean gain +$6.73) compared with 45.0% for US participants (mean gain +$0.76), despite near-identical pooled opening offers (US $24.68, Asia $24.51).

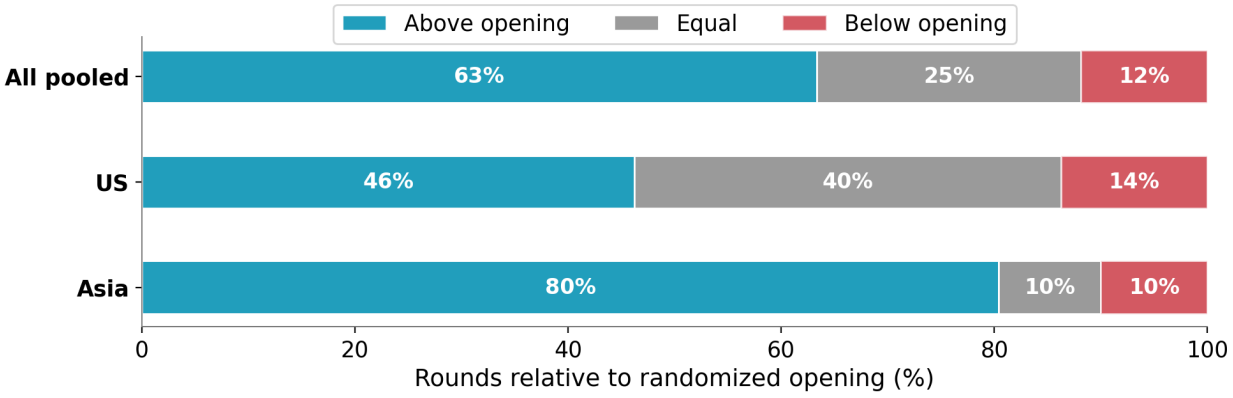


Figure 6.4: Opening-relative round outcomes by region. Asia participants finish above the opening in 80% of rounds versus 46% for US participants.

6.6.2 RQ2: How Do Coaching and Cultural Alignment Jointly Shape Outcomes?

The 35-percentage-point regional gap in above-opening outcomes is not explained by opening-offer composition alone. This gap motivated us to disentangle two pathways: whether coaching and cultural alignment shape message persuasiveness (RQ2a), and why coaching affects donation differently across regions (RQ2b).

Table 6.1: Regional coaching interaction on message persuasiveness.

Fixed effect	β	SE	95% CI	p
Intercept	**+2.03**	**0.12**	**[+1.80, +2.27]**	**¡.001*** **
Coaching	**+0.62**	**0.14**	**[+0.36, +0.89]**	**¡.001*** **
Region (Asia)	**+0.50**	**0.14**	**[+0.23, +0.77]**	**¡.001*** **
Coaching × Region	**−0.46**	**0.19**	**[−0.83, −0.08]**	**.018***

Controls for alignment, game round, and opening offer included; none significant (all $p > .38$). Full model specification in Appendix B.3.

Table 6.2: Regional coaching contrasts (derived from Table 6.1).

Derived coaching contrast	β	95% CI	p
US	**+0.624**	**[+0.357, +0.891]**	**¡.001*** **
Asia	+0.168	[−0.098, +0.434]	.2156

RQ2a: How Does Coaching Relate to Persuasiveness Across Regions?

Coaching is associated with higher message persuasiveness among US-born participants but not among Asia-born participants. We fit multilevel models with participant random intercepts to account for multiple measurements per participant, with mean persuasiveness score per round as the outcome (N = 1,613 scored messages, 80 participants). The coaching-by-region interaction is −0.46 ($p = .018$; Table 6.1): the derived coaching contrast is +0.62 ($p < .001$) for US-born but only +0.17 ($p = .22$) for Asia-born participants (Table 6.2).

Higher persuasiveness is associated with higher donations (+1.14 per score point, $p = .0001$), but coaching does not reliably increase donation after controlling for opening offer. The coaching-by-region interaction on donation is not significant (−3.97, $p = .11$), and neither regional coaching contrast reaches significance (US-born +0.60, $p = .73$; Asia-born −3.37, $p = .054$). Full model details in Appendix Tables B.3 and B.4.

RQ2b: Why Does Coaching Affect Donation Differently Across Regions?

Coaching raises persuasiveness among US-born participants but that gain does not translate reliably into final donation (RQ2a), and the direct coaching-by-region interaction on donation is marginal

(−4.47, $p = .063$; Appendix Table B.2). Because this reversal follows regional lines, we examine HC as a candidate moderator. US-born participants scored significantly higher on HC than Asia-born participants ($M = 6.87$ vs 5.93, $p = .016$), providing a plausible individual-difference pathway through which the regional coaching reversal operates.

HC is negatively associated with final donation (−3.40, $p < .001$; Table 6.3): participants with higher cooperative orientation donate less on average when uncoached. However, HC moderates the coaching-to-donation pathway. The HC-by-coaching interaction is significant (+3.58, $p = .006$): at high HC (+1 SD), coaching is associated with +$2.70 in donation; at low HC (−1 SD), coaching is associated with −$4.46. The crossover occurs at HC = 6.84, between the Asia-born mean (5.93) and the US-born mean (6.87), aligning with the regional coaching reversal.

This moderation is robust after controlling for region. When both HC-by-coaching and region are included in the same model, the regional coaching interaction attenuates by 37% (from −4.72, $p = .054$ to −2.97, $p = .25$), while HC-by-coaching remains significant ($\beta = +2.85$, $p = .021$; Table B.5, full model in Appendix B.3). Opening offer (+0.70, $p < .001$) and game round (+0.73, $p = .001$) are also positively associated with final donation.

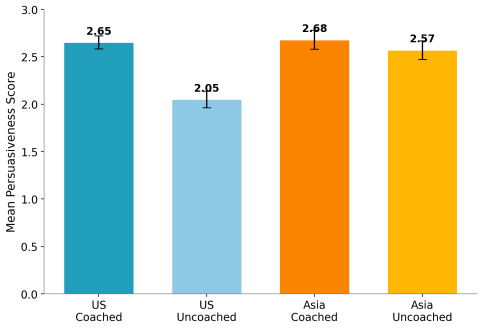


Figure 6.5: Mean persuasiveness score by region and coaching condition.

Table 6.3: HC-by-coaching moderation of final donation.

Fixed effect	β	SE	p
Intercept	**+9.05**	**1.66**	**¡.001*** **
HC	**−3.40**	**0.93**	**¡.001*** **
Coaching	−0.88	1.31	.500
HC × Coaching	**+3.58**	**1.31**	**.006** **
Opening offer	**+0.70**	**0.05**	**¡.001*** **
Game round	**+0.73**	**0.23**	**.001** **

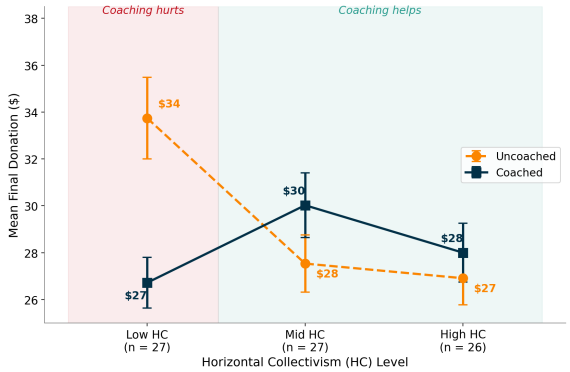


Figure 6.6: Coaching × HC interaction on final donation. At low HC, uncoached participants donate more ($34) than coached ($27); at mid and high HC, coached participants donate slightly more, consistent with the HC-by-coaching interaction ($\beta = +3.58$, $p = .006$).

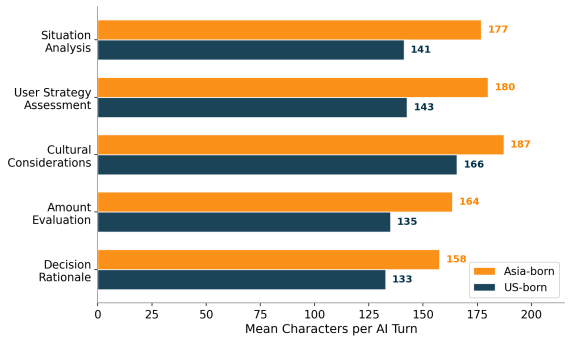


Figure 6.7: AI reasoning effort by region and CoT field. Asia-born participants consistently elicit more reasoning effort across all five fields (all $p < .001$). Only User Strategy Assessment shows a significant coaching-by-region interaction (+9.687, $p = .007$).

6.6.3 RQ3: What Is the Relationship between an AI Partner's Reasoning Effort and a Participant's Persuasive Effort?

Having established that coaching is associated with persuasiveness (RQ2a) and that HC moderates the donation pathway (RQ2b), we shift from the coaching intervention to the system's internal processing. Each AI response is generated through a structured chain-of-thought (CoT) process executed before message generation. This process decomposes the AI's internal deliberation into five reasoning components:

1. **Situation Analysis**: analyzing the current negotiation context and state.
2. **User Strategy Assessment**: evaluating the participant's persuasive approach and intent.
3. **Cultural Considerations**: reflecting on how cultural dimensions (e.g., individualism, indulgence) shape the AI's stance.
4. **Amount Evaluation**: determining why a specific donation amount is appropriate and how it should be calculated.
5. **Decision Rationale**: articulating the AI's overarching reasoning and goals behind its next

action.

This trace is never shown to participants. We operationalize both AI **reasoning effort** and participant **persuasive effort** as character count: for the AI, the character length of each CoT component captures how much elaboration the AI invested prior to responding; for participants, the character length of their visible message reflects the degree of elaboration and justification in each persuasive move. RQ3 asks whether AI reasoning effort is associated with subsequent participant persuasive effort (RQ3a), and whether participant persuasive effort is associated with donation (RQ3c).

RQ3a: How Does AI Reasoning Effort Relate to Subsequent Participant Persuasive Effort?

Only Decision Rationale is positively associated with subsequent participant message length after Holm correction across the five CoT fields: +0.386 participant characters per CoT character (Holm $p < .001$; Table B.6). The other four fields do not survive correction ($N = 1{,}161$ temporally ordered AI-response-to-next-participant-message pairs, 80 participants).

The full five-field results are in Appendix Table B.6. The Decision Rationale association is stable across both region and alignment (interaction $\beta = -0.09$, $p = .644$ for region; $\beta = +0.21$, $p = .237$ for alignment; Appendix Tables B.7, B.8). One exception: the User Strategy Assessment slope attenuates for Asia-born participants ($\beta = -0.50$, Holm $p = .011$).

RQ3b: How Does the AI Counter Player's Reasoning Effort Differ by Participant Region?

The counter player invests greater reasoning effort for Asia-born than US-born participants regardless of coaching condition (region coefficient = +136.928 characters, $p < .001$; Table 6.4). The coaching-by-region interaction is not significant (+20.858, $p = .189$). Per-field breakdown (Table B.9) shows the regional difference is broadly distributed across all five reasoning effort fields (all

Table 6.4: Region-by-coaching model of AI reasoning effort (CoT length).

Fixed effect	β	SE	95% CI	p
Intercept	**+711.83**	**13.33**	**[+685.70, +737.95]**	**¡.001*** **
Coaching	−20.80	11.39	[−43.13, +1.53]	.068
Region (Asia)	**+136.93**	**11.31**	**[+114.75, +159.10]**	**¡.001*** **
Coaching × Region	+20.86	15.89	[−10.28, +52.00]	.189
Alignment	−9.30	7.93	[−24.84, +6.25]	.241
Game round	+4.19	4.36	[−4.35, +12.73]	.336
Turn number	**+6.92**	**2.98**	**[+1.08, +12.75]**	**.020***

$p < .001$); only User Strategy Assessment has a significant coaching-by-region interaction (+9.687, $p = .007$).

