CALIBRATING SOCIAL EXPERIENCE IN HUMAN-AI COLLABORATION: TOWARD MORE INNOVATIVE AND INCLUSIVE WORK FUTURES

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Angel Hwang

August 2023

CALIBRATING SOCIAL EXPERIENCE IN HUMAN-AI COLLABORATION: TOWARD MORE INNOVATIVE AND INCLUSIVE WORK FUTURES

Angel Hwang, Ph.D.

Cornell University 2023

Recent advances in artificial intelligence (AI) have expanded the range of possibilities for how human users can make use of computers and autonomous systems, shifting the narrative from collaboration *through* technology to collaboration *with* technology. As AI holds the potential to be our new teammates, should they carry on the social experiences that often serve as double-edged swords in human-human teamwork? The present dissertation examines this core inquiry through the following approaches. I begin by reviewing the literature on social cognition and psychology to discuss how social experience arises at the individual level and how it can impact small groups and teamwork (Chapter 2). I then discuss how different AI technologies – in particular, intelligent agents – can simulate social actors and elicit social experiences during interaction with human users (Chapter 3). Based on the existing literature, I examine how social experience in human-agent teamwork can be triggered through either a top-down or a bottom-up approach. I begin with studying the influence of social experience when an agent supports individual work in dyadic settings. Specifically, I empirically test how an autonomous teamwork agent's informed identity (Chapter 4) and perceived capability (Chapter 5) can influence users' experiences and behaviors and whether such interventions would facilitate or inhibit individuals' performance. In Chapter 6, I extend to investigate the social experience of human-agent collaboration in multi-player teams. Specifically, I

examine whether and how an autonomous agent could support marginalized members in such settings. In summary, while a teamwork agent could bring on benefits commonly introduced by having human teammates, further anthropomorphism and socialization do not lead to significantly more positive teamwork outcomes. Moreover, for individuals who have previously struggled in group settings, such as those who have experienced anxiousness or marginalization, emphasizing social experiences in human-agent teamwork can hinder their performance. Together, the present dissertation takes a principled approach to the research topic and offers theoretical and practical implications that could pave new avenues for future research.

BIOGRAPHICAL SKETCH

Angel Hsing-Chi Hwang is a human-computer interaction researcher and doctoral candidate at Cornell University. With training from multiple fields, she takes an interdisciplinary approach to explore the impact of emerging technology on the future of work, culture, and social interaction. Throughout her doctoral program, she develops a strong interest in human-AI collaboration and studies the implications of AI-empowered applications on innovation and inclusivity in teamwork through both quantitative and qualitative research. Most, if not all, of her work involves human subject studies, through which she finds the constant joy of discovery and the precious opportunities to view the world through the eyes of all works of life. While much of her research is grounded in social science theories, she also adopts a practical point of view to explore how these theoretical models can inform better design of AI tools for applied settings. Such perspectives are deepened by her various experiences collaborating with several leading industry partners, including Google, Microsoft, Sony, Adobe, and Accenture.

Before pursuing her Ph.D. at Cornell University, she studied business and quantitative social sciences at The University of Hong Kong (B.B.A.) and The University of Texas at Austin (M.A.) and later worked as a data scientist. Her experiences traveling and living abroad have nurtured her interdisciplinary, intercultural points of view in research. Growing up as a daughter of an artist and a computer engineer has also naturally drawn her interest in the field of human-computer interaction.

This dissertation is dedicated to my parents

and to anyone who finds both joy and pain in this complex social world.

ACKNOWLEDGEMENTS

The recent wave of ”AI hype” took off during my doctoral program, but completing a dissertation on this topic takes much more than simply following the tides. Studying the problems of artificial intelligence (AI) often reveals issues deeply rooted in humankind, as AI is, after all, a product of our footprints in the digital space. There are times when I have nothing but doubts in this area and my work, wondering if it is even the ”right” track to vision a future of collaboration *with* AI. It is those people around me who constantly remind me of the value of my work – and that I am more than my work.

I would like to thank my committee members, Andrea Stevenson Won, Susan Fussell, Malte Jung, and Mor Naaman, for their thoughtful feedback on my research. Together, they help me shape a more holistic and human-centered view of this area. Besides, I want to thank all the research assistants that I have worked with in the past and present: Min Lee, Oana Mirestean, Yifei Wang, Piero Salas-Allende, Esther Lee, and Wei Yang. Many parts of this dissertation would not have been possible without your help.

I want to express my greatest gratitude to my colleagues, collaborators, and friends that I met within and outside of Cornell. Your insightful perspectives and intellectual conversations have helped me grow and thrive way more than any formal education could have offered. These include past and present members of the Virtual Embodiment Lab (particularly Swati Pandita, for being my ”lab sister” at times when I felt the most struggled), Qian Yang, Meir Friedenberg, and other members of the Design AI Lab, members of the Citizen and Technology Lab, members of the Robots in Groups Lab, colleagues at Cornell Center for Social Sciences (particularly, Remy Stewart and Yolanda Xue, who initiated the Machine Learning for Social Sciences Workshops from scratch with me), Ruotong Wang at the University of Washington, Laura Herman at Adobe and Oxford University, and participants of the Harvard-MIT interdisciplinary neuroscience journal club (special thanks to Chi-Ning Chou for organizing efforts and for always inviting me even though I'm not affiliated with either of the institutions!). Last but certainly not least, I have been incredibly fortunate to call the Department of Communication my home department for the past few years. I especially thank Alisius Leong for her companionship. Thank you for navigating through most parts of the Ph.D. life with me and going through lots of trials and errors to find proper Asian food in Upstate New York.

I thank my industry collaborators and mentors for (re)introducing me to the future of AI through an entirely different lens. These include members of the SPAICE team at Sony Research, the Digital Experiences team at Accenture Labs, the FATE group at Microsoft Research, the ADRS team at Adobe, and the Cerebra group at Google Research. A special shout-out to Oliver Siy and Mural UXR for their close mentorship and for reminding me the joy of doing research lies in the process per se. I am proud to call myself a Muralist.

I thank my former advisor Jeeyun Oh, at the University of Texas, for introducing me to this great grand world of research. Thank you for always having faith in my potential and for constantly reminding me why I decided to pursue a career in research in the first place.

I am forever grateful for my dearest friends and family, Katy Cheng, Lianne Li, Arbiter Yang, and my parents. Together, you form a sound net of support and shade lights into some of the darkest times throughout this wild ride.

Finally, I thank all the participants who took part in my studies. Advancing human-centered AI and teamwork research requires insights into individuals' struggles in their day-to-day work and life. Thank you for sharing with me a

fraction of your vulnerability. If one day, my work could bring even the slightest comfort to any of your challenges, I call that the greatest accomplishment I could achieve as a researcher.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

AI AS TEAMMATE: AN INTRODUCTION

As recent advances in artificial intelligence (AI) take an unprecedented pace, escalating concerns emerge around job replacement and devaluation of human work [202, 300, 304, 245, 51, 361]. Yet, the question of whether AI can one day carry out humans' jobs autonomously is less of interest in this dissertation; instead, I advocate for tracing back to the more collaborative nature of AI, re-examining how AI has supported human work since its origin in order to envision how such collaboration may evolve as the technologies continue to advance. In fact, when Alan Turing, commonly known as the father of modern computers and AI, first brought AI to life during the Second World War, it was built to decrypt the German's enigmatic code, making substantial contributions not only to the British's victory but to significantly shortening the horrendous wartime [148]. After the war, Turing continued his dedication to developing and improving the code breaker – or, in his words, a ”machine that can think” [322].

Regardless of the authenticity of such historical anecdotes, AI was designed to support work in ways humans can hardly accomplish by themselves since its very genesis. Moreover, AI has long held a unique position that is incomparable to other types of tools or machines created in the history of humanity. Later on, we saw work dedicated to building cooperative computers [71] and even intelligent agents that can autonomously operate their own ”teams” (e.g., [182]). Not to mention, a whole field of computer-supported cooperative work (CSCW) later on emerges and dedicates to building technologies that augment individual and collaborative work.

The current state-of-the-art AI has progressed far beyond what one can imagine from its blueprint. Still, the original intention of supporting and co-operating with human work remains mainstream. Increasing work has posited that **human-AI collaboration** demonstrates more promises to build benign interaction between users and AI applications in the long run [334, 333, 195, 251, 289, 178, 243, 10]. In particular, with its constantly growing technical capabilities, AI has enabled various applications to transform from passively facilitating users' work (e.g., carrying out actions upon receiving user commands) to actively participating in user experience (e.g., prompting recommendations to users through ambient sensing and affective computing). Therefore, in this dissertation, I study AI as an agent that can collaborate *with* users (i.e., collaboration *with* AI) instead of as a medium through which users can connect and work with other human players (i.e., collaboration *through* AI).

1.1 Human-AI Collaboration in the Future of Work

Among the diverse perspectives toward the Future of Work [155, 342, 165, 338], two rare consensuses stand across the board. First, given the accelerating complexity in each domain, collaboration among professionals would only become essential instead of optional. Besides, scholars and practitioners have commonly acknowledged the crucial role of AI in our work futures. According to Forbes Insight Report in 2021 [266], more than 90% of leading businesses already have substantial investments in AI technologies, while there was a 270% growth in adopting AI in corporate practices in a single year. All in all, there is little doubt that AI will become *the* norms of work futures – not just in the tech industry – but in domains of all kinds. Then, what are the implications and potential of AI as we enter this new era? How could it enrich our work and fill in areas where we lack expertise and knowledge? And, how could it support collaboration across all walks of life while taking into account individual differences? Finally, could it help address long-standing challenges in human teamwork, such as innovation bottlenecks and inclusiveness?

With a look back at history, most, if not all, of the greatest work of all times, is done through collaboration. Through collaboration, a person would most likely encounter new perspectives, practices, and knowledge they would not have thought of when working solo. However, not all such divergences benefit the outcomes of work. Different opinions may cause conflicts, social judgments may suppress individual ideas, and pressures to comply may result in groupthink. While AI enters as an additional player in teams, it will likewise increase the complexity of work. The question is whether the benefits of human-AI collaboration outweigh the additional drawbacks and how AI agents and the human-AI collaboration paradigm could be designed to maximize its benefits.

1.2 Overview of Dissertation

In this dissertation, I examine (1) whether challenges commonly emerging in human teamwork would arise in human-agent teamwork, and (2) how such challenges could be addressed through the design of AI agents and human-AI interaction paradigms. More specifically, while social experience often serves as a double-edged sword in human-human collaboration, I investigate whether highlighting or de-emphasizing the social components in human-AI collaboration allows users to obtain the benefits of teamwork while minimizing the pain points of working with other humans.

In the next chapter (Chapter 2), I begin with reviewing existing literature to understand how social experiences arise and what contributes to one perceiving and treating others as social entities. Furthermore, I take a closer look at the teamwork literature and synthesize the impact of such social experiences on teams.

Following, in Chapter 3, I review existing work on the possible roles of computer agents in human teamwork, synthesizing insights into how social experiences could shape users' collaboration with autonomous systems and agents. Through the literature review, I also aim to understand how collaboration with AI is unique from collaboration with other machines. This could again inform whether and how social experiences should be tailored in human-AI collaboration.

Based on the literature review in the previous chapters, I conducted two series of empirical work to examine social experience in different forms of human-AI collaboration. In Chapter 4, I begin investigating the topic of the simplest form of collaboration, where there is one human and one AI in teamwork. I first examine whether and how the perceived identity of an agent – namely, the notion of working with a bot rather than a human – would affect users' performance and experience in teamwork.

In Chapter 5, I further explore the importance of agents' performance in human-AI collaboration. On one hand, I address whether it is essential for an intelligent agent to possess cutting-edge affordance and offer high-quality work in order to bring contributions to teamwork. On the other hand, I study what qualities – beyond technical capabilities – users desire for an AI teammate.

Next, in Chapter 6, I study human-AI collaboration beyond dyadic interactions. Specifically, I explore the potential role of AI in facilitating teamwork among multiple humans. I place a particular focus on AI's potential to support marginalized individuals in teams. Here, I conduct empirical work to investigate whether a shared identity with the AI agent can lead to greater support.

In the final chapter (Chapter 7), I conclude with key learnings from the series of works and suggest avenues for future research.

1.3 Intended Contributions

The intended contributions of the present dissertation are threefold. First, I review existing literature and relevant theoretical frameworks that explain how social experiences arise in teamwork. Through the series of empirical studies, I investigate and summarize whether prior theories in human teamwork are applicable to studying individuals' behaviors and experiences in teamwork with AI agents. In other words, the present dissertation responds to one of the most discussed inquiries among human-computer interaction (HCI) scholars in recent years – whether and how is human-AI interaction unique from other HCI paradigms and whether we need new theories to understand this topic area? [355]. As reviewed in Chapter 3, existing literature offers various theories and perspectives to explain why machines hold the potential to serve as humans' teammates. The present work re-examines whether and when these prior theories and models can be extended to comprehend users' collaboration with more intelligent and capable AI agents while identifying avenues where new research and theories are necessary.

Second, all empirical studies in the present dissertation are designed to provide actionable practices for the design of teamwork agents. In particular, Chapter 4 and Chapter 5 highlight the advantages and drawbacks of endorsing the human-ness vs. bot-ness in agents. Furthermore, findings from the present dissertation also suggest the need to preserve an irreplaceable role for humans in collaboration settings with AI agents. As technical capabilities of AI continue to emerge, intelligent agents may be able to automate various parts of workflows. Nonetheless, this by no means implies the introduction of AI teammates should serve to replace human work. Instead, creating team experiences that allow humans to lead and make substantial contributions to teamwork remains crucial for building sustainable forms of interaction and collaboration with AI entities.

Finally, the present dissertation calls for the need to address individual differences to avoid replicating common challenges in human-human teamwork in human-AI collaboration settings. To begin with, the current work not only provides insights into how individuals' prior positive or negative teamwork experiences may affect their behaviors in collaborating with AI, such that certain groups who struggled in human-only teams may in fact benefit more from working with machine teammates. This once again suggests simulating human-human teamwork in human-AI collaboration may not be necessary, and emphasizing the bot-ness of AI teammates can also tailor to the interests of certain users. Later on in Chapter 6, I further explore the effect of having a teamwork agent that specifically serves a supporting role for those who were marginalized in teams. Considering the wide range of individual needs in teamwork, the present dissertation explores only the tip of the iceberg, but it has already shown the importance of continuously exploring different approaches to design agents, systems, and interaction paradigms to support diversity and inclusivity in human-AI collaboration. Through the present dissertation, I not only advocate the need for more attention to these topics but also reveal numerous avenues for future research.

CHAPTER 2

THE RISE OF SOCIAL EXPERIENCE: MULTI-DISCIPLINARY PERSPECTIVES

Social experience, namely, the experience of interacting socially with others, is a part of our everyday life. Social experience is also vital to teamwork and collaboration. Beyond achieving practical goals, a substantial portion of teamwork is also established on the social experience of its team members interacting with one another. However, how and where does this unique experience arise? When working with an autonomous agent, when does one perceive it as a social entity? Conversely, when does one view human-agent collaboration as identical to the experience of adopting a tool? In this chapter, I revisit and review classic literature from social sciences to examine the essence of social experience from multiple disciplines as well as their implications on the complexity of how humans collaborate and interact socially in teams.

2.1 Social Cognition in Various Disciplines

Across different domains, the concept of social cognition is commonly discussed and considered the core of social experience [181, 124, 125]. Social cognition refers to ”the ability to make sense of the world through processing social signals” from others (p. 107) [181]. In cognitive science, human development, and behavioral science, scholars examine a range of capabilities that enable humans to process and interpret information from their social world. Given these mechanisms, social psychologists extensively study how humans form the sense of self versus others and how the notion of social groups arises. Based on these notions, individuals determine how they behave and interact with others. Because social cognition is crucial to whether one can smoothly navigate their social world, it also impacts the quality of one's social experience and one's need to bond and form meaningful relationships with others.

In the following sections, I elaborate on the primary mechanism of social cognitive processing (Section 2.1.1) and how it is applied to shape the sense of "self vs. others" in our social experiences (Section 2.1.2). I summarize existing literature on how social cognition influences one's social experience through a top-down and a bottom-up pathway (Section 2.1.3).

2.1.1 Social cognitive processing

Existing literature in cognitive science and human development studies the root of social cognition through two major components. To begin with, a person needs to be able to perceive and process information from others and the environment in order to make sense of the social world [125, 181, 3]. This first component is referred to as the social perceptual processes, formulating social experiences through a bottom-up approach (i.e., understanding others and the environment through external, observable signals, such as their facial expressions and movements). Second, upon perceiving these signals, one needs to make sense of them through reasoning and interpretation. Making inferences that fit well in a social context, such as mentalizing and understanding another person's intention, is hence another critical component of social cognition that help shape one's social experiences.

To understand how humans perceive information from other social entities, face processing [114] and motion detection [254] are two of the most studied capabilities that support social cognitive processing. First, human faces carry a large amount of critical information that helps one navigate social interaction. At the minimum, human faces represent the identities of one's interactants [47]. Furthermore, facial expressions often reveal a person's emotion, which more or less conveys one's responses (e.g., positive or negative perceptions) to a social scenario [25]. Meanwhile, micro-, non-verbal behaviors on one's face, such as the direction of eye gaze, reveal one's attention and interests during interaction [105]. Scholars have found the ability of face processing emerges and develops rapidly since infancy [114, 228]. Meanwhile, when making social decisions, people are commonly influenced by nuanced information extracted from human faces during communication, including but not limited to attractiveness [267, 171, 65], social status, or power dynamics [164, 175, 176].

Similarly, humans attend to others' movements and collect much information from them during social interaction [254]. In fact, motions and contingent behaviors are important signals that help humans detect agency in a subject and determine whether they would treat this target as a social entity. Since early infancy, humans can decide how to respond to an object depending on whether it demonstrates self-propelled actions or not [257, 215]. Even with an inanimate object, humans could determine whether to treat it as an agent depending on whether it could interact contingently with a person [144, 168]. Moreover, individuals tend to automatically interpret an object as an agent should it demonstrate communicative abilities and goal-directed behaviors [275, 124].

Upon collecting information from the external world, one then needs to perform reasoning and inferencing to comprehend the meanings behind the information. Such internal cognitive processes oftentimes involve **mentalizing**

[125] and **perspective-taking** [117] in social decision-making [30, 110, 329, 328]. Mentalization refers to the capability of making sense of other people's mental states, including "their beliefs, thoughts, desires, intentions, and feelings" [181]. At the very minimum, being able to mentalize indicates a person needs to be aware that others may hold different thoughts and feelings from themselves, and it is this realization that motivates one to understand others' internal states [16, 24]. Besides, understanding the intention behind a person's behavior is also a critical function of mentalization, as intentions often inform what may seem appropriate or inappropriate to respond to the person. Furthermore, upon comprehending another person's thoughts, feelings, and intentions, whether one can incorporate such information into their decision-making process is yet another component of social cognition. This capability is referred to as perspective-taking [117]. The practice of perspective-taking allows one to interact more smoothly with others [117]. Perspective-taking is not only necessary during our one-on-one interaction with another individual but also during interactions with multiple players in group settings. Indeed, existing research has found whether each individual in a group could perform perspective-taking well is essential to enhancing fairness, trust, and cohesion at the group level [30, 110, 329, 328].

All in all, scholars have studied a wide range of capabilities that humans leverage to navigate their social world. For instance, a person cannot form meaningful connections with others if they cannot understand and empathize with the feelings and thoughts of others. And it requires the ability to mentalize in order to achieve such goals. It is through practicing these capabilities of social cognition in real life that one forms and completes their social experiences.

2.1.2 Social identity and categorization

In social psychology, Tajfel's seminal **Social Identity Theory** [309] leaves remarkable traits in countless literature studying human social experiences. According to the theory, the formation of a person's social identity (namely, the sense of who they are) sets the cornerstone for their social experiences. Specifically, one tends to form their own social identity through the social group they see themself as a part of. Tajfel referred to the act of assigning others to specific group memberships as **social categorization** [310], and the process of shaping one's own identity as well as others' identities through their corresponding social groups as **social identification** [309].

These two intersecting mechanisms are closely relevant to the human tendency to group things together in order to facilitate them to process information more efficiently [149]. Similarly, humans hold the propensity to put others into groups to help them readily process information about individuals with structures [310], which also allows them to appraise effective ways to interact with others. Interestingly, humans not only put others but also themselves into specific groups [309]. Through such group assignments, humans set up expectations for themselves and others. In other words, humans would construct the sense of who they are through (1) assimilating themselves to groups to which they identify and, at the same time, (2) differentiating themselves from groups that they do not belong. Through such processes, individuals divide their world into "us" versus "them." This division serves as the foundation of one's social experiences.

To perform social categorization, one needs a set of attributes to evaluate each individual. Formally referred to as *prototype*, this set of attributes should represent the unique features of each group and differentiate it from other groups. According to the **Self-Categorization Theory** [324], Turner et al. posited that a person would comply with the prototype of their self-identified social group, and two phenomena often emerge during this process. First, the distinction between **ingroup** versus **outgroup** becomes more vivid. That is, as we identify ourselves with a specific group, it becomes clearer who belongs to the same group as us (i.e., ingroup) and who does not (i.e., outgroup). Following, individuals have the tendency to compare between groups, demonstrating the act of **social comparison**. Moreover, one is often motivated to examine and find evidence showing one's own group is superior to others. In Chapter 5, I review the concept of social comparison with greater detail and examine how individuals' tendency to compare with other social groups may moderate the outcomes and experience of working with a more or less capable intelligent agent.

The line of work in social psychology suggests social experience arises as individuals establish the sense of "self vs. others" and "us vs. them" by placing themselves and others into various social groups. Through such a process of assigning group memberships, individuals gradually form a more concrete sense of who they are. As the distinctions between groups become more apparent, individuals are also inclined to compare their own groups to others. Together, the sense of belonging to one's own group, the pride of group superiority, and the negative perception of being defeated all add colors to our complex social experiences.

2.1.3 Two pathways to forming social experiences

To summarize the section (Section 2.1), I refer to recent reviews on social cognition in social and cognitive science literature [79, 276, 212]. Jointly, existing literature suggests three key components shape individuals' social experiences: the what, how, and why components (as shown in Figure 2.1).

	Examples of theoretical constructs	Concepts operated in the present dissertation
Why? (goal)	Self-Identification	Informed Identity (Chapter 4)
How? (rules)	Mentalization & Perspective-Taking Social Facilitation & Inhibition (Chapter 4) Social Categorization & Comparison (Chapter 5 & 6)	
What? (physical)	emotional expressions, verbal & non-verbal cues, appearance, performance	Perceived Capability (Chapter 5), Group Composition (Chapter 6)

Figure 2.1: Two pathways to forming social experiences and relevant concepts applied in the present dissertation

First, the *what* component refers to the physical, external cues that one observes in their social environment and during their interactions with others. Second, the *how* component refers to the rules one applies to process these external cues. Third, the *why* feature relates to the goals one holds during social interactions or the intentions one interprets from others' behaviors. Together, a person may form social experiences based on these three components either through a bottom-up or a top-down fashion. For instance, through a bottom-up route, one may observe linguistic or physical cues from another person (the *what* component), based on which they could categorize the person as a specific group (the *how* component), and finally, based on such outcomes of social categorization, one may either identify with the person or treat them as an outgroup (the *why* component). Conversely, a top-down pathway is driven by pre-existing notions

– for example, one may hold existing stereotypes for a certain group (the *why* component), and at the encounter of a new person, one may evaluate whether the person fits such a label or not (the *how* component), and accordingly, one determines how they would behave and interact with the social other (the *what* component).

While these three components have been studied extensively in social cognition literature, it remains an under-explored question of whether they apply to form social experiences in human-AI interaction. In the present dissertation, I used theoretical constructs from each of the three components and empirically tested them in Chapter 4 and Chapter 6. Chapter 4 examined how the informed identity of an intelligent agent could affect humans' behaviors and social experiences through the top-down path. In Chapter 5, I altered an agent's behaviors to investigate how changes in such external, observable cues may influence humans' responses from a bottom-up fashion. Similarly, in Chapter 6, I also examined the bottom-up route by introducing changes at the group level (i.e., changing the group composition to alter the external cue) and assess how such an intervention could alter individuals' behaviors and social experiences in human-agent teamwork. It is worth noting that, though various theories can be applied to explain and understand internal processing happening at the how component, it is relatively challenging to directly manipulate the rules that one adopts to process external cues in their mind (in the bottom-up route) or to determine appropriate behaviors that match their upper-level goals (in the top-down pathway). This is, however, a novel avenue that I recommend for future research.

2.2 Social Experience in Teamwork

Given these various sources where social experiences arise, I then further discuss whether and when social experiences may be beneficial or detrimental to a teamwork setting. In this section, I review existing literature to discuss the utility of social experience in teamwork from the following dimensions: the effect of social presence (Section 2.2.1), ingroups versus outgroups perceptions (Section 2.2.2), communication in different types of team composition (Section 2.2.3). Notably, because teamwork *per se* is an inherently social process, various other social components of social interaction can lead to the success or failure of a team. Here, I select these dimensions to focus on as they are also closely related to many of the theories and literature on social experiences and social cognition, as reviewed in the previous section (Section 2.1).

Previously in Section 2.1.1, I discuss how humans sense and comprehend the behaviors and intentions of others through social cognitive processing. This distinct notion of self versus others sets a foundation for the effect of social presence. Accordingly, in Section 2.2.1, I focus on reviewing the effect of having the mere presence of others (namely, even if there is minimal interaction taking place between a person and their teammates), which is often studied through the theoretical framework of social facilitation and social inhibition [5, 359]. Next, as I discuss humans' tendency to categorize others into social groups in Section 2.1.2, perceiving teammates as ingroup or outgroup has critical implications on how one behaves and interacts with their teammates, influencing their overall team performance and cohesion. Thus, I discuss the effect of ingroup and outgroup perceptions further in Section 2.2.2. Finally, given how one identifies oneself – and the intentions to live up to one's identity – individuals are driven and assigned to perform specific roles in teams. Ideally, a more diverse team would include a wide variety of individuals who thus serve all types of needs the team may encounter. However, because individuals treat ingroups and outgroups differently, a more diverse team may also heighten the tension between ingroups and outgroups, leaving whether to enhance diversity in group composition an under-debate question.

2.2.1 Social facilitation vs. social inhibition

As elaborated in Section 2.1.1, individuals' social experiences emerge with the notion of others. The notion of "Self vs. Others" indicates other people can hold different thoughts, perceptions, and behaviors than one's own. Extending this concept, **Social Facilitation Theory** [321] – one of the oldest theories in social psychology – suggests the presence of others can influence a person's performance and behaviors. Specifically, the effect of social facilitation can be elicited through three driving forces. These include the effect of *mere presence* [359, 5, 320], *cognitive-behavioral cues* [290, 101], and *social cues* [59, 90, 145]. First, as in Triplett's original study [321], the mere presence of others can motivate individuals to perform better in given tasks as long as they are aware of the presence of others in an environment. Secondly, the presence of others can also initiate cognitive-behavioral cues, such as signals of competition, which lead individuals to improve their performance in order to stand out from others [290, 101]. Finally, one may also be driven by social cues, such as appreciation, evaluative apprehension, and other social rewards that encourage improved performance [39, 315].

However, subsequent work has also revealed the negative effect of social presence, which is more formally referred to as **Social Inhibition** [49]. In particular, such negative consequences of social presence were more likely to occur when one performs less skilled, less familiar tasks under high-stress, high-stakes scenarios [160, 185, 49]. To begin with, the presence of others may also become a source of distraction, preventing one from focusing on the tasks at hand. Furthermore, there is no guarantee that competition will remain benign and the driving forces of social cues will constantly turn into positive outcomes. For instance, vicious competition could cause both parties to perform worse, as hyper-focusing on winning a competition can cause ignorance of details in the work process. This could potentially result in worse outcomes of competition. Likewise, social feedback from others may also contain negative messages. While humans are inherently sensitive to criticisms, it is particularly difficult to avoid being affected by negative comments and feedback, which may also preoccupy and distract us from focusing on carrying out the tasks per se.

Synthesizing both the positive and negative outcomes of social presence, Sanders and Baron posited the Distraction-Conflict framework [273, 22]. The theoretical model suggests that individuals' cognitive capacity is divided by the presence of others. That is, a person's attention is split between performing their task and monitoring the behaviors of their interactants. When attention is evenly divided, this conflict serves as a cognitive stimulus, allowing individuals to be more alert and more engaged while participating in the group activity. However, when either of the two sources overrides the other (either when a person is overly concerned about other team members' responses or when the person ignores others' contributions and focuses only on completing the task), such conflict can distract a team from producing favorable outcomes [273, 22, 36].

Together, the theories of social facilitation, social inhibition, and distraction conflict suggest that being aware of the presence of others could motivate a person to perform better in tasks, primarily through eliciting the sense of competition and offering social rewards [321, 290, 101, 39, 315]. However, while the positive effects of presence can lead to social facilitation, these driving forces can also inhibit a person's performance when they pay too much attention to social feedback and competition from others. As a result, the person may be distracted from focusing on the task at hand [273, 22, 36]. Similarly, in teamwork settings, the presence of teammates can encourage each member to be more dedicated to the collaboration [122], but if a person becomes too occupied either by competing with or receiving approval from their teammates, knowing the presence of others can hinder their performance [4]. In such situations, one can perform worse in teamwork than carrying out the same work alone. Furthermore, if multiple teammates encounter similar challenges, the notion of others can drastically impede group-level performance.

2.2.2 Ingroup vs. outgroup perceptions

As reviewed in Section 2.1.2, humans have the tendency to categorize individuals into social groups and to form their self-identity through assimilating with their ingroup and differentiating themselves from outgroup [309, 310]. This ingroup versus outgroup awareness also profoundly impacts teammates' behaviors, team interaction, and team experience. As suggested by their literal meanings, ingroup refers to the social groups one identifies with, while outgroup indicates those not seen as part of one's self-identified group. Moreover, individuals tend to favor their ingroup over other outgroup individuals.

Social science scholars have long debated whether positive perceptions toward ingroup members and negative perceptions against outgroup individuals operate independently or not. Building on Sumner's classic work [299], some believe positive sentiments toward the ingroup, such as "pride, loyalty, and perceived superiority" (p. 430) [45], are sources of prejudice that eventually lead to contempt and hostility against outgroup members. On the other hand, those who built on Allport's seminal work [6] posit that favorability toward ingroup does not necessarily cause outgroup hatred. Instead, positive sentiments toward one's own social groups are "psychologically primary" (p. 42) [6], guiding one to more actively show friendliness, warmth, and familiarity when interacting with ingroup members.

Aside from the theoretical discussions, empirical studies have found much evidence demonstrating the impact of ingroup and outgroup perceptions on teamwork. Various literature has found ingroup and outgroup perceptions can affect communication effectiveness during collaboration, such that individuals tend to endorse more trust and turn more frequently toward ingroup members. In contrast, outgroup perceptions can set communication barriers, such as taking more time to warm up and build rapport during conversations. Such distinct effects of ingroup versus outgroup perceptions are particularly salient when a team works in a high-pressure environment, such as in hospitals and other venues that provide urgent care [327]. As communication is vital to the success of all forms of social interaction, ingroup and outgroup perceptions influence whether or not team members can effectively achieve conversational goals – including both functional purposes (e.g., generating ideas, incorporating diverse opinions, and forming consensus) and social goals (e.g., rapport building) – and, thus, indirectly affect the outcomes of team collaboration.

Furthermore, different behavioral patterns when interacting with ingroup and outgroup teammates also have an effect on individuals' team experience and motivation, which are also important determinants of one's performance in teamwork [163]. In particular, individuals tend to feel more motivated to contribute when more ingroup members are on a team. Meanwhile, people become more insensitive to equity when most teammates are perceived as outgroup; in such cases, one can easily turn less motivated to contribute because the sentiment of unfairness (e.g., taking up too much work or being free-rode by others) more often arises. Due to these reasons, when a team consists of an imbalance team composition (i.e., people of a specific identity dominate the team), the minority members – namely, those who are treated by outgroup by the majority on the team – are more likely to report a negative team experience and demonstrate worse performance. In the following section, I further discuss the effect of group composition on teamwork.

2.2.3 Homogeneous vs. heterogeneous team composition

When people work in group settings, the aforementioned effect of social presence and the distinct perceptions toward ingroup versus outgroup typically become more salient [45]. Oakes [323] posited that the tension between ingroup versus outgroup may further arise in small group settings because one now has more specific targets to compare. Thus, the perceived distinctions with those outgroup members become more accessible in one's mind. Furthermore, each group would gradually establish its norms as interactions continue. These norms may entail assigning roles or attaching labels to specific individuals in the group, which may all contribute to perceived differences toward certain team members. Still, as each individual is unique, even if these group norms are never created, it is nearly impossible to form a team where everyone is perceived as identical, and none is viewed as an outgroup.

The question is, then, how diverse or similar teammates should one look for when forming a team? Scholars across various fields have long studied the benefits and disadvantages of homogeneous versus heterogeneous teams [327, 17, 163, 45, 299]. Overall, empirical studies, reviews, and meta-analyses have constantly shown that teams consisting of more diverse teammates can more consistently lead to team innovation due to the wider range of perspectives those teammates can bring to the table. Besides, including teammates with different personalities, soft skills, and behavioral modes can also bring social benefits to the team. For example, some can be more assertive when a decision needs to be made, while other sociable teammates can help a team by building good relationships with external partners. However, ensuring team cohesion in diverse teams is also significantly more challenging. To begin with, as perceived differences with outgroup members are more salient, conflicts are also more likely to emerge in teams with greater diversity. This is because individuals generally hold more inferior perceptions toward outgroup than ingroup members, and they are typically less tolerant when outgroup members make mistakes. As a result, though teams formed by diverse members are more likely to spark novel ideas and make the most out of each teammate's strength, such group compositions also face greater risks of collapse.

Conversely, a team that lacks diversity and is dominated by a majority group may be more likely to reach agreements. However, this type of group composition often comes with a number of underlying concerns. First of all, minority individuals often face greater pressure to conform to such teams. Therefore, opinions from the majority group remain dominant, and alternative voices may often be overlooked. This scenario is not only detrimental to team innovation – as various perspectives cannot be incorporated during teamwork – but can also discourage and cause particularly inferior team experiences for those marginalized team members. Even if a single voice does not dominate the team, this type of group composition can often suffer from the Multiple Source Effect [224]. This phenomenon occurs when people believe an idea is superior simply because more people support it, but the actual quality of the idea may be overlooked. When a majority group exists in a team, an idea can easily get more ″votes″ simply because those in the majority group are more likely to conform with one another. In other words, though teams with reduced diversity seem more cohesive on the surface, this may result from the minority individuals remaining silent and the tendency to groupthink among the majority people.

CHAPTER 3

SOCIAL EXPERIENCE IN HUMAN-AI COLLABORATION

Based on the literature review from the previous section (Section 2.2), I learned that social experiences could often serve as a double-edged sword in teamwork. The social presence of others can either make one more focused or distract them from their task performance. Likewise, social interactions can either emphasize the tension between ingroup and outgroup members or encourage them to incorporate more diverse opinions during teamwork. Then, to what extent should human-AI collaboration simulate social experiences like those that naturally emerge in human-human teamwork? And by designing intelligent teamwork agents and interventions in human-AI interaction paradigms, is it possible to adopt the benefits of socially working with others while preventing the drawbacks that social experiences bring to teamwork? Before I can get to these important inquiries, an outstanding question is whether human users can at all form social experiences with an AI teammate during collaboration. Thus, I first address this fundamental question in the present chapter by reviewing various views on the possible relationships between human users and their machine teammates. I then review prior literature on how one can introduce *humanness* to make an intelligent agent functional and a social teammate.

3.1 Can Machines be Teammates?

The question of whether machines can be teammates has been much debated in modern HCI and CSCW literature. As recent AI advances introduce new capabilities (e.g., generative AI) for what an intelligent agent can offer in a collaborative setting, this long-standing question has once again come to the forefront of scholarly discussions. However, while humans have been applying all sorts of machines throughout the history of humanity, the idea of machines as teammates has only emerged very recently. In the following, I review three key perspectives that shape this seminal concept.

3.1.1 Computers as autonomous agents

To begin with, a large body of literature posits whether machines can be treated as teammates depending on their technical capabilities. In other words, if a machine can autonomously carry out functional values that a human teammate typically offers, then it shows the potential to be one [1, 186, 132, 219, 220, 219, 89]. In the past, one may question whether a computer agent is qualified as a teammate because it often relies heavily on human guidance [83, 247]. However, recent advances in AI have enabled agents to self-direct and self-govern through functions such as reinforcement learning and situation awareness [220, 220, 219, 331, 106].

These advanced capabilities indicate machines now possess greater capabilities to co-work and co-create with users. For instance, abundant recent work showed how users could collaborate with generative AI systems to seek inspiration, perform design work, and even receive evaluative feedback [242, 211, 135, 259]. As agents' autonomy increases, they not only can guide their own behaviors, but more sophisticated agents can also guide other agents. For example, swarm robots can direct themselves and facilitate one another to work toward the same goal [40, 182]. All in all, such intelligent agents that can achieve nearly full autonomy are said to show great potential for being humans' teammates.

3.1.2 Computers as social actors

Another point of view builds on Nass and colleagues' classic "Computers are Social Actors (CASA)" paradigm [237] and the Media Equation [265]. The former theory suggests users treat computers as social entities and demonstrate social behaviors in human-computer interaction as in interpersonal scenarios [237]. The Media Equation further postulates users tend to assign human-like characters to non-human objects and mediums, including computers, televisions, radios, etc. Based on these theoretical frameworks, scholars in the late 90s have already proposed that computers could and should be viewed as teammates in collaboration settings since users would treat them as social actors regardless of the use cases [58, 50, 303, 235].

Later on, research has found adding more humanoid features (e.g., human-like appearance, body figures, facial expressions) to the design of computer agents would further enhance users' tendency to treat them like human entities [88, 291, 100]. As a result, such practices of anthropomorphism allow researchers and practitioners to build highly social robots for scenarios where providing social support is indeed crucial [43, 42]. Along the same vein, recent advances in natural language processing (NLP) [263, 196, 226, 12] and text-to-speech generation [297, 260, 55] have enabled intelligent agents (e.g., chatbot and voice assistants) to interact with users through natural forms of interpersonal communication (e.g., text messages, verbal communication). Thus, recent studies adopting these state-of-the-art agents have once again cultivated more supporting evidence for the CASA paradigm and the Media Equation, suggesting users inevitably interact socially with these agents since such human-agent interactions very much simulate human-human interactions [136, 177, 64, 15]. Together, those who built on and extended the two theoretical frameworks posit users would view agents as teammates rather than as tools because they would engage socially with computers regardless.

3.1.3 Computers as cooperative tools

Unlike Nass, who was a social scientist by training, Clarke and Smyth proposed the idea of building *"cooperative computers"* through the lens of computer science and engineering [71]. Therefore, their goal was not to create a more social experience by treating computers as teammates; instead, they believed computers should be tools, but "making them (computers) more cooperative would make them more useful tools" [71]. The authors referred to Grice's principles of human cooperation [72] to set the benchmark for cooperation, consisting of providing high-quality, high-relevance support in a human-preferred manner. They suggested the first step to achieve this goal is to improve computers' processing – or "understanding" – of what users what to do (i.e., input) so that computers' output can better align with users' interests. Interestingly, because Grice developed the definition of cooperation based on studying logic in a conversation, this set the seed for *communication* to be a key running theme throughout the line of work.

Hence, Clarke and Grice's works are oft-cited and remain influential to the research and design of today's conversational agents [206, 35]. There persists much focus on agents' capabilities to understand human communication, both verbally and implicitly. The line of work values the close relationships between communication, coordination, and cooperation, and, instead of viewing machines as teammates, it suggests *building* machines as better teammates to make them more helpful to their users.

3.2 Building Humanness in Machine Teammates

Despite their various motivations, scholars have shared the consensus that computer agents can indeed work as our teammates due to their capabilities to operate autonomously, their design to offer cooperative functions, or humans' own tendency to view them as social actors. Then, how could these teamwork agents possibly elicit social experiences in collaboration with humans? In a review of existing work, I noticed most of the work remained focused on anthropomorphizing teamwork agents; namely, eliciting social experiences in human-agent teamwork through presenting more human-like agents [330, 188, 307, 213]. Therefore, in the present section, I summarize how prior work has built humanness into the design of these intelligent agents.

Anthropomorphism refers to attributing human characteristics or behavior to non-human entities [187]. Furthermore, some argued that the process of anthropomorphizing is mostly mindless [183]: it does not necessarily reflect the user's thoughtful belief that a computer has human characteristics but is rather automatic and encouraged by cues in their interfaces. According to Epley et al. [107], anthropomorphism may be a default behavior; in order to make sense of an artifact that one is not familiar with, the person may map their own lived experiences onto the novel subject. Furthermore, individuals' tendency to anthropomorphize certain objects can change over time. For instance, with acquired knowledge about a new object, one becomes less likely to anthropomorphize it; conversely, adding humanoid features to an object can trigger one to treat it as a human-like entity. In particular, Reeves and Nass [265] have summarized three basic elements that make a computer system anthropomorphic: (i) being interactive, (ii) applying natural language, and (iii) taking on a role typically performed by a human. Nonetheless, it is worth noting that there is no single factor that makes an agent human-like [278]. In other words, building a social teamwork agent requires a number of anthropomorphic features that collectively paint its humanness.

3.2.1 Humanness in appearance & emotion

Embodiment and appearance

Embodying agents in humanoid appearances, such as a full- or half-body figure or simply a talking face, is one of the most common approaches to anthropomorphism. These humanoid features typically possess body parts (e.g., limbs, eyes, nose, mouth, and hair) and can be either cartoon-like or photo-realistic [291, 100]. However, based on the idea of Uncanny Valley [229, 140], if the appearances of agents were made *too* realistic, users would show discomfort and adversarial responses.

Emotional expression and recognition

With body figures, agents can also look more anthropomorphic by showing emotions through facial expressions and other behavioral cues, which are found particularly effective in eliciting perceived anthropomorphism [35, 107, 287, 86]. In reverse, an agent capable of recognizing users' sentiments is also deemed more human-like. In a synthetic review [107], Epley and colleagues pointed to demonstrating one's emotional states and recognizing those of others as one of the most triggering attributes of anthropomorphism. In response, humans also express more emotions when interacting with anthropomorphic artifacts. Therefore, recent work in social robotics has proposed agents expressing rich emotional cues can more effectively serve social values – particularly under the context when social values are indeed beneficial (e.g., chatbot providing mental healthcare services) [277, 209, 170].

3.2.2 Humanness in verbal & non-verbal cues

Natural language and linguistic cues

Studies have consistently found that simulating natural, interpersonal languages to communicate with users is one of the most effective ways to anthropomorphism [107, 366, 281]. Advances in natural language processing (NLP), natural language understanding (NLU), and natural language generation (NLG) have adequately enhanced agents' capabilities to engage in dialogues with users [64, 226]. Most recently, agents empowered by large language models (LLM), such as ChatGPT (https://chat.openai.com/) and Bard (https://bard.google.com/) could actively interact with users through chat and respond to their inquiries like human assistants. The high competence of these agents once again sparks lively discussions about whether AI agents possess ”general intelligence” that is often considered a uniquely human trait [48].

Besides *what* an agent could communicate, *how* it delivers the verbal content is likewise a critical factor of perceived human-ness. To begin with, humans typically show clear distinctions between their oral and written languages. When people speak, they often adopt more casual terms and have more pauses, hesitation, and repetition in their speech [23, 70, 76, 285]. Therefore, prior studies have attempted to make agents sound more conversational (such as by inserting ’um’ or ’uh’ in verbal content and removing formal phrases) and found users perceived these agents as more natural and humanoid [121, 284]. Interestingly, the level of anthropomorphism also intersects with the perceived expertise of agents [93, 184, 351]. Instead of viewing it as a know-how machine, users became more cautious about suggestions from an inarticulate agent. From a safety perspective, this greatly contributes to calibrating users’ trust in agents, reducing the chances of them thoughtlessly following agents’ suggestions.