RQ3c: How Do Participants' Persuasive Efforts Influence Negotiation Outcomes?

Participant message length is not significantly associated with final donation ($\beta = +0.002$, $p = .576$; $N = 480$ rounds, 80 participants). Longer messages do not reliably translate into higher donations, indicating that persuasive effort and persuasive effectiveness are distinct constructs in this dataset.

6.7 Post-Hoc Analyses and Discussion

As AI agents increasingly negotiate on behalf of organizations and individuals, how people perform in these encounters and how they can be helped to negotiate more effectively remain open questions. PersuaBridge addresses both questions through a two-agent system in which human fundraisers practice negotiating with culturally profiled AI donors while receiving private, real-time coaching on their persuasive moves. We recruited 80 participants from United States and East or Southeast Asian cultural backgrounds into a 2×2 design crossing real-time coaching (with vs. without) with cultural-profile alignment (matched vs. mismatched to the AI donor), and each participant negotiated six rounds.

6.7.1 RQ1: Regional Differences in Negotiation Performance

The same AI negotiation setting produced sharply unequal outcomes for participants from different cultural backgrounds, before any coaching effect is considered. Participants from East or Southeast Asian cultural backgrounds achieved above-opening donations in 80.4% of rounds, compared with 45.0% for US-born participants. This 35-percentage-point gap indicates that the impact of AI negotiation on user performance is not uniform across cultural backgrounds. One explanation for this variation is that people from different cultural backgrounds interpret exchange and obligation differently, bringing different norms for request-making, concession, and relational meaning into formally similar bargaining situations [11, 34]. Culture is located "in the world, in patterns of ideas, practices, institutions, products, and artifacts" rather than as a stable set of beliefs residing inside individuals, which means the regional gap we observe reflects context-linked variation in how participants approach the task rather than a fixed personal trait [69].

6.7.2 RQ2a: How Coaching Relates to Persuasiveness Across Regions

Coaching's association with persuasiveness showed a sharp regional gap. US-born participants gained significantly in persuasiveness (+0.62, $p < .001$), while Asia-born participants showed a smaller, non-significant gain (+0.17, $p = .22$), and the regional interaction was significant (-0.46, $p = .018$). This proximal gain did not translate directly into higher donations. Persuasiveness itself predicted donation outcomes (+1.14 per score point, $p < .001$), confirming that more persuasive messages matter for negotiation results. Yet coaching did not significantly predict donation once persuasiveness was in the model (US $p = .73$; Asia $p = .054$), suggesting that coaching's contribution passes through persuasiveness rather than operating on donation through a separate pathway. The regional pattern mirrors the RQ1 performance gap. Asia-born participants already achieved above-opening donations in 80% of rounds, while US-born participants succeeded in only 45%. Coaching appears to improve message formulation where participants' existing approach leaves the widest gap. This raises a follow-up question. If coaching helps proximal persuasiveness where room

exists, what determines whether that improvement translates into actual donation gains? The next analysis identifies an individual-level cultural orientation that moderates this translation.

6.7.3 RQ2b: Why Coaching Affects Donation Differently Across Regions

Individual cultural orientation offers a more specific account of coaching heterogeneity than a regional label alone. The coaching-to-donation association varied by participants' cooperative orientation rather than their country of origin. Horizontal collectivism (HC) moderated the coaching-donation pathway. HC measures a person's tendency to view the self as part of a cooperative in-group while valuing equality among members, as distinct from vertical collectivism, which emphasizes group membership through hierarchy and deference to authority [114]. Coaching was positively associated with donation at high HC (+\$2.70 at +1 SD) and negatively associated at low HC (−\$4.46 at −1 SD), with the crossover occurring between the Asia-born mean (5.93) and the US-born mean (6.87). HC has been validated across cultural psychology as a distinct dimension of collectivism that captures cooperative group orientation independently of hierarchical deference, and the horizontal versus vertical distinction predicts different behaviors within the same national sample, meaning individual-level HC measurement adds information beyond a country-level classification [109, 114, 121]. In our sample, US-born participants scored significantly higher on HC than Asia-born participants ($M = 6.87$ vs. 5.93, $d = 0.55$, $p = .016$), aligning with a prior cross-national comparison that found the same directional difference between US and South Korean samples [73]. The relevant design question is not whether a regional label predicts coaching benefit, but whether the participant's cooperative orientation changes how advice is received and enacted in a negotiation.

Future culturally aware coaching systems should measure or infer participant orientation rather than relying on a coarse regional proxy. Culture is constituted through practices and situations rather than residing as a stable set of beliefs inside people, which means individuals within a broad region vary in self-construal [69]. Independent and interdependent self-construal vary within any cultural group and can be situationally activated, meaning a person's orientation at the moment

of negotiation may differ from their group's average [39]. Because cooperative orientation varies within regions and shift, across situations, any personalized AI system that assigns interaction strategies based on a fixed regional profile will misfit users whose current stance differs from the regional average. For coaching systems like PersuaBridge, this means eliciting or inferring the user's communicative orientation and evaluating whether strategy choices improve recipient response, rather than assuming that a regional assignment determines interaction fit.

6.7.4 RQ3: AI Reasoning Effort and Participant Persuasive Effort

The AI counter player's reasoning effort was selectively associated with the participant's subsequent persuasive effort, but only through one specific reasoning component. We measured AI reasoning effort through the character length of five hidden chain-of-thought fields and participant persuasive effort through the character count of the next visible message. Most associations did not survive multiplicity correction. Decision Rationale was the only chain-of-thought field whose association with the next participant message's character count survived both correction families (+0.386 characters per CoT character, Holm $p < .001$). This specificity suggests that the coupling between system processing and human response is not general but localized to one component of the AI's deliberation.

The counter player was associated with longer reasoning traces for Asia-born participants across all five internal fields regardless of coaching condition. Prior work has found that conversational agents exhibit different interaction behaviors depending on the user's cultural profile, with communication patterns varying across cultural dimensions [16].

Participant effort toward elaboration did not translate into better negotiation outcomes. It aligns with prior CMC findings in which communication form and task outcomes do not always move together [105]. For future AI-mediated negotiation systems, these results suggest that the quality and strategic fit of guidance matters more than prompting participants to write longer messages. Whether the system provides real-time coaching, post-session feedback, or autonomous negotiation on the user's behalf, the design target should be strategy selection, requested-amount

calibration, timing, and repair opportunities rather than message elaboration.

Across all RQ3 models, static cultural-profile alignment did not reliably moderate any tested relation. Prior intercultural research has found that assigning cultural profiles to interaction partners does not automatically change how they communicate [126], which is consistent with the null alignment effects we observe. These results suggest that fixed profile matching is insufficient as a design strategy, and future systems should explore participant-selected or turn-contingent adaptation instead.

6.8 Implications for Human-AI Negotiation Support

PersuaBridge's staged evidence positions culturally aware human-AI negotiation support as a multi-level design problem rather than a single-outcome optimization. The system shows where coaching enters (message formulation) and where it becomes conditional (donation depends on cooperative orientation, not just coaching presence). Grounding theory argues that a contribution requires both a presentation phase and evidence of acceptance by the recipient before it is complete [20]. Future deployments should separately log each transition in the coaching pathway so that designers can distinguish whether an improved message was presented, whether the AI donor recognized it, and whether that recognition translated into a different outcome. This observability would prevent the persuasiveness score from carrying too much interpretive weight when the final donation is produced by later decision rules.

This study contributes the first empirical separation of coaching entry points, cultural-orientation moderators, AI reasoning effort, and static-profile alignment within a single human-AI negotiation testbed. By showing that coaching improves proximal persuasiveness but that its translation to donation outcomes depends on the participant's cooperative orientation rather than a regional label, PersuaBridge reframes culturally aware AI support as a personalization problem grounded in individual differences rather than demographic categories. The alignment null across all tested models further demonstrates that fixed cultural profiles are not sufficient to shape interaction dynamics,

motivating a shift toward participant-inferred and turn-contingent adaptation. As AI-mediated negotiation expands into sales, customer service, fundraising, and cross-border diplomacy, systems that expose each transition from advice to outcome and adapt to the user's current communicative stance will be better positioned to improve human performance than systems that optimize a single aggregate metric. As AI becomes embedded in everyday interaction, systems like PersuaBridge can help enable more equitable, personalized, and culturally responsive human–AI communication, provided they are built with appropriate transparency and safeguards.

Chapter 7

General Discussion

7.1 Summary of Findings

This dissertation develops a communicative process design approach for human-AI team collaboration, synthesized across four studies in two task domains.[1] The preceding chapters ask how communicative processes are shaped by culture, team structure, and designed support, and they answer through four studies spanning two task domains, brainstorming and negotiation. Across these studies, the convergent finding is that communicative processes in human-AI teams are moderated rather than uniform: the same design produces different dynamics for different users. This chapter first traces how the dissertation's three stages, characterizing communicative dynamics, examining how team structure reshapes them, and testing whether designed support helps diverse users, build on each other cumulatively, and then develops the design implications of this convergence.

Accommodation, the tendency for interlocutors to adjust their communicative styles toward one another, varies by cultural population: AI agents that accommodate well on average may fail specific cultural groups (Chapter 3). Within negotiation, affective accommodation (turn-to-turn

[1] Part of the content in this chapter is based on a paper previously published: Rong-Ching Chang and Hao-Chuan Wang. "From Unilateral Behavioral Tuning to Human–AI Co-Configuration." *ACM CHI 2026 Workshop on Human-AI Interaction Alignment: Designing, Evaluating, and Evolving Value-Centered AI For Reciprocal Human-AI Futures*, 2026.

sentiment convergence) and grounding (increasing coordination efficiency over time) operate as distinguishable processes, establishing that these dynamics are multidimensional (Chapter 4). The effect of behavioral formation, the way the AI's communicative behavior is configured, on human authorship depends on team size: formation shapes how the human contributes with one AI partner, but a second AI erases that formation effect (Chapter 5). Whether coaching translates from a message-level improvement to a task outcome such as donation amount depends on individual cooperative orientation, not regional category (Chapter 6). The shift from population-level characterization to individual-level moderation is deliberate: characterization identifies the variation, but the intervention study shows the operative moderator is individual orientation rather than population membership.

These moderated effects are not simply weaker or stronger versions of the same dynamic; they are qualitatively different communicative modes. Without design intervention, humans and AI accommodate toward each other's styles over time, but the pattern itself varies across populations. When team structure is introduced, the AI's role shifts from co-generation to coaching depending on how it is configured, and the human responds with correspondingly different engagement patterns: more board items under co-generation, more deliberative discussion under coaching. When active coaching is added, only specific components of the AI's reasoning couple with subsequent human behavior, suggesting that the connection between AI support and human performance is selective rather than global. In each case, the shift is a redistribution of communicative roles rather than a uniform amplification of human performance.

A methodological contribution of the dissertation is the measurement architecture required to capture these dynamics. The studies introduce separations between human-authored and AI-authored contributions (Chapter 5), message-level gains such as persuasiveness from task outcomes such as donation amount (Chapter 6), accommodation from grounding within the same sample (Chapter 4), and global diversity (how different the ideas are from what others in the population produced) from local diversity (how different the ideas are from each other within the session) (Chapter 5). Aggregate metrics conceal trade-offs that the communicative process design approach

makes visible: group output rises while human authorship falls (Chapter 5), and coaching improves message quality without automatically transferring to task outcomes (Chapter 6).