Minor linguistic cues, such as grammatical correctness and the use of pronouns, also have a key effect on shaping users’ perceptions of an agent. Besides, several important considerations come into play regarding whether and how an agent references itself in verbal content. First and foremost, as humans predominantly speak in the first-person perspective, adopting ”I,” ”me,” or ”myself” in speech is much more common in natural conversations, as opposed to adopting a third-person narrative [2, 305]. Furthermore, because self-referencing implies a sign of self-awareness and consciousness, which is yet another uniquely human trait, this again explains why the subtle switch from the first-person to the third-person point of view could be so effective on anthropomorphism of agents [112, 225, 241, 244].

Voice, tone, and non-verbal cues

While not every autonomous agent comes with a voice, bots that can communicate verbally with voices are often perceived as highly personified [112]. On the other hand, numerous studies have found users could make inferences of agents' attributes even when they were presented merely through voice interfaces [189, 293, 104, 107, 286, 281]. Users would often associate an agent's voice with features such as gender, age, and even mood and personality, based on which they form impressions of the agent. For example, in an experiment applying voice assistants for educational purposes, students showed preferences toward agents with lower voices as they were perceived as elder, more confident, and more knowledgeable [104]. In fact, even synthetic voices that sound rather robotic could encourage users to picture humanoid images of agents [13]. However, due to their tendency to associate agents' voices with human traits, scholars have also expressed concerns that users may more likely endorse stereotypes when interacting with voice agents [238, 162, 85, 198, 214].

Furthermore, voices *per se* could entail a wide range of variances that contribute to the anthropomorphism of agents. Prior research has found that moderating the prosody and tone of agents not only influenced their perceived anthropomorphism but also users' perceptions of these agents [347]. On one hand, these non-verbal cues often convey speakers' emotions in speech; on the other hand, listeners often rely on these hints to infer others' personality

[123, 126, 80]. Furthermore, non-verbal cues, such as a person's accent, could also represent the speaker's cultural background and social identity [82]; in particular, whether associations formed through these non-verbal cues align with the agents' functions could affect users' trust and reliance on the machines [318].

3.2.3 Implications of anthropomorphism

In view of the aforementioned research, there is no shortage of approaches to building human-like, social agents. Plentiful studies have found users demonstrating preferences and more positive experiences during one-on-one interaction with humanoid agents, and social robots were also found to positively mediate interactions among multiple humans [238, 236, 37, 204]. More recently, based on these physical, linguistic, and non-verbal cues, studies have further demonstrated the possibilities of anthropomorphizing agents through advanced cognitive features, such as an agent showing high-level reasoning, intentions, and sophisticated opinions through its speech [364, 167, 311, 48]. Meanwhile, users were found to engage with more emotions when agents entailed signs of empathizing, taking responsibility, or making judgment [306, 362, 84]. These caused particular concerns among scholars [348, 48, 166, 325, 326, 343, 128, 120, 152]. On one hand, less experienced users could easily mistake agency for authentic experience and capability (e.g., an agent saying "I feel happy" does not imply it could indeed "feel" and process happiness), thus, establishing inappropriate expectations for agents. On the other hand, anthropomorphism could also encourage users to engage with more social feedback; whether this would lead users to endorse similar biases or problematic behaviors as seen in human-human interaction remains largely under-explored.

CHAPTER 4

CALIBRATING SOCIAL EXPERIENCE IN HUMAN-AI COLLABORATION THROUGH PERCEIVED IDENTITY OF AGENT

Based on the literature review in the previous chapters, this and the following chapters examines the potential of human-AI collaboration in supporting individual work. As discussed in Section 2.1, social experiences in human-AI interaction can emerge from a top-down or a bottom-up approach. Thus, I examine whether and how social experiences may arise in human-AI collaboration through either of the two paths. In the present chapter (Chapter 4), I experimented with taking a top-down approach on calibrating social experiences in human-agent teams by informing users of the identity of an artificial agent (either as a bot or as a human) during teamwork. In the next chapter (Chapter 5), I further explore the utility of initiating social experiences through a bottom-up approach by anthropomorphizing teamwork agents through different degrees of affordance and capabilities in team performance.

The first study in this series of work[1] experiments with whether and how the perceived identity of a teamwork agent could calibrate social experience in human-AI collaboration through the top-down approach as described in Section 2.1.3. By explicitly informing users whether they were working with a human partner or an artificial agent, I examined users' experience and performance during teamwork. Building on the literature reviewed in Chapter 2 (particularly Section 2.2.1), I hypothesized that the notion of one's teamwork partner – that is, just by knowing whether one works with a human or not – could have

[1]The study described in this section has been published at the Conference on Human Factors in Computing Systems (CHI) as *IdeaBot: Investigating Social Facilitation in Human-Machine Team Creativity* [161]. This section adopts and revises content from the full published paper.

an effect on the emergence of social experience in human-AI collaboration. As a result, users may either benefit from social facilitation or be restrained by the social inhibition of the teamwork agents.

Under these hypotheses, I conducted three text-based studies where participants were paired with either a human or a bot collaborator to generate creative ideas for a brainstorming task. In Study 1, participants either interacted with a human (*real human*) or a bot partner (*real bot*) to work on a brainstorming task. In Study 2, all participants worked with a human confederate but were told that their partner was either a human (*real human*) or a bot (*fake bot*). In Study 3, all participants brainstormed with a chatbot but were told that their partner was either a human (*fake human*) or a bot (*real bot*). I performed systematic comparisons to understand whether individuals demonstrated different cognitive and behavioral patterns when the informed identity of their teamwork partner was either a human or an artificial confederate. Specifically, I examined the effect of perceived identity on four aspects: (1) task success and idea generation outcomes, (2) creative self-efficacy, (3) perception of teamworking partners, and (4) self-report teamwork experiences. Furthermore, I took into account how individual differences – specifically, participants' prior teamwork experiences and their self-report anxiousness in group settings – may moderate collaboration with either a human or an agent partner.

Results of the study show the potential for autonomous agents to effectively facilitate team performance in an idea generation task by serving as an active conversational partner. Moreover, as long as participants believed they were working with an artificial agent – regardless of whether they were indeed working with an agent or not – they showed improved performance in terms of both the quantity and quality of ideas they generated during the study. Participants who self-reported a higher level of anxiety in group settings yielded particularly positive outcomes when working with an autonomous agent. Together, findings in the study validate the theoretical groundwork of Social Facilitation and Inhibition [321] and Distraction-Conflict [22, 273] and suggest the effect of social presence can be extended to human-AI interaction.

4.1 Background & Related Work

4.1.1 Social facilitation and inhibition in human-agent teamwork

As reviewed in Section 2.2.1, since Triplett proposed the profound work on the Theory of Social Facilitation [321], the effect of social presence has been essential to the scholarly discussion across numerous fields. Scholars soon realized while the notion of others can elicit positive motivation that improves one's performance, it can also prevent a person from focusing on their own work [39, 101, 59, 90, 315]. These contradicting perspectives leave questions as to whether collaboration, compared to one working independently, can indeed bring out better quality of work.

In this line of theoretical discussion, Baron made critical progress by proposing the Distraction-Conflict framework [21] (see Figure 4.1 for an illustration). According to the theory, the presence of others creates an additional source of distraction, such that one's attention is now divided between the people around them and the task they are working on. These two types of events compete for one's cognitive bandwidth, cause conflicts, and raise the level of arousal in one's mental state. While a moderate degree of arousal can make a person stay more alert and concentrate on their task, excessive arousal can soon overwhelm them and prevent them from focusing on their own work [98]. The former explains Social Facilitation and the positive effect of social presence. At the same time, the latter mechanism shows why the notion of social others can hinder one's performance, resulting in Social Inhibition. Existing literature has provided empirical evidence to validate the theoretical framework, suggesting whether or not social presence can benefit one's work and whether collective work is superior to individual work have much to do with whether knowing the presence of others brings an appropriate level of alertness.

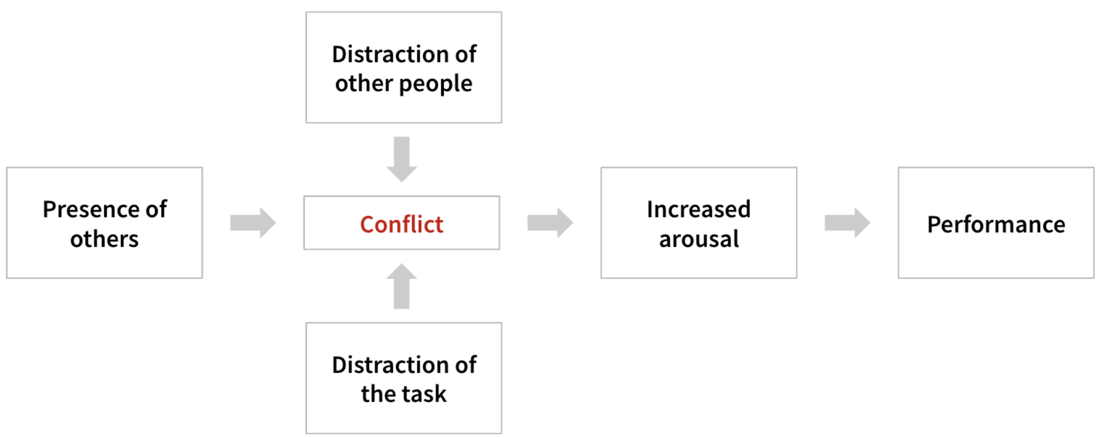

Figure 4.1: A conceptual framework of the Distraction-Conflict Theory (Adopted from [21])

Extending this concept to human-AI collaboration, I ask whether knowing one's collaborating with a bot, rather than a human, is helpful or detrimental to their performance. If knowing one's working with a bot brings the same experience as working with a human, then the make or break of collaboration may follow what the Distraction-Conflict theory proposed. If working with a bot does not trigger as much alertness as working with a human, a person may find less motivation during teamwork with an artificial agent, resulting in worse performance. Suppose the notion of working with a machine causes higher arousal and makes it even more difficult for a person to focus on one work. In that case, this may also indicate inferior outcomes of human-AI collaboration. Examining this topic helps us address fundamental questions- essentially two sides of the same coin- in human-AI collaboration: (1) Could working with an artificial agent introduce additional benefits that one can hardly attain through human-human collaboration? (2) When designing an artificial agent for collaboration, should the agent be presented as a bot or as a human entity? Together, I summarize the key inquiry that motivates the present work as follows:

RQ1: How does the informed identity of a collaborator (either as a human or as an artificial agent) influence the performance of an individual in teamwork?

In the existing literature, scholars have treated artificial agents as "passive others" and explored the effect of agents' mere presence on teamwork outcomes; namely, while these agents were present in teams, they offered a minimal amount of interaction [269, 193, 360, 146, 352, 268]. However, though these studies commonly suggest even the mere presence of artificial agents can facilitate team performance through social facilitation, whether and how humans respond socially to these agents in *real* interaction remain an active research area. Thus, a growing body of literature has emphasized the importance of studying interactive agents, positing interactive, contingent behaviors of agents may further enhance perceived anthropomorphism and lead to distinct behavioral responses from human users [282, 365]. Meanwhile, abundant literature in HRI has suggested that individuals activate a different cognitive system when interacting with robotic entities [111, 365]. Therefore, in the present study, I investigated the teamworking outcomes when participants collaborate with a confederate that is either perceived as a bot or human while assessing how the informed and perceived artificiality of the collaborative entity may elicit an effect.

4.1.2 The effect of anxiety in group communication

Whether the social presence of teammates facilitates or inhibits task success can also be moderated by individual differences; for instance, people with different personality traits (e.g., extrovert, introvert, agreeable, or critical) may demonstrate significantly different behaviors in group settings [313, 92, 97]. In particular, individual differences exist in how people feel anxious in group communication [255, 142, 292]. Previous literature [292] has summarized several conversational cues that reflect one's anxiousness in group interaction; for example, those who experience a heightened level of anxiety may apply more reserved language and demonstrate verbal intrusiveness to keep a distance from others, while those who feel more comfortable in group settings also show a higher degree of spontaneity or more expressive and sometimes even critical languages. Predictably, those who consistently demonstrated nervousness and anxiety in front of their teammates often showed inferior performance during collaboration as well [54, 340]. This phenomenon in group communication resonates with the Distraction-Conflict framework [22, 21, 273]. Since the cognitive capacity of highly anxious people is overwhelmed by concern and worries about others, anxious people cannot engage fully with the task at hand.

Along the same vein, prior HCI research had found when individuals worked with artificial agents, the degree to which they worried about social judgment and evaluation apprehensiveness was related to the agents' anthropomorphic features [180, 256]. That is, the more human-like characteristics an artificial confederate exhibited, the more concern participants demonstrated toward their partner. I propose the effect of evaluative apprehension on anxious participants' cognitive distraction will be relatively mild when a person collaborates with a partner whom they believe to be artificial or with a partner who demonstrates fewer human-like behavioral cues. However, this hypothesis has not yet been systematically tested in human-robot teamwork. Therefore, I propose the second research question:

RQ2: How does anxiety in group communication moderate teamwork and idea generation with human and bot confederates?

4.1.3 Teamwork with computer agents

Pioneering work in Human-Robot Interaction (HRI) has examined the potential of working with robots to mitigate some of the above-mentioned issues [269, 81, 146, 250]. Unlike the contributions of human collaborators, the pace and content of non-human agent responses can be fine-tuned to make other groupmates more comfortable and engaged [169]. Former research has consistently revealed that, when interacting with artificial collaborators, participants also demonstrated certain behavioral patterns similar to how they would interact with human confederates [269, 146, 81]. Classic HRI literature describes the concept of the "social robot," which suggests that the more humanoid features that are embedded in a bot (e.g., facial expressions, humanoid appearances), the more likely individuals are to view a machine-mediated agent as a human [43, 44]. Various studies have examined the effect of autonomous agents' physical appearance and concluded that the degree of embodiment could mediate the level of social facilitation [156, 250, 146]. That is, when the confederate was presented in a highly embodied form, such as in a humanoid robot, the social facilitation effect tended to be stronger. Conversely, participants were less likely to benefit from social facilitation when a non-human confederate appears in a low-level embodiment format, such as an animated image on the screen. Ongoing work in the field found robots in teams can facilitate interpersonal interaction and alleviate conflicts among other human teammates [295, 150, 319]. More recently, scholars have proposed that robots' tireless, constantly productive nature may encourage other human teammates in a group to contribute continuously as well [32, 319]. Building on these promising shreds of evidence, I examine how a team partner's informed identity – as a human or autonomous agent – could influence individuals' experience and performance in a collaborating setting.

4.2 Method

In the present research, I conducted a series of three studies to examine the effect of a teamwork partner's perceived identity. For all three studies, participants worked on a brainstorming task with a teammate through text-based communication online. In Study 1, participants worked with either a chatbot (the *Real Bot* condition) or a human confederate (the *Real Human* condition). In Study 2, all participants worked with a human teammate but were informed that their partner was either a human (the *Real Human* condition) or a chatbot (the *Fake Bot* condition). Conversely, all participants worked with a chatbot in Study 3, but their partner was described as either a chatbot (the *Real Bot* condition) or a human (the *Fake Human* condition). For all three studies, I recorded chat logs from the team conversations as behavioral data and used a post-study questionnaire to capture participants' subjective experiences during teamwork.

4.2.1 Participants and procedure

For all three studies, I recruited participants through Mechanical Turk (MTurk) (Study 1: N = 63; Study 2: N = 68; Study 3: N = 54)[2], using the following recruitment criteria: (1) participants located in the United States, (2) HIT approval rate greater than 97%, and (3) more than 500 HITs approved over the worker's lifetime. I selected MTurk as the recruitment and conducted the study online for the following reasons. First, the study was launched admist the worldwide COVID-19 pandemic, such that conducting in-person lab studies was not encouraged by my affiliated institution (Cornell University) at that time. Second, recruiting online allows me to access participants of more diverse demographic backgrounds and those who have experienced collaboration in real work scenarios. By contrast, recruiting for in-person studies on campus would mostly attract University students, resulting in a more homogeneous sample with limited experiences as working professionals. Finally, the form of communication (i.e., online chats through text) applied during teamwork sessions for the present study did not at all demand participants to interact with their teammates in person. As a result, I did not find a significant difference in the teamwork quality regardless of

[2]Before the formal data collection process, I conducted a pilot study with 28 participants to ensure the validity of measurements used in the research and refined the study protocol. Using G*Power 3.1, I also ran a power analysis for an independent sample t-test (α = .05) to determine the target sample size, which is around 30 participants per condition.

whether they were present in a lab for the study.

The experiment was distributed and conducted entirely online via Qualtrics. Participants who consented to participate first completed a pre-survey reporting their self-evaluated creativity [317] and level of anxiety in group communication [292]. They then worked on a brainstorming task with a conversational partner through online chat. In conditions where participants were assigned to work with a human partner, a research assistant would follow a script and engage in team conversations with participants via Chatplat[3]. This online chat platform has been used in a number of empirical studies (e.g., [46, 157, 349]). In conditions where participants were assigned to work with a bot, they would interact with a chatbot designed and deployed through Juji.io[4]. The chatbot was coded with the same script as the research assistant used to play the human teammate, ensuring the content and the ideas provided by the chatbot were identical as in those conditions when participants worked with a human confederate. Participants who completed the study were compensated with $7 cash. For all studies, all participants provided informed consent, and all protocols were reviewed and approved by the Institutional Review Board (IRB).

Both the research assistant and the chatbot applied the same conversational flow when interacting with all participants. They would first greet the participants, introduce itself, and invite participants to start working on the task. Following this, the chatbot or the research assistant would propose the first idea, ask participants' opinions about the idea, and ask whether they had other ideas in mind. If participants did not have an idea to contribute, the chatbot or the research assistant would move on to propose the next idea. During each 10-

[3]https://www.chatplat.com/
[4]https://juji.io/

minute brainstorming task, the chatbot or the research assistant could potentially contribute as many as ten ideas. These ideas were collected from the top ten ideas that participants most often came up with in a previous study that used the same prompt [301], and they were presented to participants in a random order. Besides, if participants said something that the chatbot could not process, the default response was, ”Tell me more about that!” It would then wait for participants’ response and move on to propose the next idea.

4.2.2 Measures

Through the study sessions, I captured participants’ team brainstorming experience and outcomes through a diverse set of behavioral and self-report variables. First, I recorded the chat logs of all team conversations chat transcripts of during all conversations and analyzed both the quantity and quality of ideas generated by participants. Besides, the pre- and post-surveys included various self-report variables capturing participants’ teamwork experience and their perceptions toward their team partners. All items in the surveys were measured on 7-point Likert scales unless specified otherwise. Table 4.1 provides descriptive statistics of all quantitative measures.

Idea generation outcomes were coded from participants’ chat transcripts and evaluated on six measures [274, 174]. These include (1) *the number of ideas generated*. (2) *The length of idea description* calculated by word-counts of each idea. (3) *Originality of ideas*, coded on a 3-point scale. An idea’s originality was coded as 1 if it is an existing idea, as 2 if participants combined an existing idea with their own new input, and as 3 if it was a completely new idea that has

Table 4.1: Descriptive Statistics of the Three Studies (Adopted from [161])

Experiment	Study 1		Study 2		Study 3	
	M	*S.D.*	*M*	*S.D.*	*M*	*S.D.*
Anxiety in group communication	3.87	1.23	4.36	1.35	4.12	1.26
Idea generation outcomes						
Count of ideas	5.31	3.55	4.03	3.06	4.65	3.36
Length of idea	19.33	10.83	14.70	7.94	16.90	9.69
Originality	1.27	0.36	5.11	3.07	3.27	2.94
Load of information	1.80	0.54	1.46	0.61	1.63	0.59
Logic with oneself	0.25	0.20	0.26	0.20	0.24	0.23
Logic with partner	0.28	0.27	0.27	0.17	0.28	0.21
Creative self-efficacy	5.06	1.29	5.38	1.01	5.24	1.18
Perception toward partner						
Perceived dominance	2.70	1.66	3.40	2.06	3.89	1.79
Perceived productivity	5.95	1.07	5.90	1.09	6.00	1.03
Perceived creativity	4.77	1.88	5.11	1.34	4.96	1.58
Concerns about partner	3.48	1.83	4.40	1.85	3.93	2.08
Conversational experiences						
Express thoughts	6.38	0.87	6.25	0.85	6.24	0.79
Pause and wait	5.92	1.29	5.78	1.12	5.44	1.69
Pressure to contribute	3.81	2.01	4.28	1.81	4.70	1.89

not been commonly practiced for water and energy conservation [**?**]. (4) *Logic with partner* was coded in a binary fashion, capturing whether an idea is (1) or is not (0) related to what their partner previously proposed. (5) Similarly, *Logic with oneself* was also a binary measure, denoting whether the idea connected to what participants previously mentioned themselves or not. (6) *Load of information* in ideas was coded on a scale of 3-point scale, measuring the amount of information embedded in the idea.

Creative self-efficacy. Participants evaluated their own performance after the teamwork session using the *Creative Self-Efficacy* scale adopted from Tierney et al.'s literature [317].

Anxiety in group communication were adopted from Stephen et al.'s literature [292]. Per the results of a Shapiro test, participants' anxiety scores in all three studies were normally distributed (Study 1: p = 0.373; Study 2: p = 0.276;

Study 3: p = 0.405).

Perception toward teammate measured participants' perceived dominance and productivity of their partner. Given the context of the present experiment (i.e., teamworking on a brainstorming task), I also asked participants to evaluate how creative their partners were.

Conversational experiences capture whether participants could freely express their thoughts in the conversation, whether they paused and waited for their partners while noting they were inserting messages as the (...) sign showed up, and whether they felt pressured to come up with ideas when they were asked to contribute to the task [240, 230].

Manipulation check. To ensure successful experimental manipulation, I asked participants whether they considered their conversational partners as "absolutely bot" or "absolutely human." Participants responded to this question using a 100% slider scale.

4.3 Results

4.3.1 Overview of data analysis

I first used two-tailed, independent sample t-tests to evaluate the main effect of experimental conditions (the informed identity of a teammate as either a human or a bot) on dependent variables. To test the significance of moderating effects, I applied general linear models from the `lmer` package in R, including participants' unique IDs as a random factor. Finally, I performed path analysis to test mediation, using the single-mediator SEM model of the lavaan package in R. For analysis measurement without a fixed scale (e.g., the total count of ideas generated, the length of idea description), I used standardized z-scores for analysis. Furthermore, based on convention in statistics literature [308], I excluded data with values outside the range of $M \pm 3.29S.D.$ to mitigate the effect of extreme values. All statistics of the three studies were reported in Table 4.2.

4.3.2 Manipulation check

I began the data analysis by assessing whether the experimental manipulation was successful. I used a two-tailed independent sample t-test to check and evaluate participants' responses to the question of whether their teammate was a bot or a human. Test results showed successful manipulation as those who were told they were working with a human indeed believed their team partner was a human, and vice versa (Study 1: $t = -9.11, p < 0.001$; Study 2: $t = 3.46, p < 0.001$; Study 3: $t3.87, p < 0.001$).

4.3.3 Idea generation outcomes

Across the three studies, participants yielded better performance in the idea generation task as long as they believed they were working with a teamwork agent instead of a human, such that they generated a significantly greater number of ideas (Study 1: $t = 6.24, p < 0.001$; Study 2: $t = 3.83, p < 0.001$; Study 3: $t = 3.83, p < 0.001$) and ideas described with greater lengths of details (Study 1: $t = 3.46, p = 0.001$; Study 2: $t = 3.88, p < 0.001$; Study 3: $t = 3.88, p < 0.001$). Furthermore, two research assistants were trained to rate the quality of these ideas and reached satisfying inter-rater reliability (with an agreement rate greater than 85% for all three studies). I again conducted t-tests with these ratings and found participants generated ideas with better qualities when the informed identity of their teammate was a bot, leading to ideas with significantly greater originality (Study 1: $t = 4.47, p < 0.001$; Study 2: $t = 2.75, p = 0.008$; Study 3: $t = 2.75, p < 0.01$), a greater load of information (Study 1: $t = 4.83, p < 0.001$; Study 2: $t = 2.93, p = 0.002$; Study 3: $t = 2.93, p < 0.001$), and increased logical connections across their own ideas (Study 1: $t = 3.43, p < 0.001$; Study 2: $t = 4.48, p < 0.001$; Study 3: $t = 4.48, p < 0.001$), while there was no significant difference in whether participants connected their own ideas with their teammate's (Study 1: $t = 1.67, p = 0.100$; Study 2: $t = 0.66, p = 0.090$; Study 3: $t = 0.66, p = 0.091$).

Additionally, I also qualitatively examined patterns in participants' idea generation processes with their team partners using linkographs [138]. These linkographs demonstrate the number of ideas using nodes and their connectivity using links. The higher density in these "webs" of ideas showed greater relevance between former and later ideas. The distance between connected dots shows whether related ideas came from adjacent ones or from those contributed much earlier in a conversation. When investigating participants' linkographs (see Figure 4.2 for examples), I saw more complex, interrelated patterns of the brainstorming process of high-performing participants when they believed they were working with a bot partner, whereas linkographs of high-performing participants who believed they were working with a human teammate were relatively sparse.

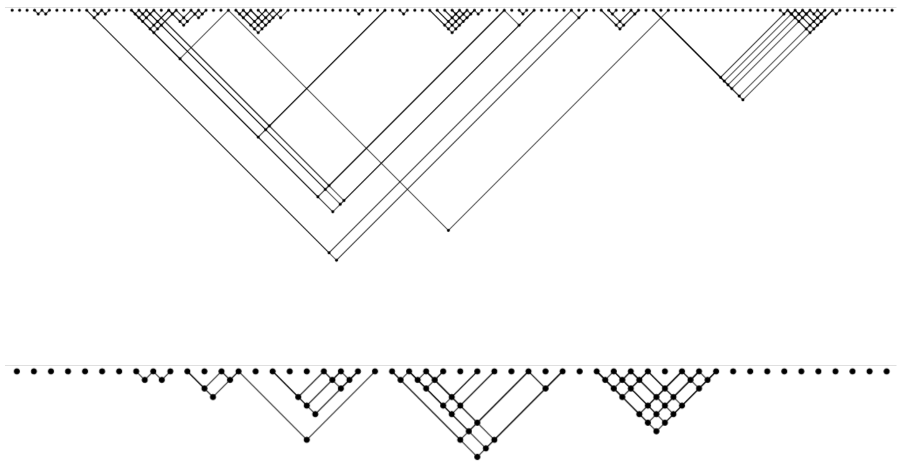

Figure 4.2: Examples of high-performing participants' Linkographs in Study 3 when the informed identity of the team partner was a bot (upper) vs. a human (lower). (Adopted from [161])

4.3.4 Creative self-efficacy

Based on an independent sample t-test, there was no direct main effect of experimental conditions on creative self-efficacy (Study 1: $t = 0.73, p = 0.500$; Study 2: $t = .53, p = 0.600$; Study 3: $t = 0.79, p = 0.500$). However, results from Study 1 and Study 3 showed that participants' creative self-efficacy could be marginally moderated by their self-reported anxiousness in group communication settings (Study 1: $F(3, 59) = 1.64, R^2 = .07, p = 0.042$; Study 2: $F(3, 59) = 0.43, R^2 = .02, p = .470$; Study 3: $F(3, 50) = 1.01, R^2 = .06, p = 0.090$), such that highly anxious participants reported significantly higher creative self-efficacy when they believed they were brainstorming with a bot, compared to those who thought they were interacting with a human confederate.

4.3.5 Perception toward teamworking partners

Through independent sample t-tests, I found that participants rated their bot partner as more dominant than the human confederate (Study 1: $t = 2.82, p = 0.006$; Study 2: $t = 1.91, p = 0.060$; Study 3: $t = 3.30, p = 0.002$). However, participants' perceived productivity (Study 1: $t = -0.34, p = 0.700$; Study 2: $t = 2.63, p = 0.010$) and creativity (Study 1: $t = -1.55, p = 0.100$; Study 2: $t = 2.84, p = 0.006$) of their partners did not show a consistent trend across the three studies.

4.3.6 Conversational experiences

Results from Study 1 and Study 2 showed that participants felt greater pressure to come up with ideas during the teamwork session when they believed they were working with a bot rather than a human (Study 1: $t = 3.08, p = 0.003$; Study 2: $t = 2.17, p = 0.033$; Study 3: $t = 0.19, p = 0.854$). However, there was no consistent trend revealed across the three studies in terms of whether participants could freely express themselves in team conversations (Study 1: $t = 1.52, p = 0.100$; Study 2: $t = 0.21, p = 0.80$; Study 3: $t = 2.60, p = 0.012$).

Table 4.2: Summary Statistics of the Three Studies (Adopted from [161])

Experiment	Study 1					Study 2					Study 3				
Condition	Real Bot	Real Human				Fake Bot	Real Human				Real Bot	Fake Human			
Perceived identity	Bot	Human	X	X*W	M	Bot	Human	X	X*W	M	Bot	Human	X	X*W	M
Conversational style	Bot	Human				Human	Human				Bot	Bot			
Idea generation															
Count of ideas	6.85	3.62	B¿H			5.29	2.70	B¿H			6.04	3.19	B¿H		
	(3.72)	(1.76)				(3.36)	(1.03)				(3.16)	(1.39)			
Length of ideas	20.10	17.44	B¿H			16.20	13.01	B¿H			18.06	15.06	B¿H		
	(9.38)	(12.27)				(7.63)	(8.06)				(8.53)	(10.82)			
Originality	1.39	.76	B¿H			1.61	1.22	B¿H			1.06	.76	B¿H		
	(.67)	(.40)				(.57)	(.35)				(.66)	(.51)			
Info load	1.95	1.07	B¿H			1.68	1.32	B¿H			1.78	1.46	B¿H		
	(.86)	(.53)				(.55)	(.58)				(.52)	(.63)			
Logic w/ self	.31	.11	B¿H			.34	.16	B¿H			.32	.14	B¿H		
	(.20)	(.09)				(.57)	(.09)				(.25)	(.18)			
Logic w/ partner	.25	.24	B=H			.25	.27	B=H			.25	.26	B=H		
	(.12)	(.13)				(.13)	(.15)				(.13)	(.14)			
Creative self-efficacy	5.17	4.94	B=H	sig.		5.45	5.32	B=H	n.s.		5.35	5.14	B=H	sig.	
	(1.25)	(1.34)				(.97)	(1.07)				(.81)	(1.44)			
Partner															
Dominance	3.25	2.13	B¿H		sig.	2.94	3.38	H¿B		n.s.	3.18	4.65	H¿B		sig.
	(1.80)	(1.31)				(1.94)	(2.10)				(1.44)	(1.81)			
Creativity	5.34	5.10	B=H		n.s.	5.07	5.58	B¡H		n.s.	5.04	5.21	B¿H		n.s.
	(1.31)	(1.25)				(1.06)	(1.25)				(1.28)	(1.26)			
Productivity	5.91	6.00	B=H		n.s.	5.57	6.24	H¿B		n.s.	6.15	5.86	B=H		n.s.
	(1.12)	(1.03)				(1.27)	(.75)				(.68)	(1.27)			
Concern	3.50	3.45	B=H		n.s.	3.91	4.91	H¿B		n.s.	3.81	4.04	B=H		n.s.
	(2.06)	(1.59)				(1.99)	(1.57)				(1.88)	(2.28)			
Conv. Experiences															
Expressiveness	6.22	6.55	B=H		n.s.	6.23	6.27	B=H		n.s.	5.96	6.50	B¿H		n.s.
	(.91)	(.81)				(.91)	(.80)				(.77)	(.74)			
Pressure	4.53	3.06	B¿H		sig.	3.83	4.76	H¿B		n.s.	4.65	4.75	B=H		n.s.
	(1.95)	(1.82)				(1.67)	(1.85)				(1.65)	(2.12)			
Pauses	5.69	6.16	B=H		n.s.	5.83	5.73	B=H		n.s.	4.92	5.93	B¿H		n.s.
	(1.42)	(1.10)				(.95)	(1.28)				(1.90)	(1.33)			

B: Bot condition; H: Human condition; X: Main effect of experimental conditions; X*W: Moderating effect of the level of anxiety in group communication; M: Mediating effect of perception toward partners or conversational experiences

4.4 Discussion

Throughout a series of three studies, I consistently found participants generated more ideas and contributed ideas with better quality when they perceived their collaborator as an autonomous agent. These findings can be explained as follows. First, based on Social Facilitation theory [5, 321, 359] and the Distraction-Conflict framework [22, 21, 273], I propose that working with a non-human agent may reduce concerns about teammates' social evaluation. This could allow users to better concentrate on their own brainstorming processes, leading to more fruitful outcomes for idea generation. Secondly, though not directly measured in the quantitative data, participants' qualitative feedback at the end of the study described working with an agent as a ”cool” and ”fun” experience and was ”more interesting than expected.” Furthermore, some participants even treated teamworking with an agent as a game and attempted to come up with more ideas than their bot teammate did. Long suggested by the creativity literature [294, 174], being able to enjoy a work process is key to one's success in innovation.

Furthermore, I found that certain outcomes of interacting with a non-human agent could be dependent on the conversational styles applied by the confederate. In particular, though I observed a moderating effect of individuals' group anxiety in Study 1 and Study 3, the pattern was not significant in Study 2. Previous research in HRI and anthropomorphism suggested humans have the tendency to view other subjects as humans when they perceive humanoid cues (i.e., a human conversational style as applied in Study 2) [365, 118]. This effect may be more prominent among highly anxious participants since they were already quite occupied with engaging in a team setting. With less cognitive control, highly anxious participants may treat their human-acting *fake bot* partner more like a human partner. Therefore, the positive effect of perceived bot identity did not affect these anxious participants in Study 2.

Findings of the present research, specifically those relevant to the conversing style of non-human confederates, offer rich insights for the design of future teamwork agents. To begin with, an autonomous agent can facilitate individuals' work without anthropomorphizing. In fact, signaling social-behavioral cues that are more robotic can be beneficial to solving a creative teamworking task. Additionally, the notion of working with an artificial entity could also bring more novel, interesting experiences as well as benign competition, which, all together, could contribute to greater engagement and innovative outcomes of human-AI collaboration.

I suggest the future design of artificial agents can adopt a greater degree of user control over bots' conversational styles to facilitate teamworking and idea generation. Specifically, when I take into account individual differences in the level of anxiety in group communication, a robotic-sounding or a human-like conversational style may benefit different groups of participants, respectively. Anthropomorphic features (e.g., physical appearance, facial expressions, emotional responses), as well as conversational cues (e.g., pace, formal or informal language, use of emojis, paraphrasing, and pausing to wait for users), can be promising areas for user-enabled customization.

Similar design implications may also be insightful for designing autonomous agents for educational purposes, though further research is required. In fact, the brainstorming task used in the current study simulates in-class activities that are commonly applied to learning. The task not only asked participants to generate creative ideas for a specified topic, but the brainstorming process also required them to connect and apply existing domain knowledge in the topic area [90]. To encourage students to speak up and express their ideas more openly, designers and engineers of bots for classrooms may consider the positive effect of robotic-signaling features on participants who demonstrate anxiety in groups.

CHAPTER 5

CALIBRATING SOCIAL EXPERIENCE IN HUMAN-AI COLLABORATION THROUGH PERCEIVED CAPABILITY OF AGENT

In the previous chapter (Chapter 4), I found that as long as participants believed they were working with an agent, reduced concerns for social evaluation allowed them to focus more on the tasks at hand, which can effectively enhance their performance at work. In this regard, to what extent should a teamwork agent offer helpful, high-quality content in order to support human work? In other words, how intelligent or capable should such teamwork agents be in order to effectively support teamwork? In this section, I thus further explore the effect of a teamwork agent's capability – or more precisely, perceived capability evaluated by its users.

To address this research inquiry, I began with reviewing the literature on human-agent team relationships, where I noticed a recurrent theme in scholarly discussions: despite acknowledging the constantly advancing capabilities of autonomous agents, users may still resist working with such machines – or, more precisely, refuse to receive the help from then artificial entities. To dive deeper into the causes of this phenomenon, I reviewed and drew insights from the Theory of Social Comparison, which is oft-cited to explain why humans do not necessarily accept help from and collaborate with the most competent. Accordingly, I designed two teamwork agents: a *Smart Bot* which provides high-quality, innovative ideas for a team brainstorming task, and a *Dumb Bot*, which offers low-quality, mundane ideas for the same task. I again observed how users interacted with these agents and recorded their self-report and behavioral data.

I found that the Smart Bot, though perceived as more capable of delivering high-quality teamwork, was perceived as less humanoid and discouraged participants' contributions to the team. This led to participants' negative perceptions of their own performance in the teamwork and decreased their willingness to cooperate with their bot partner in a subsequent task. By contrast, participants generated more ideas when they worked with the Dumb Bot, leading to more positive views of their own performance and an increased likelihood of taking suggestions from their bot partner. In the following sections, I elaborate on related work, methods, findings, and implications of the present study.

5.1 Background and Related Work

5.1.1 Beyond trust: When do humans rely on agent teammates?

Trust and reliance attract key research interests in human-agent teamwork. Based on Lee and See's oft-cited theoretical framework [197], scholars commonly conceptualize users' trust in an autonomous agent based on its purpose, process, and performance. However, as [216] proposed recently in their theoretical framework, users' trust toward intelligent agents is a multidimensional construct – beyond agents' performance, users are commonly affected by other motivational, interpersonal, and social aspects in human-agent interaction [179, 203, 154]. In this study, I focus on the perceived capability of an autonomous agent. In theory, a more competent agent should produce more trust and lead to greater compliance from users. Still, social comparison might reduce users' willingness to rely on the agent. In this regard, I propose two competing hypotheses for the present research:

H1a: Users are more likely to rely on and comply with the Smart Bot in teamwork, *due to its heightened competence and trustworthiness.*

H1b: Users are more likely to rely on and comply with the Dumb Bot in teamwork, due to positive experiences during interaction.

Existing work has studied various moderating factors in the trust-reliance relationship of human-agent interaction, providing insights into where things may "go wrong" even when agents are equipped with sufficient capacity to carry out teamwork tasks. While abundant work has investigated the influence of agent's performance (e.g., reliability, error rate), teamwork task (e.g., the level of risks and uncertainty in task scenarios), and personal characteristics of users (e.g., personality, prior experience with a bot), few studies have examined the effect of users' psychological experiences during and after working with intelligent agents. In other words, existing literature has informed how the quality of an agent can influence users' trust and reliance, while I intend to further explore how experiences of social comparison during human-agent interaction may influence users' perception of themselves and their reliance on the machine teammates. To address this topic, I built on prior work in social psychology and organizational behaviors, which has long studied how individuals' reflection of their own performance in group settings may affect their subsequent behavior through the concept of social comparison, which I elaborate further in the following section.

5.1.2 Psychological experience of social comparison

Comparing ourselves to others is one of the most innate human tendencies, and it drives a wide range of complex emotions in social settings [91, 264]. The

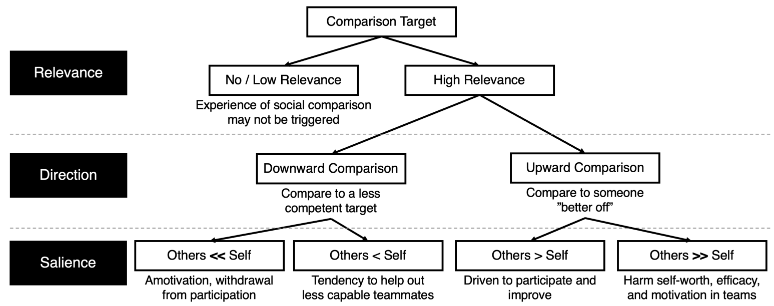

Figure 5.1: Three essential components of social comparison: Relevance, Direction, and Salience

emotional experiences of social comparison require assessment and reflection on one's own performance, resources, or trajectory compared to others. This emotional tendency often provides an instant reaction upon sensing the difference between others and ourselves. As a result, perceptions of differences between ourselves and others can, directly and indirectly, influence subsequent sentiments and behaviors [258, 264]. Festinger was one of the first to formally construct humans' tendency of social comparison in his **Theory of Social Comparison** [116]. Along with subsequent theoretical frameworks [346, 73, 232, 314], the line of research proposes three cornerstones of social comparison: relevance, direction, and salience.

Relevance concerns whether a compared target is close to one's own situation. If one perceives low or no relevance to a comparer, social comparison may not occur at all. For instance, one may feel jealous if a colleague on the same team receives a higher salary but are indifferent to Elon Musk's annual earnings. Then, do humans relate and compare themselves to computer agents? Prior work commonly established on the "Computer as Social Actors (CASA)" paradigm [237] and have shown humans did demonstrate social comparison – like how they would compare themselves to other human entities – when interacting with agents. As human users treat and interact with computer-mediated agents as social actors, I also hypothesize they would socially compare themselves to a teamwork agent.

Direction indicates whether a person compares oneself to someone better (upward comparison) or someone less competitive (downward comparison). In teamwork, individuals are more sensitive to the difference in their own performance compared to others, setting the ground for upward or downward comparison [233, 232]. Furthermore, they demonstrate distinct traits when experiencing these two opposite types of comparison, which is further moderated by the salience of perceived difference between a person and their target.

H2: Human users are likely to demonstrate upward comparison when they work with a more competent agent (i.e., the Smart Bot, which provides high-quality ideas) and downward comparison when working with a less competent agent (i.e., the Dumb Bot which provides low-quality ideas).

Salience describes the extent of the perceived difference between oneself and their compared target. Often, individuals exhibit downward comparison to secure a certain victory, especially when they perceive threats to self-esteem or in a competitive environment [314]. In teams, prior research has also found individuals are more likely to provide help to less competent colleagues. However, working with substantially less capable teammates can cause demotivation [67, 346]. On the contrary, while modest upward comparison can encourage one's improvement and contribution at work, working with a considerably more capable teammate can harm one's perceived efficacy, motivation, and self-worth in teams [316, 344, 345].

H3a: In downward comparison with the Dumb Bot, human users may become dis-

interested in working with an agent if the perceived performance difference is excessive.

H3b: In upward comparison with the Smart Bot, human users may become disinterested in working with an agent if the perceived performance difference is excessive.

5.1.3 Social comparison with intelligent agent

To date, the theories and concepts of social comparison have most often been applied to study human-to-human comparisons. In contrast, work studying human-to-agent comparison still needs to be explored. Moreover, among the limited work investigating direct comparison between humans and agents, such as [190, 231], the two parties have often been treated as competitors (i.e., a participant and a bot would work on an identical task independently then compare the outcomes of their performance) instead of collaborators. However, scenarios in which humans and autonomous agents work toward the same goal better match the usability of today's AI-empowered applications.

More recent work in human-robot or human-chatbot interaction has adopted autonomous agents as collaborators in teams [151, 169, 354, 298]. In these studies, bots may serve as conversational moderators, passive presenters, or active team contributors. Though these studies consistently demonstrated users could become more driven during teamwork when interacting and comparing themselves to bots, the complex experiences of social comparison have not been fully captured. As I elaborated in the previous section, social comparison can frequently elicit negative team experiences, mainly due to perceived inferiority in one's ability compared to teammates'. Additionally, based on work reviewed in Section 5.1.1, negative team sentiments during teamwork can serve as influential moderators, such that an agent's technical capabilities do not guarantee successful human-agent collaboration. Together, this highlights that our understanding remains scarce regarding whether and when social comparison can cause detrimental outcomes in human-agent teamwork. Therefore, I post this overarching research question for the present study:

RQ1: When and how experiences of social comparison may potentially harm human-agent teamwork?