7.2 The Three-Stage Cumulative Structure

Characterizing communicative dynamics, examining how team structure reshapes them, and testing whether designed support helps diverse users are not independent studies but a cumulative argument about how communicative processes in human-AI teams should be understood, structured, and designed. Each stage depends on what the prior stage established: characterization identifies which communicative dynamics vary and along which dimensions, giving formation a target to manipulate rather than a uniform process to amplify. Formation in turn reveals that structural choices create qualitatively different communicative modes, but these modes are set before the session begins and cannot adapt to the variation that characterization uncovered. Intervention addresses this limitation by introducing real-time support, yet the support's reach is bounded by the same population variation, closing the loop between the first and third stages.

The population variation identified in the characterize stage does not remain descriptive context; it returns as the moderation that determines who benefits from designed communicative support. The accommodation patterns that Chapter 3 identifies as culturally structured and Chapter 4 separates into distinct constructs reappear in Chapter 6, where it is horizontal collectivism, a stable individual-level cooperative orientation, that moderates whether coaching reaches the task outcome. The shift from regional to individual orientation suggests that communicative process design should infer cooperative disposition from interaction rather than assign support by demographic group. This pattern parallels a broader principle from the group brainstorming literature: cultural diversity is a resource for ideation, but the resource does not become productive without designed support that makes hidden information visible and navigable [129].

These patterns emerge across two distinct task domains, brainstorming and negotiation, with different team configurations and participant pools. In brainstorming, formation determines whether

AI co-generates or coaches, and a second AI erases the formation effect. In negotiation, coaching improves message quality where the performance gap is widest, but translation to task outcomes is moderated by cooperative orientation. Despite different tasks, the same structural logic holds: the communicative process, not just the AI's output, shapes the human outcome.

7.3 From Initial Parameterization to Iterative Co-Configuration

The dissertation's studies bracket a design frontier between two approaches to configuring AI behavior. In the first approach, initial parameterization, the AI's behavior is set before the session and remains fixed throughout. In the second approach, iterative co-configuration, both human and AI revise their communicative strategies during the session through ongoing negotiation of goals, constraints, and coordination rules. User-parameterized formation in the AddOne study (Chapter 5) exemplifies initial parameterization: the participant sets the AI's competitive orientation, evaluation apprehension, and social loafing levels through a structured walkthrough before the session begins, and the AI maintains that profile throughout. This one-time setup creates a qualitatively distinct communicative mode, the coaching profile, establishing that a single configuration step can reshape the team's communicative dynamics. PersuaBridge's coaching (Chapter 6) extends this by delivering real-time support during the task, showing that designed communicative support can improve message-level performance even with a fixed coaching strategy. Together, these studies map the design space between fixed parameterization and iterative co-configuration, and the moderated effects observed across studies establish that making this adjustment bidirectional and ongoing is the central design frontier for the field.

The dissertation proposes iterative co-configuration as the next step for human-AI team design: ongoing, bidirectional adjustment of communicative processes during collaboration. Where initial parameterization fixes the AI's behavior before the task, co-configuration would make goals, constraints, and communicative strategies revisable during the session, drawing on the kind of evidence-based coordination that the theory of grounding describes [20]. Under co-configuration,

alignment becomes an interaction property that is maintained, monitored, and revised dynamically over time, with user-verifiable evidence rather than guesswork.

To explain why current alignment practices fall short of this goal, we propose the iceberg framework (Figure 7.1), a conceptual architecture that separates visible AI behavior from the hidden layers where misattribution occurs. In human collaboration, mutual understanding rests on Theory of Mind: the capacity to attribute beliefs, intentions, and desires to a partner and to update those attributions during interaction [118]. People routinely extend these social attribution practices to AI systems, with or without believing that the system has feelings [80], and conversational LLMs intensify this tendency by producing fluent, responsive language that invites rich mental-state inferences [30]. The iceberg separates three layers: above the waterline is observable AI behavior, what the system says and does in interaction; below are two layers that users and designers often conflate, users' assumptions about the AI's stable beliefs, intentions, and values, and what current LLMs actually provide, which is next-token prediction and training-induced patterns that can approximate stances without stable or inspectable commitments [107]. This interaction is fundamentally asymmetric: systems access human action through a narrow channel, while people interpret the interaction through their own rich, situated contexts [119]. Current alignment practices such as reinforcement learning from human feedback (RLHF) and safety fine-tuning work primarily above the waterline by shaping outputs, while users read those outputs as evidence of internal states. The mismatch widens as AI behavior becomes more fluent, because better surface behavior invites stronger attributions that the system's underlying mechanisms do not warrant [87]; anthropomorphic attribution can inflate perceived capability and skew trust judgments [94]. The dissertation's own findings illustrate this asymmetry: as conversations unfold, LLMs converge with users on surface linguistic features while drifting on psychological and contextual dimensions (Chapter 3) [17], so interaction looks increasingly aligned above the waterline even as deeper convergence fails to develop.

Building on this framework, we propose three interaction design paradigms (Figure 7.2) that trace the trajectory from current practice toward co-configuration. Paradigm 0, free-form interac-

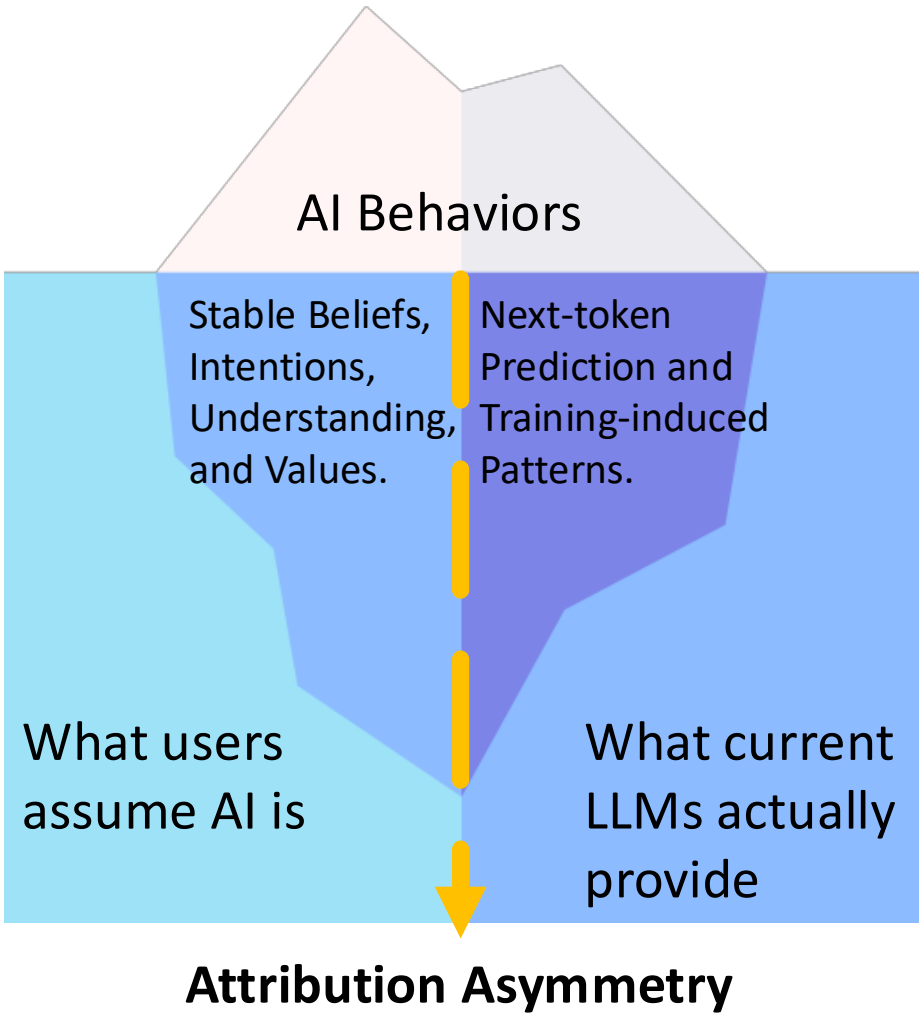


Figure 7.1: The iceberg framework contrasts visible AI behavior with hidden user assumptions and model processes, creating asymmetric inference where AI can often infer users better than users can infer AI commitments.

tion, provides no scaffolding for sustained joint work: users issue requests, models respond, and assumptions remain implicit. Paradigm 1, behavioral engineering, represents the current dominant approach: optimizing AI outputs to appear helpful and honest via RLHF and model specifications. Yet persistent failure patterns reveal structural limits with this approach: agreement bias (prioritizing user approval over accuracy), excessive refusal (blocking benign requests based on surface keywords), and gaps between stated values and actual behavior [103, 108, 112] all indicate that polished behavior can invite attributions the system cannot sustain. Behavioral engineering reproduces the planning-model error: treating alignment as a specification to encode rather than as situated action worked out in practice [119]. Paradigm 2, human-AI co-configuration, shifts from unilateral tuning to transparent, restructured interaction through co-configurable agreements where both parties shape goals, constraints, and ambiguity-handling strategies during use. The dissertation's studies occupy the transition between these paradigms: the accommodation analyses (Chapter 3) operate within Paradigm 0, characterizing communicative dynamics in free-form interaction without designed scaffolding; PersuaBridge's coaching (Chapter 6) operates within Paradigm 1, delivering real-time support through a fixed strategy set by the designer; while AddOne's CUS formation (Chapter 5) begins to approach Paradigm 2 by letting the user configure the AI's behavioral expectations before the session, though the configuration is one-time rather than iteratively revisable during collaboration.

We identify three design opportunities that co-configuration opens, each grounded in the dissertation's empirical findings. Calibration via interaction indicators would use interaction history across repeated collaborations to reveal misalignment and signal when system behaviors could be misperceived; the population variation from the characterize studies (Chapters 3 and 4) motivates such indicators, because accommodation patterns that hold on average may fail specific groups. Co-configurable interaction contracts would make task-level commitments explicit and revisable, including goals, evidence thresholds, and fallback policies, so that coordination rests on negotiated agreements rather than inferred intents; the role redistribution from the formation study (Chapter 5) motivates such contracts, because formation determines whether the AI co-generates

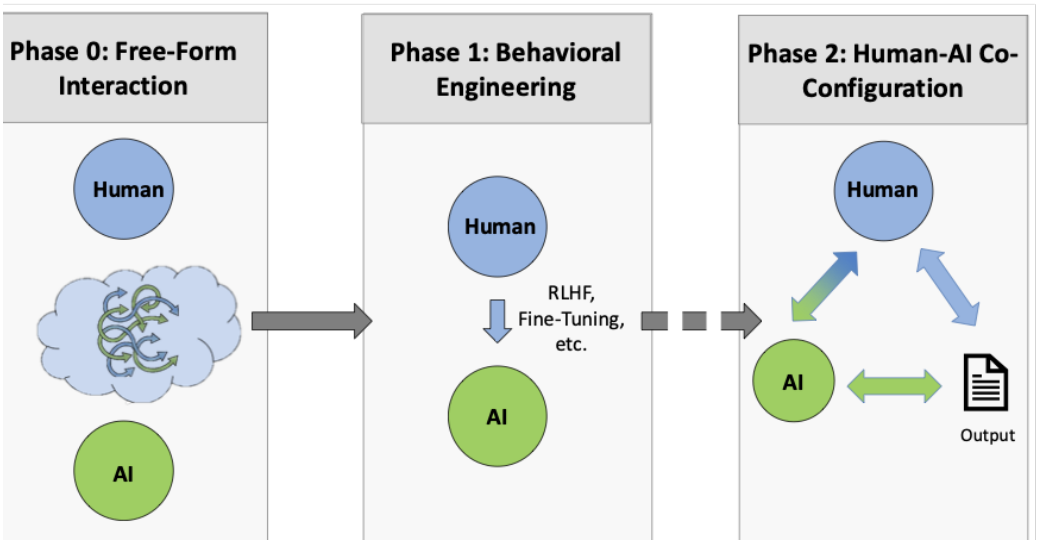


Figure 7.2: Three interaction design paradigms for human-AI interactions, from free-form interaction through behavioral engineering to human-AI co-configuration.

or coaches and the human's communicative role shifts accordingly. Shared protocols for trust recalibration would establish how both user and system respond when goals, confidence, or context shifts; the moderated coaching from the intervention study (Chapter 6) motivates such protocols, because horizontal collectivism, not regional membership, determines whether coaching reaches the task outcome, and recalibration should adjust support based on cooperative orientation rather than demographic category.

7.4 General Limitations

Several scope conditions qualify these findings. The accommodation analysis in Chapter 3 draws on the PRISM dataset, 1,500 users across 75 countries in naturalistic conversations with deployed models, whose cultural representation reflects the distribution of users on the platform. The affective accommodation and grounding analysis in Chapter 4 draws on 40 participants each completing seven negotiation rounds, a sample sized to ground the construct separation in detailed turn-level analysis of a specific task context. Both the AddOne and PersuaBridge studies draw participants

from online platforms (Cloud Research and Prolific, respectively), consistent with the controlled experimental designs that allow the factorial comparisons reported here.

The findings are tied to specific LLMs: the PRISM dataset in Chapter 3 reflects interactions with a collection of models, AddOne uses GPT-4o, and PersuaBridge uses GPT-4o-mini. As models evolve in their communicative capabilities, accommodation patterns and coaching dynamics may shift, and the specific effect sizes reported here establish baseline benchmarks against which future model generations can be compared. Because the effects documented here, surface-level accommodation, formation-dependent role shifts, and moderated coaching, all operate through the communicative channel that stronger models are designed to improve, more capable models are likely to amplify rather than eliminate these dynamics: more fluent surface behavior invites stronger attributions (the iceberg mechanism from Section 7.3), and more responsive coaching widens the gap between message-level gains and task outcomes that horizontal collectivism moderates. Moreover, the dissertation provides a systematic measurement architecture for benchmarking these dynamics across model generations: the separations between accommodation and grounding (Chapter 4), human-authored and AI-authored contributions (Chapter 5), and message-level gains and task outcomes (Chapter 6) define reusable constructs and metrics that apply regardless of which model powers the interaction.

The two task domains, brainstorming and negotiation, do not cover the full range of collaborative work. Whether the convergent patterns generalize to extended multi-session collaboration, high-stakes decision-making, or creative co-authoring remains an open question. The cross-domain convergence observed here establishes evidence for the generality of the communicative process design approach, and the convergent patterns across these domains motivate extending the progression to further task contexts.

The current studies establish the empirical conditions under which co-configuration becomes a meaningful design goal. The AddOne CUS walkthrough demonstrates that a single configuration step reshapes communicative dynamics, and PersuaBridge's coaching shows that real-time support reaches task outcomes selectively. Co-configuration, in which communicative processes become

revisable during the session through ongoing mutual grounding, is the design frontier that these findings open.

7.5 Future Directions

Implementing co-configuration requires moving beyond single-session designs toward systems that learn from interaction history and adapt communicative support in real time. A co-configurable system would allow the participant to set initial parameters, as in the AddOne CUS walkthrough, and then adjust them through ongoing interaction, with the system surfacing indicators when communicative processes drift from the negotiated agreement. The design opportunities outlined in Section 7.3 provide a starting architecture, but each would need to be tested empirically against the same kind of cross-cultural, cross-task evaluation that this dissertation applies to fixed designs.

The current studies capture single sessions, but communicative processes evolve across repeated collaborations. Longitudinal studies that track how accommodation patterns, role distributions, and coaching effects develop over multiple sessions would inform the calibration indicators that co-configuration requires. The longitudinal convergence observed in Chapter 3, where surface linguistic features align while deeper dimensions drift, suggests that calibration must operate at multiple timescales.

Extending the cultural scope beyond the US-born and Asia-born cohorts would test whether the horizontal collectivism moderation observed in Chapter 6 holds across broader global populations. The cultural variation in accommodation patterns from Chapter 3, spanning dozens of countries, provides a starting point, but intervention studies with similarly diverse samples are needed to determine whether the characterize-before-you-design principle scales to populations that current studies have not reached.

7.6 Conclusion

This dissertation articulates a communicative process design approach for human-AI team collaboration. Four studies across two task domains establish three convergent findings. Communicative processes in human-AI teams are structured and moderated by population, team structure, and individual orientation. Human communicative behavior shifts in the presence of AI partners through role redistributions rather than uniform amplification. Single-metric evaluation is insufficient because aggregate measures conceal trade-offs that only separating authorship, conversation, and quality reveals.

The progression from characterizing communicative dynamics through examining how team structure reshapes them to testing whether designed support helps diverse users demonstrates that population variation identified in early characterization returns as the moderation that determines who benefits from designed communicative support, suggesting that effective design needs to reckon with this variation rather than assume a universal user. Culturally aware systems should infer cooperative orientation from interaction rather than assign support by demographic category. Each team configuration must be evaluated as a separate design case because they produce distinct communicative dynamics. Methodologically, the approach introduces measurement schemes that separate constructs previously confounded in human-AI interaction research: human from AI authorship, affective accommodation from affective grounding, global from local diversity, and message-level gains from task outcomes.

The gap between initial parameterization and iterative co-configuration remains the central design frontier. The dissertation's studies show what communicative processes look like under fixed designs; the co-configuration paradigm proposed here outlines what it would take to make those processes revisable during collaboration. This dissertation contributes to the understanding of communicative processes in human-AI team collaboration and to the design of systems that account for population variation, and it opens new questions: how co-configuration can be implemented so that communicative support adapts during collaboration, how calibration indicators can surface

the mismatch between surface alignment and deeper convergence, and how the characterize-before-you-design principle scales to broader populations and task domains.

Appendix A

Supplementary Tables for Chapter 5: AI Team Size and Behavioral Formation

A.1 Complete OLS Models Predicting Human Board Items

We estimated separate OLS regressions predicting human-authored board items per person, each entering one AI message-act category alongside AI board volume, group size, and formation as controls (N=120). Tables A.1–A.4 report the full coefficient tables.

A.1.1 Social/Filler Model

Table A.1: OLS model predicting human-authored board items with AI social/filler share as the focal predictor (R^2=.338).

Term	β	SE	t	p
Intercept	23.01	4.09	+5.63	**<.001*** **
AI social/filler share	+26.11	11.81	+2.21	**.029***
Group size (number of AI partners)	+0.56	2.57	+0.22	.827
Formation (CUS vs. AVG)	+1.26	2.11	+0.60	.551
AI-authored board items	-0.271	0.073	-3.73	**<.001*** **

A.1.2 Generative Model

Table A.2: OLS model predicting human-authored board items with AI generative share as the focal predictor (R^2=.315).

Term	β	SE	t	p
Intercept	33.61	4.16	+8.08	**<.001*** **
AI generative share	-6.59	6.59	-1.00	.319
Group size (number of AI partners)	-1.50	2.42	-0.62	.537
Formation (CUS vs. AVG)	-3.23	2.16	-1.50	.137
AI-authored board items	-0.272	0.075	-3.61	**<.001*** **

A.1.3 Evaluative Model

Table A.3: OLS model predicting human-authored board items with AI evaluative share as the focal predictor (R^2=.317).

Term	β	SE	t	p
Intercept	30.59	2.49	+12.29	**<.001*** **
AI evaluative share	-12.04	10.49	-1.15	.254
Group size (number of AI partners)	-2.21	2.21	-1.00	.319
Formation (CUS vs. AVG)	+0.61	2.66	+0.23	.818
AI-authored board items	-0.265	0.074	-3.59	**<.001*** **

A.1.4 Questioning Model

Table A.4: OLS model predicting human-authored board items with AI questioning share as the focal predictor (R^2=.311).

Term	β	SE	t	p
Intercept	29.80	2.70	+11.04	**<.001*** **
AI questioning share	+2.49	5.61	+0.44	.658
Group size (number of AI partners)	-2.36	2.23	-1.06	.292
Formation (CUS vs. AVG)	-2.13	1.79	-1.19	.236
AI-authored board items	-0.263	0.075	-3.49	**<.001*** **

All coefficients are unstandardized. The reference category for formation is AVG; group size is coded as the number of AI partners (1 or 2). AI-authored board items is a count; message-act shares are proportions ranging from 0 to 1. Significance levels: *** p<.001, ** p<.01, * p<.05.

Appendix B

Supplementary Materials for Chapter 6: Multi-Agent Coaching

B.1 Validating Agent Culturalization and Value Embedding via Survey Comparison

To validate the cultural alignment of our models, we conducted a controlled survey-based comparison between the baseline and finetuned models under different cultural assignments. Specifically, we used the Collective Orientation Scale from the Fetzer Institute's validated self-report measures [28], which includes subscales for both individualism and collectivism.

We administered the survey to each model under two conditions: one prompting for high individualism (Idv=5) and another for low individualism (Idv=1). Each model completed the full survey five times per condition, and responses were averaged across runs. All responses were collapsed across the 16 survey items to produce a single average per cultural dimension.

As shown in Figure B.1, both the baseline and finetuned models reflected the expected shifts in cultural orientation. When assigned a low individualism score (Idv=1), both models displayed more collectivist tendencies; when assigned high individualism (Idv=5), both displayed stronger individualist responses. The baseline model showed more polarized behavior across both directions,

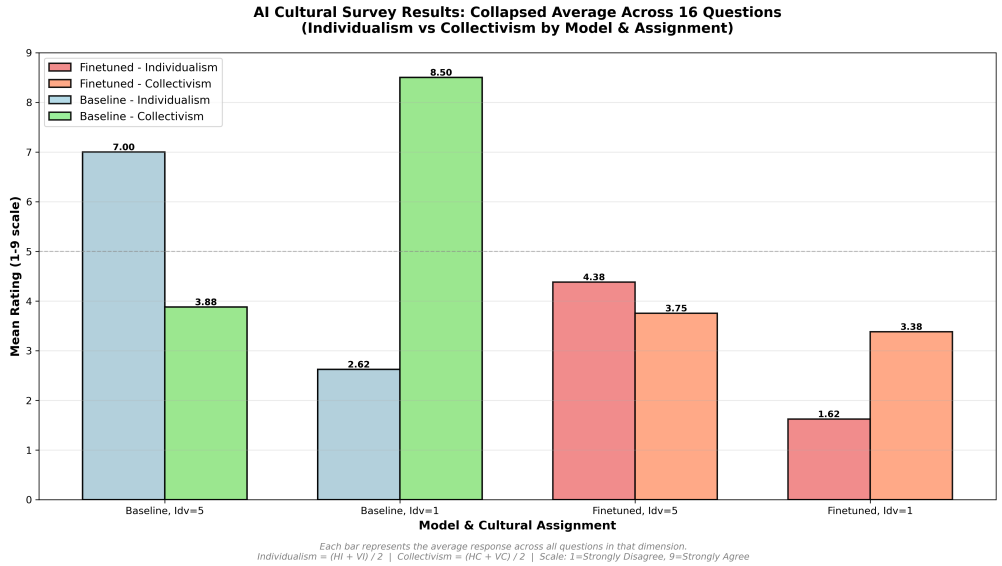


Figure B.1: AI Cultural Survey Results comparing baseline vs. finetuned models under individualistic vs. collectivist role assignments. Ratings reflect average survey response on a 1–9 scale.

particularly under Idv=1 (e.g., strong collectivism: 8.50), whereas the finetuned model exhibited more balanced and moderated responses across conditions.

This pattern is consistent with our design: the finetuned model acts as a cultural expert, capable of reasoning across perspectives rather than rigidly adhering to one. Its calibration reflects a broader understanding of human persuasion preferences and a more differentiated representation of cultural variation.