Synthesizing the literature review, I first identified several research gaps in existing HCI research on social comparison and highlighted the need to examine the impact of social comparison on human-agent teamwork, as users' experiences of interacting and comparing themselves to a machine teammate can potentially moderate their behaviors during teamwork, leading to the making or breaking of collaboration. As in prior work, I build on the CASA paradigm, assuming users would treat computer agents as social actors. Based on this theoretical foundation, I propose teamwork agents can be targets of social comparison. I manipulate the degree of agents' competence by building the Smart Bot and the Dumb Bot. This allows me to examine users' responses to agents during upward and downward comparisons. Furthermore, I examine the effect of salience in users' perceived differences with their teamwork partners. In particular, I investigate whether greater perceived differences in performance could harm human-machine partnership [116]. I operationalize this conceptual framework (Figure 5.2 to two teamwork tasks: a co-brainstorming exercise session where participants would come up with new ideas collaboratively with their bot teammates and a color block task where participants need to make collective decisions with the bot. I capture participants' behavioral changes, if any, through three dimensions, including (1) their actual team contribution measured by the number of ideas they generated during team brainstorming, (2) subjective team contribution measured by self-report questions, and (3) reliance and compliance with teammates measured by whether they adopt the agents' suggestions during the co-decision-making process.

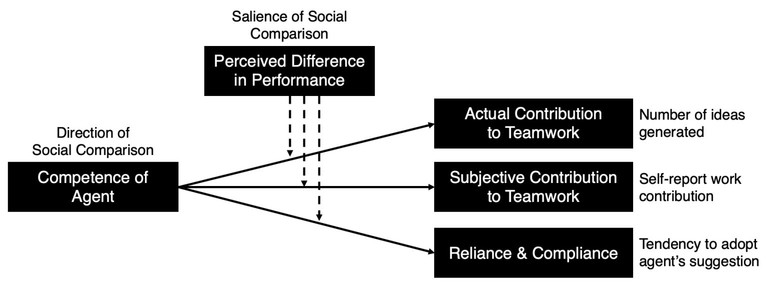

Figure 5.2: Conceptual framework for the present research

5.2 Method

The present study consists of four components: (1) A pre-survey to perform ground-truth labeling for the pool of ideas I later used to build the Smart Bot and the Dumb Bot. (2) A Wizard-of-Oz experiment to examine participants' responses to the Smart Bot and the Dumb Bot. (3) An experiment replicating findings from the Wizard-of-Oz study while using chatbots to autonomously carry out teamwork with participants. (4) A post-survey to evaluate team performance through crowdsourcing ratings. All parts of the present study were reviewed and approved by the Institutional Review Board, and all participants signed informed consent. All parts of the study protocol were reviewed and approved by the Institutional Review Board (IRB).

5.2.1 Rating idea quality

Adopting a total of 401 ideas generated by participants in the previous study (Chapter 4), I launched a survey asking 89 MTurk participants to rate these ideas (61.90% identified as male, 34.52% as female, 3.57% as other). Each participant rated 50 randomly assigned ideas by how practical, creative, and informative each idea was using 7-point Likert scales [274]. I then averaged values from the three scales for each idea ($M = 4.50, S.D. = 0.61$) and labeled those with top 25% ratings as high-quality ideas (i.e., smart ideas: $M = 5.25, S.D. = 0.26$) and bottom 25% ratings as low-quality ones (i.e., dumb ideas: $M = 3.70, S.D. = 0.34$). Results of the independent sample t-test further confirm significant mean differences between the high-quality and the low-quality ideas ($t = -34.98, p < .001$).

5.2.2 Human-agent teamwork experiments

Based on the set of high- and low-quality ideas, I conducted two between-subject, single-factor experiments with two conditions, such that participants were randomly assigned to either work with the Smart Bot or the Dumb Bot on a team brainstorming task. The first experiment was conducted through a Wizard-of-Oz protocol, where a research assistant played the teamwork agent using a pre-written script. The second experiment followed the same procedure to replicate the findings of the first experiment, while it was entirely carried out by autonomous chatbots. In the Smart Bot condition, the research assistant would randomly draw from the pool of high-quality ideas; in the Dumb Bot condition, low-quality ideas were randomly presented to participants. All study sessions were conducted online, collecting data through a Qualtrics questionnaire.

Participants.

For both experiments, I recruited participants (N = 121 for Experiment 1; N = 216 for Experiment 2)[1] on Amazon Mechanical Turk (MTurk), applying the following criteria: (1) participants had completed at least 500 MTurk HITs with an approval rate greater than 96%, (2) participants were located in the United States, and (3) participants had not previously participated in the pre-survey to provide ratings for ideas. Participants received US$7 as compensation. Due to similar reasons as described in Section 4.2.1, I was motivated to recruit and conduct the study online. All parts of the present study were conducted during the worldwide pandemic, while I still aimed to recruit participants with diverse backgrounds and work experiences.

Procedures.

The full study flow is illustrated in Figure 5.3. After consenting, participants responded to a pre-task questionnaire, which surveyed their self-efficacy in tackling the brainstorming challenge. Participants then moved to a brainstorming task (the same one used in Chapter 4), which asked them to come up with new ideas for water and/or energy conservation [301]. Participants first spent 3 minutes generating as many ideas as possible on their own. Next, participants were informed that they would be working with an AI chatbot on the same brainstorming task. In Experiment 1, they would instead be working with a research

[1] The sample size was again determined by performing a power analysis based on data collected from a pilot study with 21 participants.

assistant, while participants in Experiment 2 worked with a fully autonomous chatbot.[2] Each brainstorming session took 8 minutes. After the brainstorming session, participants completed a post-task survey about their teamwork experiences.

Participants then worked on a second task, the Color Block task (adopted and revised from [312]; also see Figure 5.4). During this task, participants saw seven color blocks and would get a unique value when clicking on each block. Their job was to pick which block yielded the highest value. All participants would get the highest value (70 points) if they clicked on the black block and the lowest scores when they clicked on the white block. They first had five trials to explore the values of different blocks on their own. Then, they joined a team discussion with a team partner (a research assistant in Experiment 1 or a chatbot in Experiment 2). They were told to jointly select **one** block that yields the highest scores for each team. During the conversation, the team agent would offer conflicting information, suggesting the white block would give the highest points. When participants recognized the conflict, the bot would suggest choosing the gray block in the middle as a compromise. After leaving the conversation, participants were asked which block they would choose as a final decision for their team. I examined how far participants shifted away from the black block to measure the extent to which they were affected and agreed with their teammate. Finally, participants reported their demographic data in an exit survey.

[2] Team interaction with the research assistant was again carried out through text-based communication on Chatplat, and the autonomous chatbot in Experiment 2 was again deployed through Juji.io.

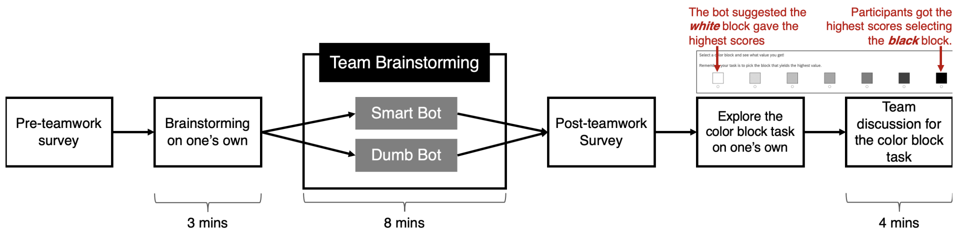

Figure 5.3: The study flow of the present experiment.

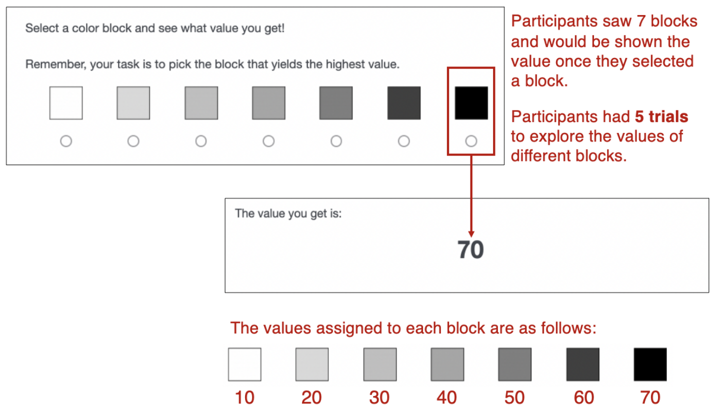

Figure 5.4: An illustration of the Color Block task. Participant had 5 trials to explore the values of different blocks. Their goal was to choose which block yields the highest value with their team partner.

Measurement

All quantitative variables in the study were measured on 7-point Likert scales unless specified otherwise. I report their descriptive statistics in Table 5.1. All multi-scale measures showed satisfactory inter-scale reliability (Cronbach's $\alpha > .70$). The variables measured in the study were selected to capture (1) *perceived difference in performance*: participants rated the performance of their teammate and themselves. I calculate the difference between the two ratings to measure the salience of performance difference in social comparison. Positive scores indicate downward comparison (i.e., participants viewed themselves as better off than their teammates), while negative scores represent upward comparison (i.e., participants considered themselves as less competent than their teammates). (2)*Subjective team contribution*. Participants rated their perceived contribution of themselves and their teammate (1= "My teammate contributed more"; 7= "I contributed more") [66]. (3) *Actual team contribution*. I coded the number of ideas generated during each team conversation based on the chat transcript of each brainstorming session. (4) *Reliance and compliance with the teammate.* I used participants' selection of color blocks as a scale to measure how willing they were to comply with the teammate's suggestions when conflicting opinions arose in teams. I also asked participants to describe their teammate using open-ended text and to report their demographic data.

	Experiment 1		**Experiment 2**	
	Dumb Bot	Smart Bot	Dumb Bot	Dumb Bot
Perceived Difference in Performance				
Partner performance	5.11 ± 1.27	5.78 ± 1.16	5.37 ± 1.05	5.4 ± 1.04
Self-performance	5.33 ± 1.15	5.31 ± 1.35	5.57 ± 0.84	5.56 ± 0.97
Subjective Team Contribution				
Work contribution	4.74 ± 1.43	3.90 ± 1.75	4.45 ± 1.36	4.65 ± 1.02
Actual Team Contribution				
Ideas generated	3.19 ± 2.37	2.13 ± 1.61	6.29 ± 3.58	8.18 ± 3.67
Reliance and Compliance with Teammate				
Color block decision (Probability to select the Black block)	0.47 ± 0.50	0.59 ± 0.50	0.54 ± 0.50	0.49 ± 0.51

Table 5.1: Descriptive statistics of variables in Experiment 1 & 2

5.3 Results

I conducted all the analyses using the lmer package in R [261] to construct linear mixed-effects models, as they are more robust in addressing individual variability [133]. For all linear models, I controlled for the random noise of individual subjects [27]. I first tested the main effect of the partner's competence (manipulated through the Smart Bot vs. Dumb Bot conditions) on perceived performance difference, subjective, and actual team contribution. I then examined participants' reliance on the teamwork agent through their responses in the subsequent Color Block task (1 = reject the agent's suggestion and stick with the Black block; 0 = comply with the agent and switch to the White block). Finally, I examined how the salience of social comparison may moderate the effect of the partner's competence by testing their interaction effect in linear mixed models. Table 5.2 summarized the statistical tests I performed for data analysis.

	Experiment 1			**Experiment 2**		
	β	$S.E.$	p	β	$S.E.$	p
Main effect of agent's competence (statistics reported for the effect of experimental condition)						
`dependent variable ~ condition + (1|subject ID)`						
Performance difference	-0.69	0.21	0.001**	-0.59	0.12	<0.001***
Work distribution	-0.84	0.30	0.005**	-0.33	0.19	0.035*
Ideas generated	-1.05	0.40	0.010*	-3.41	0.50	<0.001***
Color block decision	0.47	0.37	0.206	0.69	0.29	0.017*
Moderating effect of perceived difference (statistics reported for the interaction term)						
`dependent variable ~ condition * perceived difference + (1|subject ID)`						
Work distribution	0.39	0.23	0.098†	0.58	0.20	0.005**
Ideas generated	0.09	0.41	0.817	-0.75	0.55	0.175
Color block decision	-0.60	0.37	0.091†	-0.18	0.08	0.020*

Table 5.2: Summary of statistical tests in Experiment 1 & 2.

5.3.1 Manipulation check and perceived performance difference

I first confirmed the success of experimental manipulation, as participants' ratings for the Smart Bot's performance were significantly higher than those for the Dumb Bot in both experiments (Experiment 1: $F = 8.75, p = .004$; Experiment 2: $F = 14.15$, $p < .001$). In terms of relative performance (i.e., the difference between the agent and each participant's own performance), participants consistently rated their own performance as significantly better than the Dumb Bot and significantly worse than the Smart Bot (Experiment 1: $F = 10.49, p = .002$; Experiment 2: $F = 24.38$, $p < .001$).

5.3.2 Subjective team contribution

Though participants reported a better team experience working with the Smart Bot than with the Dumb Bot (Experiment 1: $F = 7.94, p < .001$; Experiment 2: $F = 4.50$, $p = .035$), they reported that they made fewer contributions when teaming up with the Smart Bot than with the Dumb Bot (Experiment 1: $F = 6.35$, $p = .013$; Experiment 2: $F = 3.73$, $p = .055$).

5.3.3 Actual team contribution

Participants' actual team behavior during the team brainstorming task aligned with their self-report work distribution – they came up with more ideas when they spoke with the Dumb Bot than with the Smart Bot (Experiment 1: $F = 6.85$, $p = .010$; Experiment 2: $F = 48.00$, $p < .001$).[3]

5.3.4 Reliance and compliance with teamwork agent

In the subsequent Color Block task, most decided either to persist with the black block (59.80%) or to follow the bot's suggestion (i.e., switching to the white block; 27.94% of total participants). Whether they chose the White Block or not, 40.20% of participants shifted some degree away from their original choice (the black block). In particular, participants were significantly more likely to stick with their original choice of the black block after interacting with the Smart Bot,

[3]Though ideas presented by the Smart Bot was more elaborate, participants did *not* spend more time reading the smart ideas nor had less time to generate ideas, as I checked the chat log and there was no significant difference in the number of iterations between the two conditions.

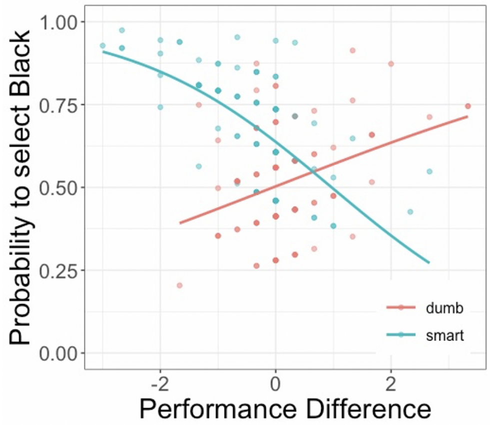

Figure 5.5: The relationship between perceived performance difference and likelihood to choose the black block.

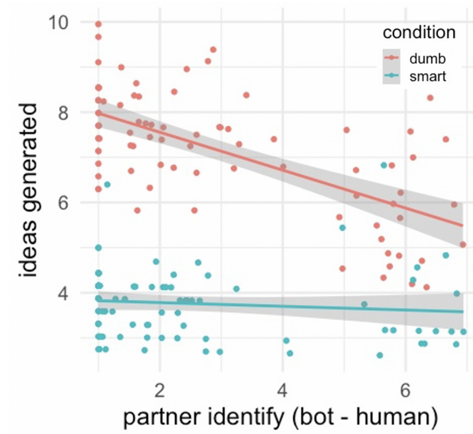

Figure 5.6: The relationship between perceived identity of partner and ideas generated by participants.

while they were more likely to accept the Dumb Bot's suggestion (Experiment 1: $F = 7.19$, $p = .008$; Experiment 2: $F = 7.19$, $p = .008$).

5.3.5 Moderating effect of social comparison

Participants' perceived performance difference between themselves and the team partner could moderate their decision to comply with the teamwork agent in the subsequent Color Block task (Experiment 1: $\beta = -.60$, $S.E. = .37$, $p = .090$; Experiment 2: $\beta = -.86$, $S.E. = .38$, $p = .023$). Among those interacting with the Dumb Bot, participants who perceived themselves as performing *better* than the bot (i.e., the performance difference is positive) were more likely to persist with choosing the black block. Conversely, among those interacting with the Smart Bot, those who perceived themselves as performing *worse* than the Smart Bot (i.e., the performance difference is negative) were more likely to ignore its suggestions.

5.3.6 Effect of perceived human-ness vs. bot-ness

Participants tended to perceive the Dumb Bot as providing more of a "human touch" (M = 3.26, $S.D.$ = 2.09) than the Smart Bot (M = 2.86, $S.D.$ = 2.03), though the difference was not statistically significant (F = 1.77, p = 0.18). Nonetheless, I found that the perceived identity of the agent (bot: lower value; human: higher value) also significantly predicted the number of ideas contributed by participants in the brainstorming session (Experiment 2: $\beta = -.42$, $S.E.$ = .18, p = .018). Specifically, those who firmly perceived the Dumb Bot *as* a bot were more likely to carry on with the teamwork rather than free-riding their teammate. On the other hand, for those who worked with the Smart Bot, the number of ideas generated by participants did not shift much, regardless of the perceived bot-ness or human-ness of their agent teammate. As a result, across the two conditions, those who believed their partner was a bot generated more ideas (M = 5.79, $S.D.$ = .77) than those who believed their partner was a human (M = 1.63, $S.D.$ = .72).

5.3.7 Evaluating human-agent teamwork outcomes

Collecting all the chat transcripts from Experiment 2, I recruited N = 135 participants on MTurk (who have not participated in any of the previous parts of the study) to rate the brainstorming outcomes of each human-agent team. Each participant was randomly presented with the full dialogues from eight different teams. In each transcript, I clearly marked the contributors as a chatbot or an anonymous human for each part of the conversation. Participants were asked to read through each dialogue and rate the performance of the human.

Again, they used the same three scales (creativity, practicality, and informativeness) to evaluate the quality of brainstorming outcomes. In the results, I did not see a significant difference in performance ratings across conditions. Namely, participants were not rated as generating ideas with better quality when they interacted with either the Smart Bot or the Dumb Bot.

5.4 Discussion

Through a four-part study, consisting of a pre-survey, a Wizard-of-Oz prototype, an experiment implemented fully through chatbots, and a post-survey, I examined the effect of social comparison in human-agent teamwork when users were working with intelligent machines that are ”better off” (the Smart Bot) or worse than (the Dumb Bot) their own performance. I saw that users reflected more positive self-performance and generated more ideas when working with the Dumb Bot; in a subsequent co-decision-making task, they were also more likely to adopt suggestions from their bot partner. I discuss the theoretical and design implications below.

5.4.1 Designing teamwork agents with comparable performance

The present study suggests that an agent does not need to be highly intelligent or technologically advanced to benefit a human-agent team. Not only did participants generate more ideas when they worked with the Dumb Bot, but the quality of their ideas was no worse than those generated by participants working with the Smart Bot. Similarly, in previous human-AI co-decision-making studies, researchers have investigated how users work with agents when the humans and agents are equally competent at solving a given task [20, 18, 19]. These studies found that agents providing additional information did not enhance accuracy in decision-making. In this scenario (as in the current research) – team dynamics can differ when human-agent teamwork is not situated in high-risk, high-stake, and highly knowledgeable domains (e.g., healthcare, military). Rather than simply offering intellectual and informative resources, agents can enhance team performance by carrying on the flow of team interaction and even boosting the self-efficacy of human participants. In the creative sector, these aspects (e.g., social exchange, flow experience, and self-perception) are particularly critical to innovation [274]. Informed by the present research, I lay out three design principles to consider when designing future teamwork agents with comparable competence:

1. Build team players, not individual contributors. When humans and agents possess similar competence, I suggest the design goal of teamwork agents should shift from offering a rich knowledge base to facilitating team synergy. As I see in the case of working with the Smart Bot, offering more informative, high-quality ideas did not improve human productivity or innovation but instead encouraged social loafing. To avoid such outcomes, I encourage designers and developers to consider how human-agent teamwork can accomplish more than either human-only or agent-only teams.

2. Leverage the "non-human" quality of autonomous agents. Introducing an agent to a team provides different opportunities from introducing yet another human team member. In the present study, the perceived bot-ness of the agent predicts positive teamwork outcomes. In fact, the less personified design led to more fruitful ideation outcomes. I encourage future research to explore other forms of non-humanoid design and their usefulness in human-agent teamwork.

3. Keep user engagement in the loop. In the case of working with the Smart Bot, I saw undesirable teamwork outcomes when participants felt detached from the teamwork, raising the importance of keeping users engaged in team interaction. However, the present experiment also showed this could hardly be achieved through dialogues alone. During the team conversation, though the bot would keep asking whether participants had more ideas to share, such requests for contribution did not effectively prevent social loafing. Regarding this outstanding challenge, I encourage future work to experiment with different forms of content design, such as testing whether participants would be more engaged if the bot offers ideas that associate with participants' previous ones instead of proposing new, random ideas.

4. Assess user experience beyond one-shot interaction. While testing the robustness of empirical findings through various tasks has become a popular practice in HCI research, multi-sectional or even longitudinal studies using human subjects remain scarce, particularly in the research area of human-bot teamwork [279]. Nonetheless, when it comes to applying computer-mediated agents in real-world teamwork settings, interactions between users and agents would certainly go beyond a single occurrence. In the present research, I identify instances where perceived bot competence can backfire in a subsequent task, illustrating the critical need to consider team relationships in the long term.