B.2 Validating Cultural Value in Generated Messages via Human Ratings

To validate whether the cultural conditioning embedded in GPT-4o's outputs aligned with human perception, we conducted two mixed-effects linear regressions. Specifically, we tested whether the original GPT-assigned cultural scores (used during message generation) significantly predicted how human raters perceived the cultural orientation of each message. For each regression, we used a mixed-effects model with a random intercept for rater ID to account for individual variability in

interpretation. Results showed a significant positive effect ($\beta = 0.35$, $p < .001$), indicating that higher GPT-assigned individualism levels corresponded to higher perceived individualism by raters. In the second model, we examined GPT-assigned indulgence scores and found a similarly significant positive effect on perceived indulgence ($\beta = 0.20$, $p < .001$). These findings confirm that messages generated by GPT-4o reliably embedded the intended cultural signals, and that raters consistently interpreted them in line with their assigned conditioning. This validation step supports the use of these messages and rater ratings as a basis for inferring individual-level cultural preferences and generating culturally calibrated fine-tuning examples. Moreover, the significant between-rater variance observed in both models suggests meaningful individual differences in cultural interpretation.

B.3 Supplementary Model Tables

Table B.1: Regional coaching interaction on message persuasiveness. The outcome is the AI-evaluator persuasiveness score. The model includes coaching condition, participant region, their interaction, alignment condition, game round, and the counter player's opening offer as fixed effects, with participant random intercepts (N = 1,613 scored messages, 80 participants).

Fixed effect	β	SE	95% CI	p
Intercept	+2.034	0.122	[+1.795, +2.272]	$< .001$
Coaching	+0.624	0.136	[+0.357, +0.891]	$< .001$
Region (Asia)	+0.501	0.136	[+0.234, +0.768]	.0002
Coaching × Region	−0.456	0.192	[−0.833, −0.079]	.018
Alignment	−0.083	0.096	[−0.271, +0.106]	.390
Game round	+0.004	0.010	[−0.014, +0.023]	.642
Opening offer	+0.002	0.002	[−0.002, +0.006]	.387

Table B.2: Regional coaching interaction on final donation, anchor-adjusted. The model controls for the counter player's opening offer, alignment condition, and game round ($N = 480$ rounds, 80 participants).

Fixed effect	β	SE	p
Intercept	+4.467	1.834	.015
Coaching	+1.418	1.698	.403
Region (Asia)	+7.599	1.698	$< .001$
Coaching × Region	−4.465	2.404	.063
Alignment	+0.386	1.199	.747
Opening offer	+0.774	0.044	$< .001$
Game round	+0.441	0.213	.039

Table B.3: Final donation conditional on current-message persuasiveness. The outcome is the round's final agreed donation amount, repeated on each scored participant-response row. This message-weighted specification controls for the persuasiveness score, within-round response order, and the counter player's opening offer ($N = 1{,}613$ response rows covering 480 rounds, 80 participants).

Fixed effect	β	SE	95% CI	p
Intercept	+1.157	1.708	$[-2.191, +4.505]$	.498
Coaching	+0.603	1.765	$[-2.856, +4.063]$	.732
Region (Asia)	+7.158	1.758	$[+3.711, +10.604]$	$< .001$
Coaching × Region	−3.974	2.481	$[-8.837, +0.889]$	.109
Alignment	+0.652	1.238	$[-1.775, +3.079]$	.599
Game round	+0.470	0.111	$[+0.253, +0.688]$	$< .001$
Response index	+0.281	0.180	$[-0.071, +0.634]$	.118
Opening offer	+0.781	0.023	$[+0.736, +0.827]$	$< .001$
Persuasiveness score	+1.142	0.294	$[+0.565, +1.719]$	.0001

Table B.4: Regional coaching contrasts conditional on persuasiveness, derived from Table B.3.

Derived coaching contrast	β	95% CI	p
US	+0.603	$[-2.856, +4.063]$	.732
Asia	−3.371	$[-6.796, +0.055]$	.054

Table B.5: HC × coaching interaction with region control. Participant region is added as a covariate alongside the horizontal collectivism moderation ($N = 480$ rounds, 80 participants).

Fixed effect	β	SE	p
Intercept	+6.269	1.682	$< .001$
HC	−2.324	0.907	.010
Coaching	−0.843	1.211	.486
HC × coaching	+2.847	1.232	.021
Region (Asia)	+4.413	1.277	$< .001$
Opening offer	+0.775	0.044	$< .001$
Game round	+0.440	0.213	.039

Table B.6: Internal AI chain-of-thought fields predicting next-human-message length. Each row is the focal coefficient from a separate multilevel model predicting the character count of the next visible participant message from one CoT field's character count, participant region, alignment, and turn number, with participant random intercepts ($N = 1{,}161$ pairs, 80 participants). Holm correction applied within the five-test family.

CoT field	β	SE	95% CI	Raw p	Holm_5 p
Situation Analysis	+0.114	0.082	[−0.046, +0.275]	.163	.164
User Strategy Assessment	+0.188	0.079	[+0.034, +0.343]	.017	.068
Cultural Considerations	+0.136	0.078	[−0.017, +0.290]	.082	.164
Amount Evaluation	+0.188	0.087	[+0.018, +0.357]	.030	.091
Decision Rationale	**+0.386**	**0.092**	**[+0.206, +0.566]**	**< .001**	**< .001**

Table B.7: CoT field × Region interactions on next-human-message length. Five-test Holm family. Same specification as Table B.6 with additive control replaced by interaction.

CoT field × Region	β	SE	95% CI	Raw p	Holm p
Situation Analysis × Region	−0.234	0.169	[−0.564, +0.097]	.166	.331
User Strat. Assess. × Region	**−0.500**	**0.163**	**[−0.819, −0.181]**	**.002**	**.011***
Cultural Considerations × Region	−0.352	0.160	[−0.664, −0.039]	.028	.110
Amount Evaluation × Region	−0.348	0.185	[−0.710, +0.014]	.060	.179
Decision Rationale × Region	−0.087	0.189	[−0.457, +0.282]	.644	.644

Table B.8: CoT field × Alignment interactions on next-human-message length. Five-test Holm family. Same specification as Table B.6 with additive control replaced by interaction.

CoT field × Alignment	β	SE	95% CI	Raw p	Holm p
Situation Analysis × Alignment	−0.057	0.160	[−0.371, +0.257]	.721	1
User Strategy Assessment × Alignment	−0.075	0.155	[−0.378, +0.229]	.630	1
Cultural Considerations × Alignment	+0.071	0.155	[−0.234, +0.375]	.650	1
Amount Evaluation × Alignment	+0.174	0.170	[−0.159, +0.508]	.306	1
Decision Rationale × Alignment	+0.215	0.182	[−0.141, +0.571]	.237	1

Table B.9: Region effect and coaching × region interaction by individual CoT field. Each row is from a separate multilevel model with one CoT field's character count as outcome. $N = 1{,}161$ pairs, 80 participants.

CoT field	Region (Asia) β	SE	p	Coaching × Region β	p
Situation Analysis	**+31.825**	**2.807**	**< .001**	+6.810	.084
User Strategy Assessment	**+32.262**	**2.563**	**< .001**	**+9.687**	**.007**
Cultural Considerations	**+20.957**	**2.729**	**< .001**	+1.553	.685
Amount Evaluation	**+28.777**	**2.561**	**< .001**	−0.431	.905
Decision Rationale	**+23.172**	**2.562**	**< .001**	+3.090	.391

Bibliography

[1] M. F. Adilazuarda, S. Mukherjee, P. Lavania, S. S. Singh, A. F. Aji, J. O'Neill, A. Modi, and M. Choudhury. Towards measuring and modeling "culture" in LLMs: A survey. In Y. Al-Onaizan, M. Bansal, and Y.-N. Chen, editors, *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 15763–15784, Miami, Florida, USA, Nov. 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.emnlp-main.882. URL `https://aclanthology.org/2024.emnlp-main.882/`.

[2] Y. Aperstein, Y. Cohen, and A. Apartsin. Generative AI-based platform for deliberate teaching practice: A review and a suggested framework. *Education Sciences*, 15(4):405, 2025.

[3] J. Ashkinaze, J. Mendelsohn, Q. Li, C. Budak, and E. Gilbert. How AI ideas affect the creativity, diversity, and evolution of human ideas: Evidence from a large, dynamic experiment. In *Proceedings of the Collective Intelligence Conference*. Association for Computing Machinery, 2025. doi: 10.1145/3715928.3737481.

[4] O. Belova. Technologies for promoting animal welfare NGOs on the Telegram social network. *Vestnik Nauki*, 4(5 (86)):1105–1118, 2025.

[5] D. Bennett and E. D. Mekler. Beyond intrinsic motivation: The role of autonomous motivation in user experience. *ACM Transactions on Computer-Human Interaction*, 31(5):1–41, 2024. doi: 10.1145/3689044.

[6] A. U. Bolin and G. A. Neuman. Personality, process, and performance in interactive

brainstorming groups. *Journal of Business and Psychology*, 20(4):565–585, 2006. doi: 10.1007/s10869-005-9000-7.

[7] F. Bomba, M. Menéndez-Blanco, P. Grigis, M. Cremaschi, and A. De Angeli. The choreographer-performer continuum: A diffraction tool to illuminate authorship in more than human co-performances. *ACM Transactions on Computer-Human Interaction*, 31(6):1–23, 2024. doi: 10.1145/3689040.

[8] R. L. Boyd, A. Ashokkumar, S. Seraj, and J. W. Pennebaker. The development and psychometric properties of LIWC-22. *Austin, TX: University of Texas at Austin*, 10:1–47, 2022.

[9] S. E. Brennan. LEXICAL ENTRAINMENT IN SPONTANEOUS DIALOG. *Proceedings of ISSD*, 96:41–44, 1996.

[10] S. E. Brennan and H. H. Clark. Conceptual pacts and lexical choice in conversation. *Journal of experimental psychology: Learning, memory, and cognition*, 22(6):1482–1493, 1996.

[11] J. M. Brett. *Negotiating globally: How to negotiate deals, resolve disputes, and make decisions across cultural boundaries*. John Wiley & Sons, 2007.

[12] S. M. Breum, D. V. Egdal, V. G. Mortensen, A. G. Møller, and L. M. Aiello. The persuasive power of large language models. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 18, pages 152–163, 2024.

[13] E. Brynjolfsson, D. Li, and L. R. Raymond. Generative AI at work. Working Paper 31161, National Bureau of Economic Research, 2023. URL `https://www.nber.org/papers/w31161`.

[14] Y. Cao, M. Chen, and D. Hershcovich. Bridging cultural nuances in dialogue agents through cultural value surveys. In Y. Graham and M. Purver, editors, *Findings of the Association for Computational Linguistics: EACL 2024*, pages 929–945, St. Julian's, Malta, Mar. 2024. Association for Computational Linguistics. URL `https://aclanthology.org/2024.findings-eacl.63/`.

[15] M. Chandra, S. Naik, D. Ford, E. Okoli, M. De Choudhury, M. Ershadi, G. Ramos, J. Hernandez, A. Bhattacharjee, S. Warreth, and J. Suh. From Lived Experience to Insight: Unpacking the Psychological Risks of Using AI Conversational Agents. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, pages 975–1004, 2025. doi: 10.1145/3715275.3732063. URL https://doi.org/10.1145/3715275.3732063.

[16] R.-C. Chang and H.-C. Wang. Communication accommodation between large language models and users across cultures (student abstract). *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(28):29331–29333, 2025. doi: 10.1609/aaai.v39i28.35241. URL https://doi.org/10.1609/aaai.v39i28.35241.

[17] R.-C. Chang and H.-C. Wang. Convergence, reciprocity, and asymmetry: Communication accommodation between large language models and users across cultures. In *Proceedings of the 47th Annual Meeting of the Cognitive Science Society*, 2025.

[18] R. H. Chen. Effects of deliberate practice on blended learning sustainability: A community of inquiry perspective. *Sustainability*, 14(3):1785, 2022.