Abundant recent work in human-AI interaction has examined how to build trust in an AI entity [20, 252, 341, 283]. Common approaches include framing and offering explanations for an AI's behaviors [153, 210, 207]. In scenarios where humans rely on information offered by agents, highlighting these heuristics is critical for users to evaluate an agent's reliability in co-decision-making. However, in the generative task used in the present study, even though the Smart Bot, which presented more informative content in their responses, was indeed perceived as more trustworthy, this did not improve team brainstorming outcomes. Furthermore, the Smart Bot's competence in the first task did not translate to participants' greater reliance on them in the subsequent task. In light of this result, I again advocate the need to examine the effectiveness of trust-building through more than a single task or session since users' reliance on agents' competence can be constantly changing and highly context-dependent.

Beyond the challenges of designing agents, the present research leads me to consider yet another level of complexity —the perception of an agent (bot-like or human-like) —in human-agent interaction. On top of the variance in agents' responses and users' individual differences, I saw that the perceived bot-ness and human-ness of agents also played a key role in their team performance. In the present research, I did not identify any single variable that can solely predict how participants will perceive the identity of an agent. Even within each condition, I saw a wide variance of ratings for the perceived bot-ness/humanness. Therefore, future research can continue to explore contributing factors to explain an agent's perceived identity. Besides, while the teamwork literature traditionally presents in-group identities as serving critical functions in team success, here, I saw more positive outcomes when users viewed the bot as more distinct (i.e., less human-like). This echoes recent calls to re-think human-machine partnership [87, 68] —human-agent collaboration does not necessarily need to simulate human-human teamwork. Emphasizing the non-human aspects of agents can also benefit teams.

5.4.2 Limitations and future work

Below, I list some limitations to this series of studies. To begin with, I used the Color Block task as a simple, quantifiable measure to examine participants' behaviors after an initial teamwork interaction. However, how humans interact with AI in co-decision-making tasks can be highly context-dependent [335]. For instance, in tasks involving complex computation and a large amount of information, humans may benefit more from cooperating with agents. Thus, future research should not only apply different collaborative topics but also investigate the effect of explainability [283]. For example, in the present study, I saw participants were less willing to compromise with the Smart Bot, leading me to wonder whether the agent further explaining their suggestions could mitigate users' unwillingness to cooperate.

To begin with, I acknowledge that the present recruitment platform, MTurk, is not representative of all potential users of AI, nor does it equally represent populations of diverse ethnicity, cultural background, sexuality, and more [63]. Moreover, in the context of the present study, MTurk limits the chance to investigate the effect of expertise [288]. Previous research has found amateurs were more likely to be affected by the framing and content design of information provided by AI [153, 210], while experts seek more information and explanations from AI-mediated content [20, 52]. In this regard, will domain experts be more likely to accept and cooperate with the Smart Bot? Or will they be more likely to completely dismiss suggestions from the agent? Future research that recruits experts to generate new pools of ideas and ground truth labels will also allow us to learn more about the role of data as a design material for future AI.

Due to the worldwide pandemic, I was not able to conduct any in-person lab experiments. This means that besides having limited control over participants' attention spans during online experiments, I was not able to compare how participants interacted with a smarter chatbot to a physical robot. In particular, while I see participants' perceived "bot-ness" of their team partners had an impact on their experiences and behaviors, I do not yet know how a more embodied, anthropomorphic form would manipulate participants' perceptions of agents' human-nesss.

Furthermore, while the present study is built heavily on the Theory of Social Comparison [116], more granular psychological experiences of social comparison are difficult to capture in the present empirical settings. For instance, participants typically wouldn't report themselves as envious of the Smart Bot, though they were placed in a condition that triggered upward comparisons. In the present study, I used participants' ratings of their own performance in relation to those of the agents as a proxy to assess how they compared themselves to their teammates. However, without other forms of measurement for validation, it remains an open question whether there could be other confounding factors – besides the psychology of social comparison – that drove participants' behaviors and decisions during the study. Therefore, I encourage future research to explore other behavioral and/or physiological measures to validate these hypotheses and findings. Beyond the present study, this also calls for the need to explore new measures besides self-report scales and participants' task performance for future human-agent teamwork research.

Finally, previous research investigating autonomous agents in human-human teams has found humans were less likely to blame each other for team failures when they could ”blame” the bot [169, 298]. Similarly, future studies could examine how Smart and Dumb Bots affect human-human interaction. Will multiple human teammates free-ride together when they realize the Smart Bot can produce high-quality work? Or, will humans ally to ”defeat” the agent, bringing up the whole team’s productivity? Though I have not accounted for the effect of ”other humans” in the present study, I look forward to answers to these intriguing questions in future research.

CHAPTER 6

SUPPORTING MARGINALIZED INDIVIDUALS IN HUMAN-AI COLLABORATION THROUGH SOCIAL PRESENCE

In the previous chapter (Chapter 4), I explored and found evidence for the potential of adopting autonomous agents to support individuals' work, forming human-agent teams that could productively yield fruitful and creative outcomes. However, in our modern work environment, the majority of work requires collaboration with others. Considering this reality, I thus further examine whether human-AI collaboration can also effectively support teams of multiple individuals. In particular, I focus on whether the presence of teamwork agent could benefit marginalized individuals who, according to abundant prior literature [262, 127, 17], tend to lack support and more often struggle in group settings.

In this chapter, I conducted a study placing participants in either a majority group (i.e., working on a team where there are teammates with shared identities) or a marginalized setting (i.e., working on a team where a participant possesses a unique identity). In each scenario, I examined how the presence of a teamwork agent – either with a shared or different identity – may affect individuals' experience and behaviors during teamwork as well as the overall performance of the entire team. To operate these ideas in an experiment, I randomly assigned female and male participants to either a female-dominant or male-dominant group and randomized the presence of either a female-sounding or male-sounding voice agent to teamwork sessions.

It is worth noting while marginalization in group settings can occur based on various personal identities (e.g., gender, ethnicity, nationality, or age), I focus on gender-based marginalization in the present study for two major reasons. First, the issue remains common in modern workplaces and requires timely remedies. Second, as an initial attempt to explore the use of autonomous agents for supporting female and male marginalized members, I intend to reduce the complexity of the agent's design to minimize confounds. As female- and male-sounding voices of agents are one of the simplest ways to attach identities to autonomous agents without the need for further anthropomorphism [239], this also serves as a prominent reason why I chose to focus on gender-based marginalization.

Results of the study showed significant gender differences in participants' self-report and behavioral responses. In particular, female participants showed more positive perceptions toward a same-gender agent. They also performed better (in terms of coming up with more ideas in a team brainstorming task) with the presence of a female agent. However, while male participants did not show a preference for a team agent's gender, they performed worse when they were the only male member on a team and when a male agent facilitated the teamwork session.

In the following sections, I first review related work for this research topic. I began with synthesizing existing literature on the causes and consequences of marginalization and consolidating prior HCI work that attempted to support marginalized members through technological interventions. Based on existing work, I formulated hypotheses and research questions for the present study and investigated them through an empirical study. Besides reporting and interpreting findings from the present research, I also discussed relevant theoretical and practical implications at the end of this chapter.

6.1 Background and Related Work

6.1.1 Marginalization in group settings

Interacting with those alike is a universal propensity [147, 134, 61]. As discussed in Chapter 2, humans are inclined to categorize others into various social groups, as this cognitive process could serve functional values in social settings and effectively guide how one navigates in their social environment [310]. Furthermore, social categorization also results in some being viewed as more similar to ourselves, while others were more distinct from ourselves [163, 17**?** , 299]. Accordingly, the notion of ingroup and outgroup arises wherein individuals tend to favor ingroup members more than outgroup others. This also explains why individuals are more likely to interact with and befriend those who share common identities [191, 96]. Abundant research has observed such tendencies in a wide variety of naturalistic settings, such as in classrooms [262], in dorms [127], and in offices [17], individuals tend to interact more with those seen as ingroup members. Similarly, based on the Theory of Homophily (i.e., birds of a feather flock together [221]), plentiful studies applying social network analyses have found employees are more likely to work, interact, and collaborate within and across organizations with those deemed as ingroup members.

However, these phenomena also imply individuals identifying as minority group members often face more significant challenges in forming and maintaining bonds with others since they are often left out in interpersonal settings [11, 108, 227]. For instance, research studying collaboration in academia has revealed minority individuals often faced more obstacles to finding collaborators and forming research teams [53, 192]. Along the same vein, longitudinal research examining gender disparity in the STEM fields has found female students not being able to obtain the same level of social resources has turned into the primary obstacle, which eventually discourages a large portion of female students from pursuing careers in STEM [141, 296, 336].

Likewise, research on organizational behavior has also revealed the threat of marginalization commonly faced by minority group members [11, 108, 227]. Moreover, behavioral differences in how one interacts with ingroup versus outgroup members can be captured through various means of communication at work [57, 94, 332, 339]. For instance, in a study utilizing corporate email records, Milkman and colleagues found that individuals were less likely to respond or took longer to respond to emails from colleagues of gender or racial outgroups [223]. Numerous studies about interpersonal communication have also suggested that individuals can demonstrate distinct – and sometimes discriminative – behaviors toward outgroup, minority individuals through both verbal and non-verbal cues [60, 137, 271]. Large-scale analyses of text-based communication on online forums have found that, just by viewing others' account names on the platforms, one can make judgments of their identities [159]. As a result, users tend to entail greater positivity in their languages when interacting with those perceived as the same social group. In contrast, interactions with outgroup users adopt more neutral, distant, and even hostile tones. In offline settings, scholars had found that individuals also showed subtle but discriminative behaviors when they sensed distinct identities from their interactants [102]. Specifically, through analyzing the speeches of political candidates, the authors found that politicians were more likely to adopt less competent word choices when delivering speeches to primarily minority audiences. This tendency was also observed during interracial interactions in group settings.

While such distinct and even discriminative behaviors are omnipresent in a wide range of naturalistic settings, research in organizational behavior suggests they may not necessarily be carried on to interactions with team members at work – especially in teams of smaller scales. This is because team interaction at workplaces embodies a unique social scenario such that individuals are more motivated to suppress their personal opinions and behave professionally as they are concerned about social evaluations from their colleagues [26, 363]. In this regard, one may intend to avoid any signs of discrimination and demonstrate an interest in working with all individuals in the same team. However, the extent to which a group tolerates discriminative behaviors also serves as a critical guide for individuals' behaviors [270]. For instance, if the entire team is ignorant of inclusiveness and disrespects minority team members, such signals may further exacerbate discrimination and marginalization in small groups.

6.1.2 Technological support for marginalized individuals

Existing work in the HCI field has dedicated much effort to marginalized individuals; a large portion of the line of research focused on addressing diversity challenges in educational programs – particularly computer science, engineering, and related subjects – as these fields have long faced the criticisms of gender and racial imbalance. Several ground-up, long-term approaches aim to enhance diverse individuals' interests and motivations by exposing them to these technical fields from a young age (e.g., hosting high school engineering summer camps or workshops) or introducing gamified learning methods [218, 75, 119]. Furthermore, [34] incorporated a perspective-taking strategy into course curricula, which was found to enhance students' empathy and intent to support marginalized individuals.

Besides, a number of studies proposed new techniques that introduce changes in team formation and group dynamics in educational and professional settings, such as changing the gender ratio of small teams in a workplace [131, 78]. Attempts have also been made to design networking tools to connect individual workers or students, supporting marginalized individuals who may otherwise feel isolated in a work or learning environment [337, 158]. For instance, [158] developed a networking tool to help connect individuals with disabilities with mentors who can provide personalized guidance [158].

As in many other fields, AI has been applied to design some of these supportive tools, though non-AI tools still dominate most of the work in this area. In more recent work, [337] has designed an AI assistant that helps recommend users whom to reach out to, mitigating isolation in online learning programs. Furthermore, the use of AI techniques has been adopted in a specific line of proposed solutions that have less been of interest prior to the recent AI hype. These studies (e.g., [201, 56, 139]) leveraged AI, specifically natural language processing (NLP), to detect toxicity and discrimination in online environments, protecting most users from being harassed by these problematic behaviors. However, it is worth noting that despite the strong potential to utilize AI for personalized/customized design, these existing solutions are not targeted to provide particular help to specific individuals.

However, direct interventions in small-group interactions remain relatively scarce. Moreover, interventions targeting to support marginalized individuals in team conversations could risk putting them on the spot, causing them to be more reluctant to adopt such remedies [99, 129, 130]. Alternatively, studies have attempted to make the majority group more aware of the issues of marginalization, such as performing real-time linguistic analyses and visualizing the proportion of each member's speech in a team conversation to make the entire team aware of whether certain teammates could be dominating the conversation while some could be encouraged to speak up [199, 200].

More recently, scholars have explored the potential of having intelligent agents participate in team conversations to support marginalized members. For instance, [205] adopted a conversational agent to open up the speaking floor and interrupt certain teammates when they have taken their turn to speak a couple of times in a row. [280] experimented with having agents in teams to create "pseudo presence." For example, if there is only one female on a team, including a few female agents during teamwork could potentially deemphasize the solo status of the female human teammate. Nonetheless, these more generic approaches have limited considerations for the different experiences and need that marginalized members may possess in group settings. As the present research examines gender-based marginalization in teamwork, I further discuss how female and male members respond differently to marginalization in teams in the following section.

6.1.3 Gender difference in response to marginalization

Abundant work in psychology, communication, organizational behavior, management science, and more recently, HCI, has investigated gender differences in teamwork [350, 222, 115, 29, 109, 103]. Much of the literature was motivated to understand the importance and possible benefits of gender diversity in teams, as gender imbalance remains a common phenomenon in a large number of fields. It is worth noting that, though gender is a socially constructed concept (namely, one often forms their gender identity through interacting with others in their social world) and goes beyond the female-male dichotomy [14], the majority of research in this topic area focuses on differences between self-identified males and females. Therefore, the following literature review and the study design follow the same convention. Specifically, existing literature has studied gender differences in teamwork through two closely related lenses: first, plentiful discussions emerge from gender differences at the individual level [14, 9, 208, 159], which then drive female and male teammates' distinct behaviors in collaboration; second, researchers have further explored how group composition (i.e., the portion of female vs. male teammates in a team) amplify or suppress such individual differences. In the following, I further unpack these two lines of work respectively.

Prior research draws heavily on the Theory of Social Identity [309] to understand gender difference at the individual level, suggesting that gender-based behavioral differences are developed through the expectations, role models, and even stereotypes one gradually "learns" through interacting with others in their society. Therefore, gender differences become particularly salient in an environment where gender-based role assignments are clearly defined – as individuals tend to develop and acquire relevant skills and characteristics that support them to fit well with these expected roles [9]. As a result, research has consistently found women exhibiting a significantly higher level of social sensitivity, based on their greater ability to read nonverbal cues and make accurate inferences about what others are feeling or thinking [353]. This trait also transforms their behaviors in teams, leading female teammates to be significantly more interpersonally oriented than men in group settings [103]. Accordingly, women pay more attention to ensuring equality and inclusiveness in teams, resulting in more even turn-taking when there are more female members on a team [31, 217, 353].

By contrast, men are more likely to adopt an autocratic style, which can be demonstrated through behaviors such as giving orders, interrupting others, and applying more assertive languages [103, 249]. During teamwork, men also tend to be more task-oriented, prioritizing whether a team has achieved its goal over the experiences of individual teammates [217, 31]. Research in management science has found men exhibited a stronger tendency to claim leadership in teams – as males expect a clearer role specialization in a team structure when no one plays the leadership role, they are more likely to jump in and provide directions to the team [234]. Meanwhile, compared to females, male teammates applied a considerably distinct set of verbal and non-verbal cues in team conversations; for instance, men display more direct and dominant behaviors when speaking in teams, such as chin thrusts, gesturing, and direct eye contact, whereas women engage in more smiling whether they are speaking or listening [95].

These behavioral differences at the individual level also cause females and males to react differently – and receive different reactions – when they encounter gender-imbalanced teams, while such gender-based traits become particularly salient when one experiences a solo status (i.e., a person is the only male or the only female on a team) [172, 173]. Female solos tend to experience greater visibility and scrutiny of their work, as well as confinement to tasks that are stereotypically feminine [38, 172, 173, 356, 357, 358], whereas male solos tend to be evaluated more positively than female solos for task performance [77, 113, 246, 272]. Empirical research has built on this prior literature and found women in male-dominant groups tend to speak less and be less assertive than men; on the contrary, male teammates in female-dominant groups tend to speak more and become even more dominant and task-oriented [234, 28]. Once again, these differences between females' and males' responses are magnified when they are the sole group members [234, 28].

Synthesizing the above literature, I hypothesize that female and male participants would respond differently when they are the sole marginalized members of a team. However, it remains unknown whether having a same-gender or different-gender agent on the team could provide support for marginalized individuals. Namely, I examine the following inquiries in the present research:

RQ1a: How will having a ***same-gender*** *agent on the team influence* ***female*** *participants' team experiences and performance when they are marginalized in teamwork?*

RQ1b: How will having a ***same-gender*** *agent on the team influence* ***male*** *participants' team experiences and performance when they are marginalized in teamwork?*

RQ2a: How will having a ***different-gender*** *agent on the team influence* ***female*** *participants' team experiences and performance when they are marginalized in teamwork?*

RQ2b: How will having a ***different-gender*** *agent on the team influence* ***male*** *participants' team experiences and performance when they are marginalized in teamwork?*

6.2 Method

6.2.1 Participants

In the present study, I recruited participants through Prolific (N = 178) and randomly assigned them to gender-imbalanced teams (either a female-dominant or a male-dominant team) based on their self-reported gender identities. Each teamwork session includes three participants; a female-dominant team consists of two females and one male, while a male-dominant team has two males and one female. I ran 20 groups of participants as a pilot to ensure the validity of the study protocol. As the pilot already showed significant patterns with the behavioral data, instead of performing a power analysis, I set the target sample size as 180 such that each condition at least includes 30 participants (i.e., having 30 marginalized females, 30 marginalized males, 60 non-marginalized females, and 60 non-marginalized males). According to the Central Limit Theorem in the statistics literature, it requires at least n = 30 for a sample to reach normal distribution, which is the statistical assumption for various tests performed in the current data analyses [308]. In the end, data from 2 participants were removed since they did not complete the post-teamwork survey, resulting in N = 178 as the final sample size. Participants' demographic data is recorded and further elaborated in the Measurement section.

I again recruited and conducted the study entirely online in hope to get access to participants with real-world work experiences. This consideration became even more important in the present study as one of the teamwork tasks requires participants to evaluate whether several employees should be offered job promotions given their work profiles. In terms of recruitment platform, I shifted from MTurk to Prolific because recent studies have constantly found participants on Prolific providing better and more consistent data quality than those on MTurk (See a review at [248]).

6.2.2 Procedure

Upon signing up to take part in the study, participants first filled out a short pre-survey to self-report their past experiences of teamwork and marginalization, as well as their self-identified demographic profiles and familiarity with AI technologies. Participants were then randomly assigned to one of the two types of gender-imbalanced teams and attended a teamwork session based on their availability. All teamwork sessions were held remotely on Zoom. In each of the teamwork sessions, participants were asked to work on two teamworking tasks in a randomized order (an idea generation task [274] and a decision-making task; see more details about the tasks in the next section on ”Teamwork Tasks”). During each task, there was an AI agent that participated in teamwork. I also applied random assignment for the gender identity of the agent, such that in each session, participants worked with a female agent in one task and a male agent in the other task. The order in which the female agent and the male agent were presented randomly. This experimental design resulted in four possible task procedures:

1. In the <u>first</u> task, participants worked with a **female** agent on an idea generation task. In the <u>second</u> task, participants worked with a **male** agent on a decision-making task.
2. In the <u>first</u> task, participants worked with a **male** agent on an idea genera-

tion task. In the <u>second</u> task, participants worked with a **female** agent on a decision-making task.

3. In the <u>first</u> task, participants worked with a **female** agent on a decision-making task. In the <u>second</u> task, participants worked with a **male** agent on an idea generation task.
4. In the <u>first</u> task, participants worked with a **male** agent on a decision-making task. In the <u>second</u> task, participants worked with a **female** agent on an idea generation task.

Immediately after completing each task, participants were asked to fill out a short survey to reflect on their teamwork experiences just then and their perceptions toward their teammates. After they completed both tasks and filled out both surveys, participants completed a short exit survey to share their overall experiences and final thoughts. The entire study took around 45 ~ 60 minutes to complete, and participants received $15 cash compensation for their time and work. The study protocol was reviewed and approved by the the Institutional Review Board (IRB).

6.2.3 Teamwork tasks

In the study, participants worked on two teamwork tasks in a randomized order. One task was an idea generation task where participants were asked to come up with ideas for water and/or energy conservation with their teammates [274]. Participants had 10 minutes to work on this task. The task has been used in several prior teamwork studies (e.g., [301]), as the topic requires ideas to be not only new but also practical, sufficing the definition of creativity as ”novel and functional” [294].

The other task was a decision-making task where participants viewed the profiles of five employees and determined who to offer a promotion to. The task and these employee profiles were adopted from Binns et al.'s study [33], which is oft-cited and used in prior research studying human-AI co-decision-making as well [194]. To ensure participants indeed came to a decision for each employee profile, the AI agent explicitly asked participants to vote on whether they determined to promote each candidate or not before moving on to the next one. Participants had 15 minutes to work on this task. The lengths of both teamwork tasks were determined by conducting pilot studies with other research assistants.

During both tasks, participants saw a screen presenting the teamwork task content (see Figure 6.1 for screen capture). They interacted with their human teammates and the AI agent merely through voice interaction. The AI agent was not embodied in a physical form or humanoid figure. The only visual cue of the AI agent was an animated, abstract shape shown on the screen. I deliberately chose to minimize the tangible form and visual representation of the AI in order to eliminate the confounding factor of embodiment and anthropomorphism [239].

6.2.4 AI agent in teamwork

The AI agent in the present study was implemented through a Wizard-of-Oz protocol, such that the agent's responses were pre-scripted and pre-recorded.