[19] R. B. Cialdini and N. J. Goldstein. The science and practice of persuasion. *Cornell Hotel and Restaurant Administration Quarterly*, 43(2):40–50, 2002.

[20] H. H. Clark and S. E. Brennan. Grounding in communication. In L. B. Resnick, J. M. Levine, and S. D. Teasley, editors, *Perspectives on socially shared cognition.*, pages 127–149. American Psychological Association, 1991. doi: 10.1037/10096-006. URL https://doi.org/10.1037/10096-006.

[21] H. H. Clark and D. Wilkes-Gibbs. Referring as a collaborative process. *Cognition*, 22(1): 1–39, 1986.

[22] J. R. Curhan, H. A. Elfenbein, and H. Xu. What do people value when they negotiate?

mapping the domain of subjective value in negotiation. *Journal of personality and social psychology*, 91(3):493–512, 2006.

[23] L. Debter. Retailers Are Testing An AI Bot That Haggles With Customers Over Price. *Forbes, Sep*, 2023.

[24] F. Dell'Acqua, E. McFowland, E. Mollick, H. Lifshitz, K. C. Kellogg, S. Rajendran, L. Krayer, F. Candelon, and K. R. Lakhani. Navigating the jagged technological frontier: Field experimental evidence of the effects of artificial intelligence on knowledge worker productivity and quality. *Organization Science*, 37(2):403–423, 2026. doi: 10.1287/orsc.2025.21838.

[25] M. Deutsch. A theory of co-operation and competition. *Human Relations*, 2(2):129–152, 1949. doi: 10.1177/001872674900200204.

[26] D. A. Dickey and W. A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366a):427–431, 1979.

[27] M. Diehl and W. Stroebe. Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology*, 53(3):497–509, 1987. doi: 10.1037/0022-3514.53.3.497.

[28] J. E. Driskell, E. Salas, and S. Hughes. Collective orientation and team performance: Development of an individual differences measure. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 52(2):316–328, 2010. doi: 10.1177/0018720809359522.

[29] O. J. Dunn. Multiple comparisons among means. *Journal of the American statistical association*, 56(293):52–64, 1961.

[30] N. Epley, A. Waytz, and J. T. Cacioppo. On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, 114(4):864–886, 2007. doi: 10.1037/0033-295X.114.4.864.

[31] K. A. Ericsson. Deliberate practice and acquisition of expert performance: a general overview. *Academic emergency medicine*, 15(11):988–994, 2008. doi: 10.1111/j.1553-2712.2008.00227.x.

[32] K. A. Ericsson, R. T. Krampe, and C. Tesch-Römer. The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100(3):363–406, 1993.

[33] V. Figueiredo. Designing smarter conversational agents for kids: Lessons from cognitive work and means-ends analyses. *ACM Transactions on Computer-Human Interaction*, 32(6):1–40, 2025. doi: 10.1145/3765284. URL `https://doi.org/10.1145/3765284`.

[34] A. P. Fiske, S. Kitayama, H. R. Markus, and R. E. Nisbett. The cultural matrix of social psychology. In D. T. Gilbert, S. T. Fiske, and G. Lindzey, editors, *The Handbook of Social Psychology*, volume 2, pages 915–981. McGraw-Hill, 4th edition, 1998.

[35] Y. Fu, H. Peng, T. Khot, and M. Lapata. Improving language model negotiation with self-play and in-context learning from AI feedback. *arXiv preprint arXiv:2305.10142*, 2023.

[36] A. D. Galinsky, G. Ku, and T. Mussweiler. To start low or to start high? The case of auctions versus negotiations. *Current Directions in Psychological Science*, 18(6):357–361, 2009.

[37] G. Gao, S. Y. Hwang, G. Culbertson, S. R. Fussell, and M. F. Jung. Beyond information content: The effects of culture on affective grounding in instant messaging conversations. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW):1–18, 2017.

[38] P. Ghosh and S. Rintel. YES AND: A generative AI multi-agent framework for enhancing diversity of thought in individual ideation for problem-solving through confidence-based agent turn-taking. In *Extended Abstracts of the 2025 CHI Conference on Human Factors in Computing Systems*, pages 1–13, New York, NY, USA, 2025. Association for Computing Machinery. doi: 10.1145/3706599.3720142.

[39] M. Giacomin and C. Jordan. Interdependent and independent self-construal. In *Encyclopedia of Personality and Individual Differences*, pages 1–7. Springer, 2017. doi: 10.1007/978-3-319-28099-8_1136-1.

[40] D. T. Gilbert, S. T. Fiske, E. J. Finkel, and W. B. Mendes, editors. *The Handbook of Social Psychology*. Situational Press, 6th edition, 2025. doi: 10.70400/NYKH3013. URL https://doi.org/10.70400/NYKH3013.

[41] H. Giles and T. Ogay. Communication accommodation theory. In *Explaining communication: Contemporary theories and exemplars*, pages 293–310. Lawrence Erlbaum, 2007.

[42] H. Giles, N. Coupland, and J. Coupland. Accommodation Theory: Communication, Context, and Consequence. In *Contexts of Accommodation: Developments in Applied Sociolinguistics*, pages 1–68. Cambridge University Press, 1991. doi: 10.1017/CBO9780511663673.001. URL https://doi.org/10.1017/CBO9780511663673.001.

[43] H. Giles, A. L. Edwards, and J. B. Walther. Communication accommodation theory: Past accomplishments, current trends, and future prospects. *Language Sciences*, 99:101571, 2023.

[44] C. W. J. Granger. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, 37(3):424–438, 1969. doi: 10.2307/1912791. URL https://doi.org/10.2307/1912791.

[45] C. Gutwin and S. Greenberg. A descriptive framework of workspace awareness for real-time groupware. *Computer Supported Cooperative Work*, 11(3–4):411–446, 2002. doi: 10.1023/A:1021271517844.

[46] E. T. Hall. *Beyond Culture*. Anchor Press, Garden City, NY, 1st edition, 1976. ISBN 9780385087476.

[47] E. T. Hall and M. R. Hall. *Understanding Cultural Differences*. Intercultural Press, Yarmouth, ME, 1990. ISBN 9780933662841.

[48] G. Hofstede. Cultures and organizations: Software of the mind. *McGraw Hills*, 1991.

[49] G. Hofstede, G. Hofstede, and M. Minkov. *Cultures and Organizations: Software of the Mind, Third Edition*. McGraw Hill LLC, 2010. ISBN 9780071770156. URL `https://books.google.com/books?id=o4OqTgV3V00C`.

[50] G. H. Hofstede and G. J. Hofstede. *Cultures and Organizations: Software of the Mind*. McGraw-Hill, New York, 2nd edition, 2005. ISBN 9780071439596.

[51] S. Houde, K. Brimijoin, M. Muller, S. I. Ross, D. A. Silva Moran, G. E. Gonzalez, S. Kunde, M. A. Foreman, and J. D. Weisz. Controlling AI agent participation in group conversations: A human-centered approach. In *Proceedings of the 30th International Conference on Intelligent User Interfaces*. Association for Computing Machinery, 2025. doi: 10.1145/3708359.3712089.

[52] R. Inglehart, M. Basanez, J. Diez-Medrano, L. Halman, and R. Luijkx. World values surveys and european values surveys, 1981-1984, 1990-1993, and 1995-1997. *Ann Arbor-Michigan, Institute for Social Research, ICPSR version*, 2000.

[53] M. E. Ireland and J. W. Pennebaker. Language style matching in writing: synchrony in essays, correspondence, and poetry. *Journal of personality and social psychology*, 99(3):549–571, 2010.

[54] E. Y. Ji. Large language models: A historical and sociocultural perspective. *Cognitive Science*, 48(3):e13430, 2024.

[55] J. G. Johnson, R. S. Huang, K. Nguyen, J. Y. Nam, and M. Nebeling. "i felt bad after we ignored her": Understanding how interface-driven social prominence shapes group discussions with GenAI. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, pages 1–22, New York, NY, USA, 2026. Association for Computing Machinery. doi: 10.1145/3772318.3791881.

[56] H. Ju and S. Aral. Collaborating with AI agents: Field experiments on teamwork, produc-

tivity, and performance. *arXiv preprint arXiv:2503.18238*, 2026. URL https://arxiv.org/abs/2503.18238.

[57] M. F. Jung. Affective grounding in human-robot interaction. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pages 263–273, 2017.

[58] S. Kim, J. Bak, and A. Oh. Do you feel what i feel? social aspects of emotions in Twitter conversations. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 6, pages 495–498, 2012.

[59] H. Kirk, A. Whitefield, P. Röttger, A. Bean, K. Margatina, J. Ciro, R. Mosquera, M. Bartolo, A. Williams, H. He, B. Vidgen, and S. Hale. The PRISM Alignment Dataset: What Participatory, Representative and Individualised Human Feedback Reveals About the Subjective and Multicultural Alignment of Large Language Models. In *Advances in Neural Information Processing Systems 37*, pages 105236–105344. Neural Information Processing Systems Foundation, Inc. (NeurIPS), 2024. doi: 10.52202/079017-3342. URL https://doi.org/10.52202/079017-3342.

[60] A. D. Kramer, J. E. Guillory, and J. T. Hancock. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24):8788–8790, 2014.

[61] R. E. Kraut. Applying social psychological theory to the problems of group work. In J. M. Carroll, editor, *HCI Models, Theories, and Frameworks*, pages 325–356. Morgan Kaufmann, 2003. doi: 10.1016/B978-155860808-5/50012-5.

[62] H. Kumar, J. Vincentius, E. Jordan, and A. Anderson. Human creativity in the age of LLMs: Randomized experiments on divergent and convergent thinking. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, 2025. doi: 10.1145/3706598.3714198.

[63] C. Li, D. Teney, L. Yang, Q. Wen, X. Xie, and J. Wang. CulturePark: Boosting Cross-cultural Understanding in Large Language Models. In *Advances in Neural Information Processing Systems 37*, pages 65183–65216. Neural Information Processing Systems Foundation, Inc. (NeurIPS), 2024. doi: 10.52202/079017-2082. URL `https://doi.org/10.52202/079017-2082`.

[64] N. Li and M. B. Rosson. At a different tempo: what goes wrong in online cross-cultural group chat? In *Proceedings of the 2012 ACM International Conference on Supporting Group Work*, pages 145–154, 2012.

[65] H. Lim, D. Choi, S. Nam, B. Kim, and H. Hong. Understanding human–multi-agent team formation for creative work. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, 2026. doi: 10.1145/3772318.3791166.

[66] H. Liu. We knead you: An animal welfare-themed educational board game. Master's thesis, Pratt Institute, 2025.

[67] S. M. C. Loureiro, F. Ali, and M. Ali. Symmetric and asymmetric modeling to understand drivers and consequences of hotel chatbot engagement. *International Journal of Human–Computer Interaction*, 40(3):782–794, 2024.

[68] H. R. Markus and S. Kitayama. Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, 98(2):224–253, 1991. doi: 10.1037/0033-295X.98.2.224.

[69] H. R. Markus and S. Kitayama. Cultures and selves: A cycle of mutual constitution. *Perspectives on Psychological Science*, 5(4):420–430, 2010. doi: 10.1177/1745691610375557.

[70] R. C. Mayer, J. H. Davis, and F. D. Schoorman. An integrative model of organizational trust. *Academy of management review*, 20(3):709–734, 1995.

[71] J. E. McGrath. *Groups: Interaction and Performance*. Prentice-Hall, Englewood Cliffs, NJ, 1984.

[72] G. Mills and G. Redeker. Self-repair increases referential coordination. *Cognitive Science*, 47 (8):e13329, 2023.