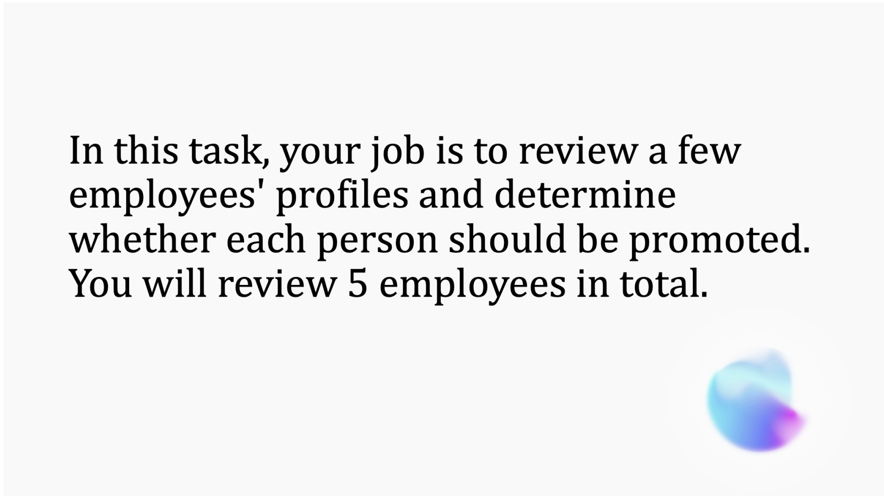

Figure 6.1: A screen capture of the interface used for the present study.

The researcher (myself) played the agent in each teamwork session by playing the pre-scripted content, which was all generated by OpenAI's ChatGPT [1]. The decision to adopt a Wizard-of-Oz paradigm instead of using an existing voice assistant product (e.g., Apple Siri, Amazon Alexa, or Google Assistant) was made deliberately so that the agent's speech content in each team conversation was identical. When participants' responses could not be covered using the pre-recorded content, the agent would respond with "I'm sorry, but I don't understand what you just said."

The key manipulation in the present study is the gender identity of the AI agent. I used the agent's voice (i.e., either a female-sounding or male-sounding voice) to implement such manipulation. After obtaining the speech content from ChatGPT, I then created the female- and male-sounding responses through Google Cloud's text-to-speech function. It is worth noting that the sound quality of the voice agent is not the key focus of this study. Instead, what matters is the gender identity that participants associated the agent with. Existing

[1] https://chat.openai.com/

work has found that users would either categorize an agent's voice as female-sounding or male-sounding [253]. Therefore, I did not include a "gender neutral" condition because adopting a gender-ambiguous voice for the agent may simply demand additional cognitive loads of the participants. Still, they are likely to identify the agent as female or male.

6.2.5 Measurement

All teamwork sessions were video recorded through Zoom and, later on, transcribed through Otter.ai[2]. These materials served as the main data source to analyze participants' behaviors during the teamwork sessions. Additionally, I captured a wide range of self-report responses from participants through the pre- and post-task surveys. Specifically, measurement in the present study can be categorized into the following four themes.

Behavioral responses.

Transcripts of participants' conversations during teamwork were labeled with their subject IDs and timecode. I first examined the behavioral data by looking at the *portion of speech* of each participant in their team (M = 0.27, $S.D.$ = 0.16). Namely, this measure captures how much time each participant talked during a team conversation divided by the total amount of time all members of the same team spoke in the conversation. For the idea generation task, I further coded *the number of ideas* generated by each participant (M = 3.10, $S.D.$ = 2.28). The coding result was verified by a second coder going through the transcripts to code the

[2]https://otter.ai/

number of ideas generated as well. The inter-rater reliability reaches 89.99%. For any discrepancy, I took the mean value between the two coding results.

Teamwork outcomes.

In the brainstorming task, participants' team performance was captured by the number of ideas generated. In each brainstorming session, each team produced 10.11 ideas on average ($S.D. = 4.86$). In the decision-making task, participants were asked to vote on whether they would promote an employee or not after reviewing each profile. On average, participants were able to reach a consensus among 84.84% of all decisions discussed during the teamwork conversations (i.e., when all teammates agreed to promote or not to promote an employee). As these behavioral data were not measured on fixed scales (e.g., unlike self-report measures using 7-point Likert scales), I used their standardized values for data analyses to prevent results from being biased by extreme values.

Self-report responses.

I capture a set of self-report measures from participants in both the pre- and post-study questionnaires. All self-report items were measured on a 7-point Likert scale unless specified otherwise (1 = strongly disagree; 7 = strongly agree).

To understand how individuals' prior experience of marginalization may influence their current team behaviors and the effect of having an agent teammate, I asked participants to report their existing experience of *marginalization* in the pre-survey through three scales from existing literature [62] ($M = 2.85$,

$S.D. = 1.16$). These scales capture key dimensions of marginalized experience, including whether individuals have felt a sense of ignorance in group settings (e.g., feeling "invisible" or "unheard") and whether their peers have demonstrated micro-aggressive behaviors against their identities.

During the study, participants filled out two post-surveys, each taking place immediately after finishing each teamwork task. The post-survey captured whether participants experienced *marginalization* during teamwork using the same items as in the pre-study questionnaire ($M = 2.17$, $S.D. = 1.99$). They reflected on whether they found themselves being marginalized by their human teammates ($M = 2.47$, $S.D. = 2.03$) or by their AI teammates respectively ($M = 2.64$, $S.D. = 2.05$). Furthermore, participants reported their overall *team experience* ($M = 6.07$, $S.D. = 0.79$), with higher values indicating more positive team experience [41]. Besides, participants reported their *perception toward human teammates* ($M = 6.20$, $S.D. = 0.78$) and *perception toward AI teammates* respectively ($M = 4.75$, $S.D. = 1.49$) wherein higher values indicate a more positive perception of their teammates [205]. Participants also specified their they perceived *team support* either from their human teammates ($M = 5.13$, $S.D. = 0.68$) or from their AI teammates ($M = 4.40$, $S.D. = 0.95$) [74]. Finally, I adopted six items from the NASA task load scale [143] to measure participants' perceived workload and difficulties of the teamwork tasks ($M = 2.67$, $S.D. = 0.98$).

Demographic data.

In the pre-survey, participants were asked to self-report their basic demographic profiles, including age, gender, ethnicity, years of work experience, and fluency in the English language (since the entire study, including the team interaction and the questionnaires, was conducted in English). Based on participants' self-report profiles, there are 90 females and 88 males in the current sample. Their average age is 34.90 ($S.D.$ = 11.39). The most identified ethnicity groups are Caucasian (40.82%), Black American (17.75%), and Asian (16.57%). On average, participants have 10.98 years of work experience ($S.D.$ = 11.17). 70.69% of participants mentioned English as their first language, while all of them were able to communicate fluently in English during the team conversations. All except for 6 participants have used a voice assistant before, and 23.60% of participants reported themselves as regular users.

Validity check. At the end of the second post-survey (i.e., the final questionnaire participants filled out after completing both teamwork tasks), I included an open-ended question for a validity check, asking participants to describe what they thought the study was about in text. This question is used to ensure participants were not fully aware of the experimental manipulation. In their responses, none of the participants mentioned team marginalization as the topic of the study. Furthermore, to check the effectiveness of the Wizard-of-Oz study protocol, I asked participants whether they believed the AI agent was indeed an AI or was played by a human. Again, all participants believed the teamwork agents were indeed AI instead of humans.

6.3 Results

6.3.1 Overview of data analysis

I performed all data analyses in R [261]. I began the analyses by examining whether participants' self-report and behavioral responses differed by (1) their gender, (2) whether they were marginalized or not, and (3) whether the team agent adopted a female or male voice. For each independent variable, I first ran a factorial ANOVA to check whether there is a significant mean difference between (1) female vs. male participants, (2) marginalized and non-marginalized individuals, and (3) same- vs. different-gender agents. For independent variables that showed significant mean differences, I then used linear models to examine the relationships between variables. Depending on whether a dependent variable is a repeated measure or not, I used either the `lm` or `lmer` function in R to perform linear modeling. To examine any effect on dependent variables that are repeated measures (i.e., self-report scales in the two post-task questionnaires), I used linear mixed models from the `lmer` package [27] to account for within-subject variances by including the `(1|subjectID)` term in the models. To examine any effect on dependent variables that are *not* repeated measures (i.e., teamwork outcomes in each of the two tasks), I used the `lm` function to run a linear regression. All linear models (including both `lm` and `lmer` models) were controlled for the task order (i.e., whether the data was collected from the first or second teamwork task in a study) to account for a potential order effect resulting from the between-within-subject design.

Then, I examined whether female and male participants responded differently in marginalized situations by testing a two-way interaction term of *partic-*

Purpose	Model	Independent Variable (IV)	Dependent Variable (DV)	Moderating Variable (MV)
To examine the main effect of experimental conditions on self-report scales (repeated measures)	lmer(DV ~ IV + task order + (1|subjectID))	1. participant's gender 2. marginalized 3. agent's gender	1. team marginalization 2. team experience 3. support from teammate 4. perception of teammate 5. task load	n.a.
To examine the main effect of experimental conditions on behavioral measures (non-repeated measures)	lm(DV ~ IV + task order)	1. participant's gender 2. marginalized 3. agent's gender	1. portion of speech 2. # of ideas generated	n.a.
To examine how female and male participants responded when being marginalized based on their self-report scales (repeated measures)	lmer(DV ~ IV + task order + (1|subjectID))	participant's gender x marginalized condition	1. team marginalization 2. team experience 3. support from teammate 4. perception of teammate 5. task load	n.a.
To examine how female and male participants responded when being marginalized based on their behavioral data (non-repeated measures)	lm(DV ~ IV + task order)	participant's gender x marginalized condition	1. portion of speech 2. # of ideas generated	n.a.
To examine whether and how having same- or different-gender agents has an effect on marginalized teammates based on their self-report scales (repeated measures)	lmer(DV ~ IV + task order + (1|subjectID))	participant's gender x support type (human, same-gender agent, different-gender agent)	1. team marginalization 2. team experience 3. support from teammate 4. perception of teammate 5. task load	n.a.
To examine whether and how having same- or different-gender agents has an effect on marginalized teammates based on their behavioral scales (non-repeated measures)	lm(DV ~ IV + task order)	participant's gender x support type (human, same-gender agent, different-gender agent)	1. portion of speech 2. # of ideas generated	n.a.
To examine whether and how existing team experience moderates marginalized teammates' self-report experiences (repeated measures)	lmer(DV ~ IV*MV + task order + (1|subjectID))	participant's gender x support type (human, same-gender agent, different-gender agent)	1. team marginalization 2. team experience 3. support from teammate 4. perception of teammate 5. task load	existing experience of marginalization
To examine whether and how existing team experience moderates marginalized teammates' behavioral responses (non-repeated measures)	lm(DV ~ IV*MV + task order)	participant's gender x support type (human, same-gender agent, different-gender agent)	1. portion of speech 2. # of ideas generated	existing experience of marginalization

Table 6.1: Summary of models tested in the present study

ipant's gender (female/male) x *marginalized teammate* (yes/no) in linear models. Likewise, I am interested in whether the presence of a same- or different-gender agent affects female and male marginalized teammates' experience and performance in teamwork. I performed this part of the analysis by coding each participant's assigned team scenario into three types. Participants who were randomly assigned as the majority group were labeled as receiving ***human support***. Participants who were randomly assigned as the marginalized members were coded as receiving either ***AI support*** or ***no support***. To be specific, *AI support* indicates when a marginalized participant worked with a same-gender agent; *no support* indicates when a marginalized individual worked with a different-gender agent.

After coding these three support types, I again tested whether the support types have any significant main effect on participants' self-report experience and behavioral responses. Furthermore, I examined whether there is a significant two-way interaction between participants' gender and the support types they received. I adopted this analytic approach instead of testing a three-way interaction of participant's gender (female/male) x marginalized teammate (yes/no) x agent's gender (female/male) as the former can help with interpreting the data more handily. For instance, with a three-way interaction test, looking at the effect of marginalized females or marginalized males working with a female agent has different syntactic meanings, making interpretations of the findings more challenging.

Finally, I examined whether participants' prior team and marginalization experiences would affect their current human-agent teamwork experiences and behaviors. I did so by including their self-report responses in the pre-survey as a moderating variable in the linear models and testing a three-way interaction effect across participants' gender, their prior marginalization experience, and the support types they received during teamwork. In Table 6.1, I summarize all models tested for the present study.

6.3.2 Main effect of marginalization

A significant main effect of marginalized conditions on participants' self-report marginalized experience verified successful experimental manipulation, such that participants who were randomly assigned as the marginalized member in a team indeed reported a significantly higher degree of marginalization post-teamwork ($\beta = 0.72$, $S.E. = 0.15$, $t = 4.96$, $p < 0.001$). Furthermore, participants reported marginally worse perceptions of their human teammates when they were the marginalized team members ($\beta = -0.22$, $S.E. = 0.13$, $t = -1.73$, $p = 0.085$).

6.3.3 Main effect of participants' gender

Regardless of their assigned conditions, participants' self-report and behavioral responses showed significant gender-based differences. To begin with, male participants reported significantly more positive perceptions toward their human teammates ($\beta = 0.26$, $S.E. = 0.12$, $t = 2.19$, $p = 0.029$) and also marginally more favorable team experience ($\beta = 0.22$, $S.E. = 0.13$, $t = 1.67$, $p = 0.097$). Meanwhile, female participants reported a marginally higher level of perceived task load during teamwork ($\beta = 0.28$, $S.E. = 0.16$, $t = 1.74$, $p = 0.083$). Demonstrated in their behavioral data, male participants talked significantly more in team conversations (statistical test results based on transformed data: $\beta = 0.07$, $S.E. = 0.01$, $t = 10.54$, $p < 0.001$; statistical test results based on raw data: $\beta = 0.47$, $S.E. = 0.04$, $t = 10.54$, $p < 0.001$). Regarding their performance in the idea generation task, male participants also contributed significantly more ideas than female teammates in the brainstorming task (statistical test results based on transformed data: $\beta = 0.07$, $S.E. = 0.04$, $t = 1.71$, $p = 0.031$; statistical test results based on raw data: $\beta = 0.02$, $S.E. = 0.01$, $t = 1.71$, $p = 0.003$).

6.3.4 Main effect of voice agent's gender

Test results of linear models did not show a significant main effect of the teamwork agent's assigned identity (i.e., presented in either a male-sounding or female-sounding voice) on any of the self-report or behavioral variables.

6.3.5 Gender difference in response to marginalization

Furthermore, I found female and male participants showed several different patterns in marginalized situations; that is, results of linear models showed a significant interaction effect between participants' gender and their assigned marginalized conditions (i.e., whether one was the marginalized member or not) on their subjective experience and behavior during teamwork. I first found a significant gender difference in participants' self-report marginalization ($\beta = -1.36$, $S.E. = 0.21$, $t = -6.59$, $p < 0.001$); as shown in Figure 6.2, female participants reported a significantly higher degree of marginalized experience when they were indeed the marginalized member in a team, whereas male participants were less likely to self-report being marginalized when they were the only male member in a team. Meanwhile, when being marginalized, male participants found the teamwork more burdening and reported a significantly higher degree of task load compared to scenarios when they were the majority group in teams; however, female participants did not reveal a similar pattern when they experienced marginalization ($\beta = 0.75$, $S.E. = 0.35$, $t = 2.17$, $p = 0.032$).

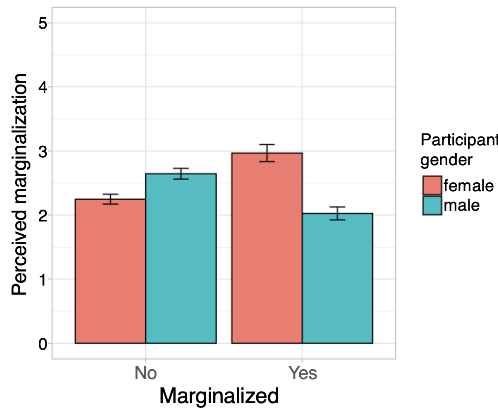

Figure 6.2: Self-report marginalized experience by participants' assigned condition and their gender.

Marginalization can also have an effect on one's perception of their teammates. When being marginalized, female and male participants as well showed different perceptions toward their AI teammates ($\beta = -2.38$, $S.E. = 1.07$, $t = -2.21$, $p = 0.036$). To be specific, female participants expressed more positive sentiment toward their AI teammates when they were the only female on the team, whereas male participants reported worse perceptions of AI teammates when they experienced marginalization in teamwork (see Figure 6.3).

Analyzing data from participants' team conversations, I found female and male participants also showed different behavioral patterns when they were marginalized. In particular, there is a significant interaction effect between participants' gender and their assigned condition (i.e., whether they were assigned to the majority or the minority group) on the portion they spoke up during team conversations (statistical test results based on transformed data: $\beta = 0.08$, $S.E. = 0.01$, $t = 5.76$, $p < 0.001$; statistical test results based on raw data: $\beta = 0.54$, $S.E. = 0.09$, $t = 5.76$, $p < 0.001$). When being marginalized, female participants

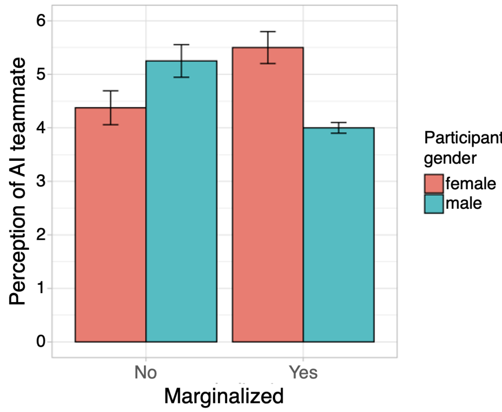

Figure 6.3: Self-report perception of AI teammates by participants' assigned condition and their gender.

tended to become less active during team discussions, but male participants instead talked more when they were the only male member of a team. However, talking more did not indicate greater team contribution or productivity. In fact, though male participants talked more when they were marginalized, they generated fewer ideas compared to their male counterparts who were assigned to the non-marginalized condition. Conversely, though marginalized female participants were relatively quiet, they were able to focus on the task at hand and generated more ideas than females who were not marginalized. As a result, I also saw a significant interaction effect between participants' gender and marginalized conditions on the count of ideas they produced (statistical test results based on transformed data: $\beta = -0.22$, $S.E. = 0.09$, $t = -2.29$, $p = 0.022$; statistical test results based on raw data: $\beta = -0.07$, $S.E. = 0.03$, $t = -2.29$, $p = 0.022$). Together, gender-based behavioral differences in marginalized versus non-marginalized conditions were illustrated in Figure 6.4.

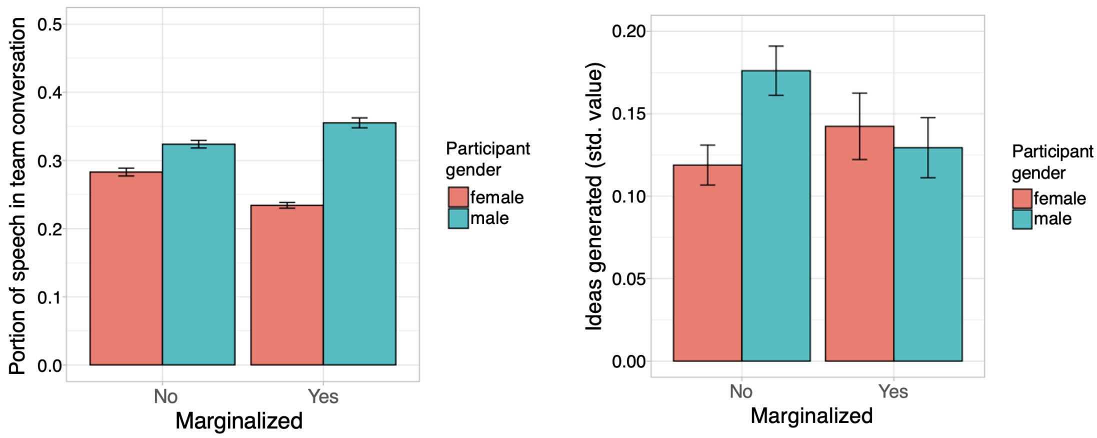

Figure 6.4: Behavioral difference in team conversations by participants' assigned condition and their gender.

6.3.6 Agent's support for marginalized teammates

Participants' responses to working with a same- or different-gender agent in marginalized conditions again showed significant gender differences. To begin with, there was a significant interaction effect between participants' gender and the types of team support they received on their perception toward the teamwork agent ($\beta = 3.06$, $S.E. = 1.44$, $t = 2.13$, $p = 0.044$). In particular, female participants reported significantly more positive perceptions of the AI teammate when they were marginalized – with the highest average ratings for same-gender agents. Conversely, male participants showed a worse perception of the AI teammate when they were marginalized, and they also did not have a preference between the female agent and the male agent. Compared to participants' perception of their human teammates, I did not observe significant gender difference when working with a same- or different-gender agent ($\beta = 0.09$, $S.E. = 0.41$, $t = 0.23$, $p = 0.818$). Gender difference in subjective experience during teamwork was also reflected in participants' perceived task load during teamwork ($\beta = -1.11$, $S.E. = 0.44$, $t = -2.49$, $p = 0.014$). As per the analysis from the previous section, while male teammates perceived a greater task load when they were marginalized, they felt more burdened when working with a different-gender agent than working with a same-gender agent. Resonating with results from the previous section, females were, by contrast, able to focus better on the task at hand when they were marginalized, resulting in a lower perceived task load.