[73] C. Moon, G. A. Travaglino, and A. K. Uskul. Social value orientation and endorsement of horizontal and vertical individualism and collectivism: An exploratory study comparing individuals from North America and South Korea. *Frontiers in Psychology*, 9:2262, 2018. doi: 10.3389/fpsyg.2018.02262.

[74] A. Mortazavi, F. F. Taheran, and D. Amiri. Behind Chatbot Dialogues: The Power of Message Length Similarity and its Effect on Attitudes Toward Chatbot. SSRN, 2024. URL `https://doi.org/10.2139/ssrn.4712679`.

[75] K. Muir, A. Joinson, R. Cotterill, and N. Dewdney. Linguistic style accommodation shapes impression formation and rapport in computer-mediated communication. *Journal of Language and Social Psychology*, 36(5):525–548, 2017.

[76] B. Mullen, C. Johnson, and E. Salas. Productivity loss in brainstorming groups: A meta-analytic integration. *Basic and Applied Social Psychology*, 12(1):3–23, 1991. doi: 10.1207/s15324834basp1201_1.

[77] M. Muller, S. Houde, G. Gonzalez, K. Brimijoin, S. I. Ross, D. A. Silva Moran, and J. D. Weisz. Group brainstorming with an AI agent: Creating and selecting ideas. In *Proceedings of the 15th International Conference on Computational Creativity*, pages 156–165. Association for Computational Creativity, 2024.

[78] P. W. Mulvey and H. J. Klein. The impact of perceived loafing and collective efficacy on group goal processes and group performance. *Organizational Behavior and Human Decision Processes*, 74(1):62–87, 1998. doi: 10.1006/obhd.1998.2753.

[79] M. Namvarpour, B. Brofsky, J. Y. Medina, M. Akter, and A. Razi. Understanding Teen Overreliance on AI Companion Chatbots Through Self-Reported Reddit Narratives. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, 2026. doi: 10.1145/3772318.3790597.

[80] C. Nass and Y. Moon. Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1):81–103, Jan. 2000. doi: 10.1111/0022-4537.00153.

[81] K. G. Niederhoffer and J. W. Pennebaker. Linguistic style matching in social interaction. *Journal of Language and Social Psychology*, 21(4):337–360, 2002. doi: 10.1177/026192702237953.

[82] B. A. Nijstad and W. Stroebe. How the group affects the mind: A cognitive model of idea generation in groups. *Personality and Social Psychology Review*, 10(3):186–213, 2006. doi: 10.1207/s15327957pspr1003_1.

[83] M. Nomura, T. Ito, and S. Ding. Towards collaborative brain-storming among humans and AI agents: An implementation of the IBIS-based brainstorming support system with multiple AI agents. In *Proceedings of the ACM Collective Intelligence Conference*, pages 1–9, New York, NY, USA, 2024. Association for Computing Machinery. doi: 10.1145/3643562.3672609.

[84] S. Noy and W. Zhang. Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654):187–192, 2023. doi: 10.1126/science.adh2586.

[85] OpenAI, A. Hurst, A. Lerer, A. P. Goucher, A. Perelman, A. Ramesh, A. Clark, A. Ostrow, A. Welihinda, A. Hayes, A. Radford, A. Madry, A. Baker-Whitcomb, A. Beutel, A. Borzunov, A. Carney, A. Chow, A. Nichol, A. Paino, A. Renzin, A. T. Passos, A. Kirillov, A. Christakis, A. Conneau, A. Kamali, A. Jabri, A. Moyer, A. Tam, A. Crookes, A. Tootoonchian, A. Kumar, A. Vallone, A. Karpathy, A. Braunstein, A. Cann, A. Codispoti, A. Galu, A. Kondrich, A. Tulloch, A. Mishchenko, A. Baek, A. Jiang, A. Pelisse, A. Woodford, A. Gosalia, A. Dhar, A. Pantuliano, A. Nayak, A. Oliver, B. Zoph, B. Ghorbani, B. Leimberger,

B. Rossen, B. Sokolowsky, B. Wang, B. Zweig, B. Hoover, B. Samic, B. McGrew, B. Spero, B. Giertler, B. Cheng, B. Lightcap, B. Walkin, B. Quinn, B. Guarraci, B. Hsu, B. Kellogg, B. Eastman, C. Lugaresi, C. Wainwright, C. Bassin, C. Hudson, C. Chu, C. Nelson, C. Li, C. J. Shern, C. Conger, C. Barette, C. Voss, C. Ding, C. Lu, C. Zhang, C. Beaumont, C. Hallacy, C. Koch, C. Gibson, C. Kim, C. Choi, C. McLeavey, C. Hesse, C. Fischer, C. Winter, C. Czarnecki, C. Jarvis, C. Wei, C. Koumouzelis, D. Sherburn, D. Kappler, D. Levin, D. Levy, D. Carr, D. Farhi, D. Mely, D. Robinson, D. Sasaki, D. Jin, D. Valladares, D. Tsipras, D. Li, D. P. Nguyen, D. Findlay, E. Oiwoh, E. Wong, E. Asdar, E. Proehl, E. Yang, E. Antonow, E. Kramer, E. Peterson, E. Sigler, E. Wallace, E. Brevdo, E. Mays, F. Khorasani, F. P. Such, F. Raso, F. Zhang, F. von Lohmann, F. Sulit, G. Goh, G. Oden, G. Salmon, G. Starace, G. Brockman, H. Salman, H. Bao, H. Hu, H. Wong, H. Wang, H. Schmidt, H. Whitney, H. Jun, H. Kirchner, H. P. de Oliveira Pinto, H. Ren, H. Chang, H. W. Chung, I. Kivlichan, I. O. Connell, I. Osband, I. Silber, I. Sohl, I. Okuyucu, I. Lan, I. Kostrikov, I. Sutskever, I. Kanitscheider, I. Gulrajani, J. Coxon, J. Menick, J. Pachocki, J. Aung, J. Betker, J. Crooks, J. Lennon, J. Kiros, J. Leike, J. Park, J. Kwon, J. Phang, J. Teplitz, J. Wei, J. Wolfe, J. Chen, J. Harris, J. Varavva, J. G. Lee, J. Shieh, J. Lin, J. Yu, J. Weng, J. Tang, J. Yu, J. Jang, J. Q. Candela, J. Beutler, J. Landers, J. Parish, J. Heidecke, J. Schulman, J. Lachman, J. McKay, J. Uesato, J. Ward, J. W. Kim, J. Huizinga, J. Sitkin, J. Kraaijeveld, J. Gross, J. Kaplan, J. Snyder, J. Achiam, J. Jiao, J. Lee, J. Zhuang, J. Harriman, K. Fricke, K. Hayashi, K. Singhal, K. Shi, K. Karthik, K. Wood, K. Rimbach, K. Hsu, K. Nguyen, K. Gu-Lemberg, K. Button, K. Liu, K. Howe, K. Muthukumar, K. Luther, L. Ahmad, L. Kai, L. Itow, L. Workman, L. Pathak, L. Chen, L. Jing, L. Guy, L. Fedus, L. Zhou, L. Mamitsuka, L. Weng, L. McCallum, L. Held, L. Ouyang, L. Feuvrier, L. Zhang, L. Kondraciuk, L. Kaiser, L. Hewitt, L. Metz, L. Doshi, M. Aflak, M. Simens, M. Boyd, M. Thompson, M. Dukhan, M. Chen, M. Gray, M. Hudnall, M. Zhang, M. Aljubeh, M. Litwin, M. Zeng, M. Johnson, M. Shetty, M. Gupta, M. Shah, M. Yatbaz, M. J. Yang, M. Zhong, M. Glaese, M. Chen, M. Janner, M. Lampe, M. Petrov, M. Wu, M. Wang, M. Fradin, M. Pokrass, M. Castro, M. O. T. de Castro,

M. Pavlov, M. Brundage, M. Wang, M. Khan, M. Murati, M. Bavarian, M. Lin, M. Yesildal, N. Soto, N. Gimelshein, N. Cone, N. Staudacher, N. Summers, N. LaFontaine, N. Chowdhury, N. Ryder, N. Stathas, N. Turley, N. Tezak, N. Felix, N. Kudige, N. Keskar, N. Deutsch, N. Bundick, N. Puckett, O. Nachum, O. Okelola, O. Boiko, O. Murk, O. Jaffe, O. Watkins, O. Godement, O. Campbell-Moore, P. Chao, P. McMillan, P. Belov, P. Su, P. Bak, P. Bakkum, P. Deng, P. Dolan, P. Hoeschele, P. Welinder, P. Pronin, P. Tillet, P. Dhariwal, Q. Yuan, R. Dias, R. Lim, R. Arora, R. Troll, R. Lin, R. G. Lopes, R. Puri, R. Miyara, R. Leike, R. Gaubert, R. Zamani, R. Wang, R. Donnelly, R. Honsby, R. Smith, R. Sahai, R. Ramchandani, R. Huet, R. Carmichael, R. Zellers, R. Chen, R. Chen, R. Nigmatullin, R. Cheu, S. Jain, S. Altman, S. Schoenholz, S. Toizer, S. Miserendino, S. Agarwal, S. Culver, S. Ethersmith, S. Gray, S. Grove, S. Metzger, S. Hermani, S. Jain, S. Zhao, S. Wu, S. Jomoto, S. Wu, S. Xia, S. Phene, S. Papay, S. Narayanan, S. Coffey, S. Lee, S. Hall, S. Balaji, T. Broda, T. Stramer, T. Xu, T. Gogineni, T. Christianson, T. Sanders, T. Patwardhan, T. Cunninghman, T. Degry, T. Dimson, T. Raoux, T. Shadwell, T. Zheng, T. Underwood, T. Markov, T. Sherbakov, T. Rubin, T. Stasi, T. Kaftan, T. Heywood, T. Peterson, T. Walters, T. Eloundou, V. Qi, V. Moeller, V. Monaco, V. Kuo, V. Fomenko, W. Chang, W. Zheng, W. Zhou, W. Manassra, W. Sheu, W. Zaremba, Y. Patil, Y. Qian, Y. Kim, Y. Cheng, Y. Zhang, Y. He, Y. Zhang, Y. Jin, Y. Dai, and Y. Malkov. GPT-4o System Card. *arXiv preprint arXiv:2410.21276*, 2024. URL `https://arxiv.org/abs/2410.21276`.

[86] A. F. Osborn. *Applied Imagination: Principles and Procedures of Creative Problem-Solving*. Charles Scribner's Sons, New York, rev. edition, 1957.

[87] R. Parasuraman and D. H. Manzey. Complacency and bias in human use of automation: An attentional integration. *Human Factors*, 52(3):381–410, 2010. doi: 10.1177/0018720810376055.

[88] R. Parasuraman, T. B. Sheridan, and C. D. Wickens. A model for types and levels of human

interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, 30(3):286–297, 2000. doi: 10.1109/3468.844354.

[89] P. B. Paulus and V. R. Brown. Toward more creative and innovative group idea generation: A cognitive-social-motivational perspective of brainstorming. *Social and Personality Psychology Compass*, 1(1):248–265, 2007. doi: 10.1111/j.1751-9004.2007.00006.x.

[90] P. B. Paulus and M. T. Dzindolet. Social influence processes in group brainstorming. *Journal of Personality and Social Psychology*, 64(4):575–586, 1993. doi: 10.1037/0022-3514.64.4.575.

[91] S. Pawar, J. Park, J. Jin, A. Arora, J. Myung, S. Yadav, F. G. Haznitrama, I. Song, A. Oh, and I. Augenstein. Survey of Cultural Awareness in Language Models: Text and Beyond. *Computational Linguistics*, 51(3):907–1004, 2025. doi: 10.1162/coli.a.14. URL `https://doi.org/10.1162/coli.a.14`.

[92] J. W. Pennebaker, M. E. Francis, and R. J. Booth. *Linguistic Inquiry and Word Count: LIWC 2001*. Lawrence Erlbaum Associates, Mahwah, NJ, 2001.

[93] A. Plaat, M. Van Duijn, N. Van Stein, M. Preuss, P. Van der Putten, and K. J. Batenburg. Agentic Large Language Models, a Survey. *Journal of Artificial Intelligence Research*, 84, 2025. doi: 10.1613/jair.1.18675. URL `https://doi.org/10.1613/jair.1.18675`.

[94] A. Placani. Anthropomorphism in AI: hype and fallacy. *AI and Ethics*, 4(3):691–698, Feb. 2024. doi: 10.1007/s43681-024-00419-4.

[95] J. Prather, B. N. Reeves, P. Denny, B. A. Becker, J. Leinonen, A. Luxton-Reilly, G. Powell, J. Finnie-Ansley, and E. A. Santos. "It's weird that it knows what I want": Usability and interactions with Copilot for novice programmers. *ACM transactions on computer-human interaction*, 31(1):1–31, 2024. doi: 10.1145/3617367.

[96] P. Qin, C.-L. Yang, J. Li, J. Wen, and Y.-C. Lee. Timing matters: How using LLMs at different timings influences writers' perceptions and ideation outcomes in AI-assisted ideation.

In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pages 1–16, New York, NY, USA, 2025. Association for Computing Machinery. doi: 10.1145/3706598.3713146.

[97] K. Quan, D. Albassam, M. Wu, Z. Ding, and J. Chin. Towards AI as colleagues: Multi-agent system improves structured ideation processes. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, 2026. doi: 10.1145/3772318.3790375.

[98] H. Raiffa. *Negotiation analysis: The science and art of collaborative decision making*. Harvard University Press, 2007.

[99] S. Rasouli, M. Ghafurian, and K. Dautenhahn. Co-design and user evaluation of a robotic mental well-being coach to support university students' public speaking anxiety. *ACM Transactions on Computer-Human Interaction*, 32(3):1–70, 2025. doi: 10.1145/3718084.

[100] J. Rezwana and M. L. Maher. Designing creative AI partners with COFI: A framework for modeling interaction in human-AI co-creative systems. *ACM Transactions on Computer-Human Interaction*, 30(5):1–28, 2023. doi: 10.1145/3519026.

[101] L. Riefle, S. Alamo Alonso, M. Jungnitz, A. Rothmaier, H. Wiegand, R. Zacateco Herrera, and C. Benz. Adaptive conversational agents: Exploring the effect of individualized design on user experience. In *Proceedings of the Thirty-First European Conference on Information Systems (ECIS)*., 2023.

[102] E. Roth. DoNotPay is launching an AI chatbot that can negotiate your bills. The Verge, Dec. 2022.

[103] P. Röttger, H. Kirk, B. Vidgen, G. Attanasio, F. Bianchi, and D. Hovy. XSTest: A test suite for identifying exaggerated safety behaviours in large language models. In *Proceedings*

of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), pages 5377–5400, 2024.

[104] L. E. Scissors, A. J. Gill, K. Geraghty, and D. Gergle. In CMC we trust: The role of similarity. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 527–536, 2009.

[105] L. D. Setlock and S. R. Fussell. Culture or fluency? Unpacking interactions between culture and communication medium. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 1137–1140. ACM, 2011. doi: 10.1145/1978942.1979112.

[106] O. Shaer, A. Cooper, O. Mokryn, A. L. Kun, and H. Ben Shoshan. AI-augmented brain-writing: Investigating the use of LLMs in group ideation. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, 2024. doi: 10.1145/3613904.3642414.

[107] M. Shanahan, K. McDonell, and L. Reynolds. Role play with large language models. *Nature*, 623(7987):493–498, Nov. 2023. doi: 10.1038/s41586-023-06647-8.

[108] M. Sharma, M. Tong, T. Korbak, D. Duvenaud, A. Askell, S. R. Bowman, E. Durmus, Z. Hatfield-Dodds, S. R. Johnston, S. Kravec, T. Maxwell, S. McCandlish, K. Ndousse, O. Rausch, N. Schiefer, D. Yan, M. Zhang, and E. Perez. Towards Understanding Sycophancy in Language Models. In *The Twelfth International Conference on Learning Representations*, 2024. URL `https://openreview.net/forum?id=tvhaxkMKAn`.

[109] S. Shavitt, A. K. Lalwani, J. Zhang, and C. J. Torelli. The horizontal/vertical distinction in cross-cultural consumer research. *Journal of Consumer Psychology*, 16(4):325–342, 2006. doi: 10.1207/s15327663jcp1604_3. URL `https://doi.org/10.1207/s15327663jcp1604_3`.

[110] S. Shavitt, T. P. Johnson, and J. Zhang. Horizontal and vertical cultural differences in

the content of advertising appeals. *Journal of International Consumer Marketing*, 23(3–4): 297–310, 2011. doi: 10.1080/08961530.2011.578064.

[111] R. Shea, A. Kallala, X. L. Liu, M. W. Morris, and Z. Yu. ACE: A LLM-based negotiation coaching system. In Y. Al-Onaizan, M. Bansal, and Y.-N. Chen, editors, *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 12720–12749, Miami, Florida, USA, November 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.emnlp-main.709. URL `https://aclanthology.org/2024.emnlp-main.709/`.

[112] H. Shen, N. Clark, and T. Mitra. Mind the value-action gap: Do LLMs act in alignment with their values? In C. Christodoulopoulos, T. Chakraborty, C. Rose, and V. Peng, editors, *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 3097–3118, Suzhou, China, Nov. 2025. Association for Computational Linguistics. ISBN 979-8-89176-332-6. doi: 10.18653/v1/2025.emnlp-main.154. URL `https://aclanthology.org/2025.emnlp-main.154/`.

[113] A. Shepelenko, P. Shepelenko, K. Panidi, V. Kosonogov, and A. Shestakova. How the emotions evoked by homeless pets induce online charitable giving. *Journal of Philanthropy and Marketing*, 29(2):e1842, 2024.

[114] T. M. Singelis, H. C. Triandis, D. P. S. Bhawuk, and M. J. Gelfand. Horizontal and vertical dimensions of individualism and collectivism: A theoretical and measurement refinement. *Cross-Cultural Research*, 29(3):240–275, 1995. doi: 10.1177/106939719502900302.

[115] M. Skjuve, A. Følstad, K. I. Fostervold, and P. B. Brandtzaeg. My chatbot companion-a study of human-chatbot relationships. *International Journal of Human-Computer Studies*, 149:102601, 2021.

[116] T. Song, Y. Tan, Z. Zhu, Y. Feng, and Y.-C. Lee. Multi-agents are social groups: Investigating

social influence of multiple agents in human-agent interactions. *Proceedings of the ACM on Human-Computer Interaction*, 9(7), 2025. doi: 10.1145/3757633.

[117] L. Spillner and N. Wenig. Talk to me on my level–linguistic alignment for chatbots. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction*, pages 1–12, 2021.

[118] J. W. A. Strachan, D. Albergo, G. Borghini, O. Pansardi, E. Scaliti, S. Gupta, K. Saxena, A. Rufo, S. Panzeri, G. Manzi, M. S. A. Graziano, and C. Becchio. Testing theory of mind in large language models and humans. *Nature Human Behaviour*, 8:1285–1295, July 2024. doi: 10.1038/s41562-024-01882-z.

[119] L. Suchman. *Human-Machine Reconfigurations: Plans and Situated Actions*. Cambridge University Press, 2nd edition, 2007.

[120] R. I. Swaab, W. W. Maddux, and M. Sinaceur. Early words that work: When and how virtual linguistic mimicry facilitates negotiation outcomes. *Journal of Experimental Social Psychology*, 47(3):616–621, 2011.

[121] H. C. Triandis and M. J. Gelfand. Converging measurement of horizontal and vertical individualism and collectivism. *Journal of Personality and Social Psychology*, 74(1):118–128, 1998. doi: 10.1037/0022-3514.74.1.118.

[122] S. Trott, C. Jones, T. Chang, J. Michaelov, and B. Bergen. Do large language models know what humans know? *Cognitive Science*, 47(7):e13309, 2023.

[123] M. M. van Pinxteren, M. Pluymaekers, J. Lemmink, and A. Krispin. Effects of communication style on relational outcomes in interactions between customers and embodied conversational agents. *Psychology & Marketing*, 40(5):938–953, 2023.

[124] T. Wagner, N. Punyanunt-Carter, and E. McCarthy. Rules, reciprocity, and emojis: An

exploratory study on flirtatious texting with romantic partners. *Southern Communication Journal*, 87(5):461–475, 2022.

[125] Y. Wan and Y. M. Kalman. Diverse AI personas can mitigate the homogenization effect in human-AI collaborative ideation. *Computers in Human Behavior: Artificial Humans*, 8: 100289, 2026. doi: 10.1016/j.chbah.2026.100289.

[126] H.-C. Wang and S. Fussell. Groups in groups: Conversational similarity in online multicultural multiparty brainstorming. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, pages 351–360, 2010.

[127] H.-C. Wang, S. R. Fussell, and L. D. Setlock. Cultural difference and adaptation of communication styles in computer-mediated group brainstorming. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 669–678, 2009.

[128] H.-C. Wang, D. Cosley, and S. R. Fussell. Idea Expander: Supporting group brainstorming with conversationally triggered visual thinking stimuli. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, pages 103–112, 2010. doi: 10.1145/1718918.1718938.

[129] H.-C. Wang, S. R. Fussell, and D. Cosley. From diversity to creativity. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*, pages 265–274. Association for Computing Machinery, 2011. doi: 10.1145/1958824.1958864.

[130] X. Wang, W. Shi, R. Kim, Y. Oh, S. Yang, J. Zhang, and Z. Yu. Persuasion for good: Towards a personalized persuasive dialogue system for social good. In A. Korhonen, D. Traum, and L. Màrquez, editors, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5635–5649, Florence, Italy, July 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1566. URL `https://aclanthology.org/P19-1566/`.

[131] B. Yan, Z. Zhou, L. Zhang, L. Zhang, Z. Zhou, D. Miao, Z. Li, C. Li, and X. Zhang. Beyond

self-talk: A communication-centric survey of LLM-based multi-agent systems. *arXiv preprint arXiv:2502.14321*, 2025.

[132] S. Yao, J. Zhao, D. Yu, N. Du, I. Shafran, K. Narasimhan, and Y. Cao. ReAct: Synergizing reasoning and acting in language models. In *International Conference on Learning Representations (ICLR)*, 2023.

[133] M. Young-Ng, Q. Zhu, J. Liao, and H.-C. Wang. Balancing human agency and AI autonomy in human-AI idea selection. In *Extended Abstracts of the 2025 CHI Conference on Human Factors in Computing Systems*, pages 1–6, New York, NY, USA, 2025. Association for Computing Machinery. doi: 10.1145/3706599.3719880.

[134] H. D. Zajac, T. O. Andersen, E. Kwasa, R. Wanjohi, M. K. Onyinkwa, E. K. Mwaniki, S. N. Gitau, S. S. Yaseen, J. F. Carlsen, M. Fraccaro, M. B. Nielsen, and Y. Chen. Towards Clinically Useful AI: From Radiology Practices in Global South and North to Visions of AI Support. *ACM Transactions on Computer-Human Interaction*, 32(2):1–38, 2025. doi: 10.1145/3715115. URL `https://doi.org/10.1145/3715115`.

[135] R. Zhang, H. Li, H. Meng, J. Zhan, H. Gan, and Y.-C. Lee. The dark side of AI companionship: A taxonomy of harmful algorithmic behaviors in human-AI relationships. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pages 1–17, 2025.

[136] L. Zhou, J. Gao, D. Li, and H.-Y. Shum. The design and implementation of XiaoIce, an empathetic social chatbot. *Computational Linguistics*, 46(1):53–93, 2020.