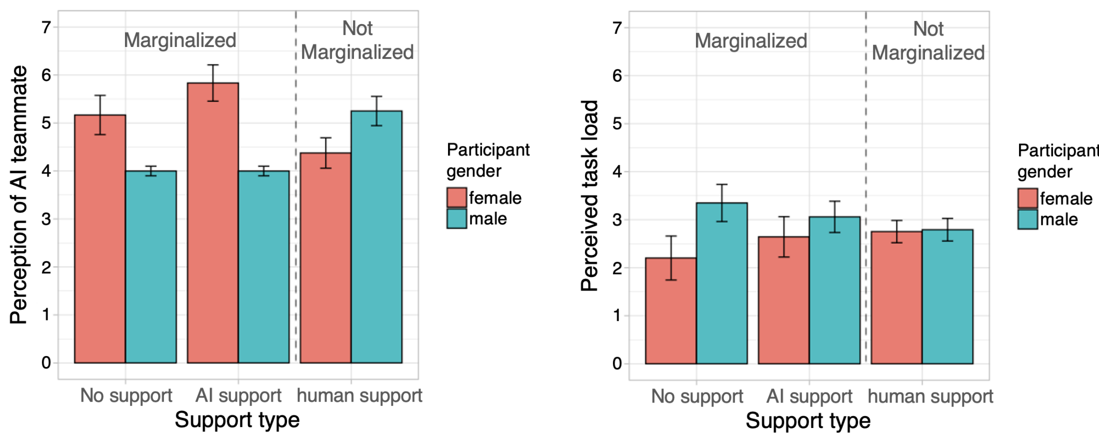

Figure 6.5: Effect of agent's support on participants' subjective experiences in teamwork.

Similarly, females and males demonstrated distinct behavioral responses to same- and different-gender agents when they were marginalized. First, there is a significant interaction effect between participants' gender and received support types on the portion of speech they participated in during team conversations (statistical test results based on transformed data: $\beta = -0.09$, $S.E. = 0.02$, $t = -5.32$, $p < 0.001$; statistical test results based on raw data: $\beta = -0.47$, $S.E. = 0.12$, $t = -3.90$, $p < 0.001$). Once again, male participants talked more when they were marginalized, and they were particularly talkative when working with a different-gender agent in a marginalized condition. Conversely, while female participants talked less in a team where they were the only female member, the difference between working with a same- or different-gender agent was less prominent. Again, how much a participant talked in team conversations did not necessarily transfer to their contributions. Upon observing a significant interaction effect between participants' gender and their support types on the number of ideas they generated during teamwork (statistical test results based on transformed data: $\beta = -0.55$, $S.E. = 0.16$, $t = -3.51$, $p < 0.001$; statistical test results based on raw data: $\beta = -0.18$, $S.E. = 0.05$, $t = -3.51$, $p < 0.001$), I saw marginalized females thrived with the support of a same-gender agent, producing even more ideas than those females working in a female-dominant group. While marginalized males did not show a preference for either a same- or different-gender agent (see results reported in the above paragraph), they were able to generate the most ideas without the presence of any same-gender entity.

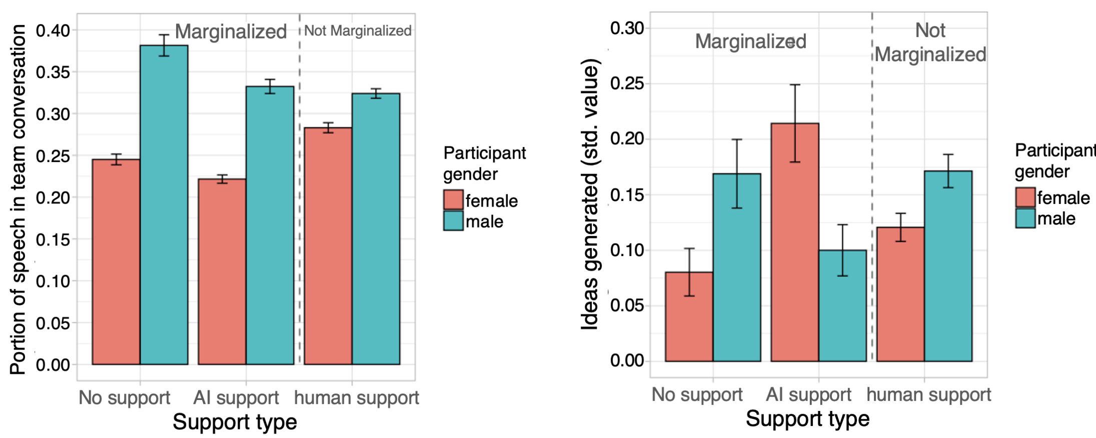

Figure 6.6: Effect of agent's support on participants' behaviors in teamwork.

6.3.7 Moderating effect of prior marginalization experience

In the following sections, I examined whether participants' prior experience in teamwork had any effect on how they performed in the present teamwork. I began with examining the effect of whether participants have experienced marginalization in the past by testing three-way interaction terms: (1) *gender x prior marginalized experience*: a two-way interaction between participants' gender and their self-report prior marginalized experience, (2) *gender x marginalized condition x prior marginalized experience*: a three-way interaction across participants' gender, whether they were assigned as the marginalized member in teamwork during the present study, and whether they have experienced marginalization in the past, and (3) *gender x support type x prior marginalized experience*: a three-way interaction across participants' gender, the type of support they received during the present study (from a human, same-gender agent, or different-gender agent), and whether they have experienced marginalization in the past.

Before examining the above test results, it is worth noting that participants' self-report prior experience of marginalization is not normally distributed (based on a Shapiro-Wilk normality test: $W = 0.96$, $p < 0.001$), and the majority of participants did not report experiencing intense marginalization in the past (see Figure 6.7). Accordingly, I used the log-transformed values of this variable for analyses. However, findings concerning those who have experienced more marginalization in the past may still be subject to greater variances.

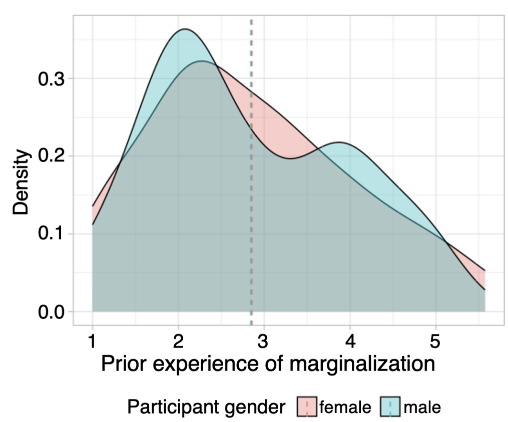

Figure 6.7: Density plot of participants' prior experience of marginalization

Results of the above-mentioned analysis first showed gender differences in the moderating effect of one's prior marginalization on team behavior. First, I observed a significant interaction effect between participants' gender and their ratings of prior marginalized experience on the portion of the speech in team conversations during the current study (statistical test results based on transformed data: $\beta = 0.05$, $S.E. = 0.02$, $t = 2.25$, $p = 0.026$; statistical test results based on raw data: $\beta = 0.05$, $S.E. = 0.17$, $t = 0.27$, $p = 0.024$). As shown in Figure 6.8, female participants who experienced more severe marginalization in the past tended to talk less during teamwork sessions in the present study, while male participants who experienced more marginalization in the past were more talkative in the study sessions. Looking at the number of ideas generated during the present teamwork experiment, participants' prior experience of marginalization also moderated their performance in teams (statistical test results based on transformed data: $\beta = 0.64$, $S.E. = 0.31$, $t = 2.03$, $p = 0.042$; statistical test results based on raw data: $\beta = 0.21$, $S.E. = 0.10$, $t = 2.03$, $p = 0.042$). As shown in the downward-trending line on the right of Figure 6.8, while, in general, participants who have experienced more severe marginalization in the past tended to contribute fewer ideas during the brainstorming sessions, this trend was more salient among female participants.

Next, I examined whether and how past marginalization experience may influence female and male participants who worked either as the majority or minority during the current teamwork study. Results showed a marginally significant three interaction across participants' gender, their marginalization condition, and their prior experience of marginalization on the portion of which they talked during team conversations (statistical test results based on transformed data: $\beta = -0.09$, $S.E. = 0.05$, $t = -1.87$, $p = 0.063$; statistical test results based

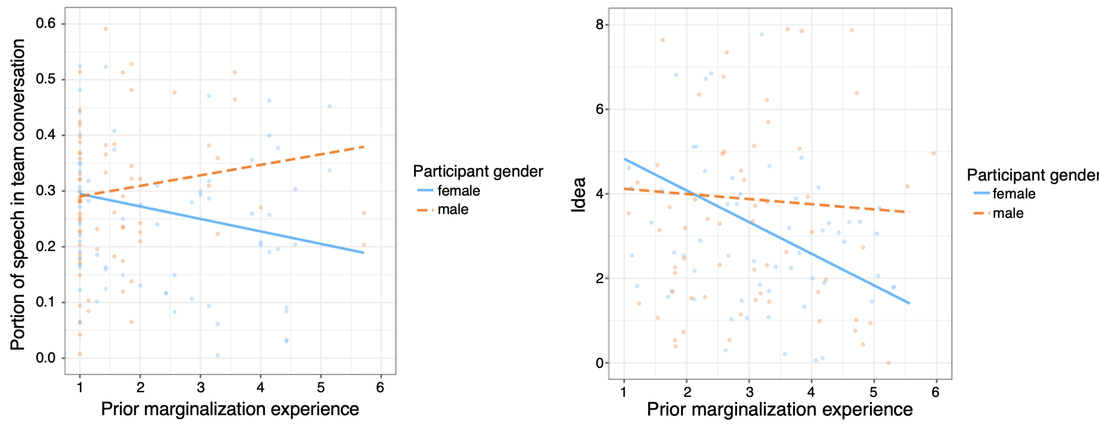

Figure 6.8: Moderating effect of participants' prior marginalization experience on team behaviors.

on raw data: $\beta = -0.15$, $S.E. = 0.12$, $t = -1.16$, $p = 0.024$) as well as the number of ideas generated during the team brainstorming session (statistical test results based on transformed data: $\beta = 1.88$, $S.E. = 0.81$, $t = 2.32$, $p = 0.022$; statistical test results based on raw data: $\beta = 0.58$, $S.E. = 0.08$, $t = 0.71$, $p = 0.048$). Specifically, female and male participants who experienced more marginalization in the past contributed fewer ideas when assigned to the majority groups. However, when they were indeed placed in a marginalized situation, male participants who experienced more intense marginalization in the past contributed more ideas during the teamwork sessions. In contrast, females who have been marginalized in the past tended to contribute less to the marginalized condition during the present study as well.

Finally, there was a marginally significant three-way interaction effect across participants' prior marginalization experience, their gender, and the types of support they received on the number of ideas generated during teamwork (statistical test results based on transformed data: $\beta = -1.88$, $S.E. = 1.09$, $t = -1.73$, $p = 0.086$; statistical test results based on raw data: $\beta = -0.60$, $S.E. = 0.13$, $t = -0.46$, $p = 0.045$). Specifically, when considering participants' prior experi-

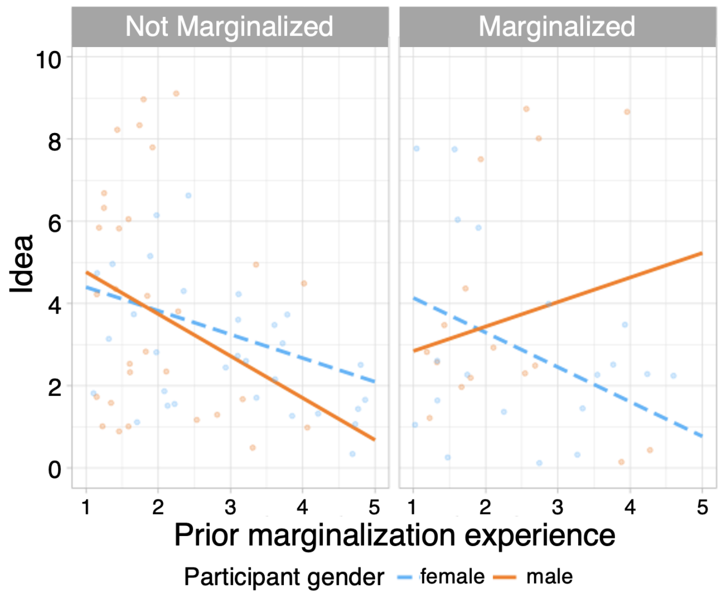

Figure 6.9: Moderating effect of participants' prior marginalization experience on their responses to marginalized conditions.

ences, male participants who have gone through more intense marginalization benefited more from having an autonomous agent on the team, regardless of the agent's gender. As shown in Figure 6.10, male participants generated more ideas with the presence of a teamwork agent when they were indeed marginalized in the present experiment. Conversely, female participants who have been marginalized in the past did not show a significant difference in their work performance (evaluated by the number of ideas generated) regardless of their assigned conditions.

6.4 Discussion

In the present study, I assigned participants to gender-imbalance teams to simulate experiences of marginalization. Under such context, I observed how female and male teammates might respond and whether the presence of a same- or

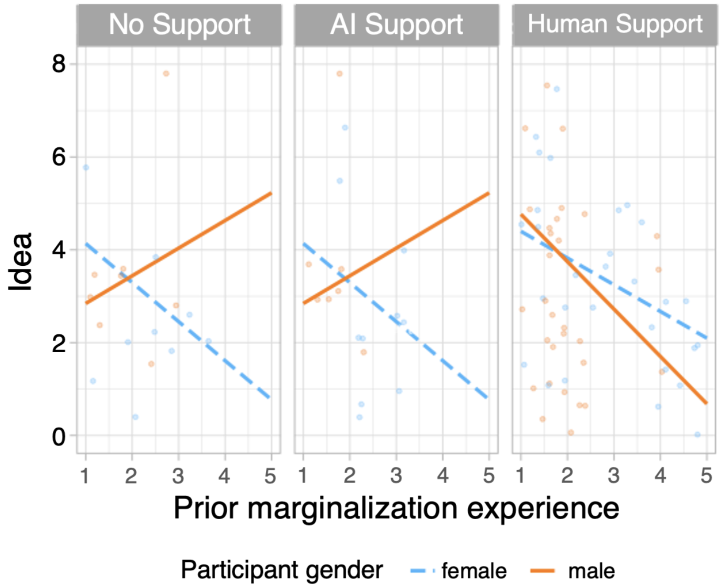

Figure 6.10: Moderating effect of participants' prior marginalization experience on the effect of support types.

different-gender agent on the team may introduce new dynamics to the teams that can potentially benefit these marginalized members. Results of the present work first support existing literature and show significant gender differences in how females and males behave when placed in a marginalized scenario, such that marginalized males tend to be more active and talkative, whereas females become more reserved when they are the minority.

According to the empirical results of the present study, this phenomenon persists even when an agent participates in teamwork. However, whether participants found shared identities with the agent led to different perceptual and behavioral outcomes. In particular, marginalized females showed more positive perceptions toward a female agent, while marginalized males had no preference for the agent's gender. Meanwhile, marginalized females became more productive in terms of generating ideas to contribute to teamwork with the presence of a female agent. Moreover, participants' prior experiences in teamwork and marginalization at work could potentially moderate their behaviors. To be specific, I found the presence of a same-gender agent was particularly effective in supporting marginalized males who also had the experience of being severely marginalized in their past experiences.

The findings of the present study have several important implications for understanding and designing applications for teamwork interactions. To begin with, I found participants' subjective experience of marginalization remains similar when working with autonomous agents in teams. For instance, females became more reserved while males talked more when they were marginalized. This suggests rich insights from existing literature on behaviors of minority versus majority groups are likely applicable to understanding their behaviors in human-agent teamwork as well. Furthermore, though these prior studies often emphasize the importance of gender diversity and the benefits of having gender-equal teams, in reality, it is not always possible to alter the group composition at work – and mostly not feasible to achieve such ideals within a short period of time.

Interestingly, participants' open-text responses at the end of the post-task survey, as well as follow-up interviews with some marginalized participants, did not show a preference for having a male or female-sounding agent. Per their qualitative feedback, participants often cared little about the identity, presence, and even performance of the agent, and they often described their reasons as "after all, it is *just* an AI." Such responses may suggest the potential of leveraging the agent's presence to support marginalized individuals in teams without putting them on the spot, as users, in general, had little concerns about the agent. In other words, while the presence of an agent can lead to behavioral changes in a subliminal fashion, the use of an agent is also not perceived as an intervention targeting to support marginalized members.

From an empirical viewpoint, the present study also shows the importance of triangulating multiple measures – including behavioral and self-report data. To begin with, I saw a significant gender difference in the extent to which females and males experienced marginalization. On one hand, this could be explained as females being more sensitive to a marginalized situation; on the other hand, this could also indicate males are less likely to report their being marginalized. However, when ratings of participants' self-report marginalization are jointly examined with their behavioral data, I still observed quite salient behavioral differences in *both* males and females' reactions to being marginalized and working with a same- or different-gender agent, despite the former reported rather mild responses to their solo status. Besides, investigating multiple types of behavioral data is also helpful in extracting more comprehensive views of marginalized individuals' experiences. Specifically, for both marginalized males and females, the amount of speech they participated in did not directly translate to their actual productivity and contribution to teams, but it instead indicates whether or not a person was able to stay focused on resolving the tasks at hand.

Despite its various findings, the study is not free from limitations. As it is challenging to target recruiting those who indeed had severely marginalized experiences, as mentioned in the results above, the majority of the sample did not consider themselves as struggling seriously with marginalization. Therefore, future work should at least consider recruiting an even larger sample size or conducting field studies to capture more experiences in real life. Besides, while the current study only entails two teamwork sessions lasting $10 \sim 15$ minutes, team dynamics can often continue to change over time. For instance, one may develop more trust and reliance on their human teammates [197]. In this regard, whether a marginalized teammate could persistently gain support from the teamwork agent after multiple rounds of team interaction remains an open-ended question.

Results of the present work show the potential to leverage the presence of intelligent agents in teams as a temporal remedy, which could be particularly encouraging for female teammates in general or male teammates who had intense marginalization experience in the past. Nonetheless, it is worth noting that, overall, the effect of the support from a same-gender agent remains relatively limited when it comes to those who have experienced intense marginalization and is particularly ineffective among females who have been marginalized intensely. Therefore, adopting supportive agents in teams should not be the permanent solution to marginalization, and seeking foundational improvement in team diversity and workplace inclusiveness remains critical in the long run.

CHAPTER 7

CONCLUSION

The present dissertation explores the role of social experiences during human-AI collaboration and examines whether and how intelligent agents prompting different levels of social experiences during interaction with users may influence their perceptions and behaviors during teamwork. According to existing literature across various fields of social and behavioral sciences, humans' social experiences are shaped through two possible pathways: (1) a bottom-up approach where individuals adopt external cues to navigate through their social environments and learn about how to interact with social others in such environments, or (2) a top-down approach where individuals hold pre-assumptions or existing beliefs about certain groups or individuals, guiding their behaviors during interactions with these entities. Likewise, humans pick up cues from or hold pre-dominating notions about autonomous agents. These two pathways again shape their behaviors, responses, and the level of social experiences they endorse during human-agent interaction.

While social experiences in human-AI collaboration could be formed through these different means, the question is whether eliciting such experiences in human-agent teamwork could benefit its outcomes. I began addressing this inquiry by examining the role of social experiences in human-AI collaboration, where an intelligent agent supports individual work. In a series of three studies, I manipulated social experiences in human-agent teamwork through either a top-down approach by informing users of the identities of their teamwork partners or through a bottom-up approach where users could sense different levels of agent's capabilities and personification. Across these studies, I found the participation of an agent can facilitate work processes, such as pushing through an idea-generation session. However, triggering social experiences in human-agent teamwork does not necessarily benefit users' experiences and performance. Instead, concerns for social evaluation arise with emerging social experiences in human-agent teamwork, distracting individuals from focusing on their tasks at hand and particularly hindering the performance of those who were already anxious in group settings.

Meanwhile, having intelligent agents participate in teams of multiple players also introduces new dynamics. In the study where I recruited and assigned participants to gender-imbalance teams (Chapter 6), the presence of a teamwork agent – particularly an agent perceived with a shared identity of its users – has the potential to support marginalized individuals' work, while such effects show gender difference. When placed in a marginalized scenario, females showed more positive perceptions toward a teamwork agent and were able to be more task-focused, yielding more productive performance with the presence of a same-gender agent. A similar behavioral pattern was observed among males who have experienced intense marginalization in the past. However, overall, the effect remains limited when it comes to supporting those who have been seriously marginalized in the past, leaving an unsolved challenge for future research.

Based on insights from its empirical findings, I further summarize the theoretical and practical implications of the present dissertation in the following sections.

7.1 Theoretical Implications

As reviewed in Section 3.1, the existing literature has proposed various perspectives (i.e., computers as autonomous agents, computers as social actors, and computers as cooperative tools) addressing why machines hold the potential to serve as teammates in collaboration settings. More specifically, these previous studies suggest computer agents could be teammates as they are often designed to be cooperative tools since their origin [71, 72, 69]. If these machines could autonomously and independently carry out the type of work humans do, they are deemed even more qualified teammates [1, 186, 132, 219, 220, 219, 89]. Meanwhile, users already hold the tendency to treat computers as social entities during human-computer interaction [237, 265, 58, 50, 303, 235]. The outstanding question is whether these prior work and theoretical frameworks can still be extended to explain the feasibility of having AI teammates and how users may respond.

Recent research has posited that human-AI interaction may be unique from other forms of human-computer interaction due to advanced technical capabilities and the probabilistic nature (i.e., the same input does not yield identical output in responses) of AI [355, 8, 7, 302]. Insights from the present dissertation resonate with these propositions, such that existing HCI theories need to be updated and extended in order to more accurately capture users' experiences and explain their behaviors in human-AI collaboration. In particular, while the Computers as Social Actors (CASA) paradigm has long served as the key pillar for scholars to apply social science theories to explain individuals' responses to computer agents, findings across the three studies in the present dissertation suggest users have now shifted their views toward intelligent agents and did not treat them like how they would with human teammates. This phenomenon is most directly reflected in findings from the three experiments in Chapter 4 where I consistently found participants showed different responses when they believed they were working with humans versus with chatbots. Similar perspectives have been posited in recent work (e.g., [128, 120, 152]), suggesting people may view intelligent agents with more unique angles – distinctively from treating agents as human-like social actors – as technologies and their own user experiences continue to expand.

Nonetheless, the extent to which users showed differing responses when working with agent teammates depends on the design of the agent and the interaction paradigm as well. In particular, introducing more anthropomorphic and more social experiences to human-AI collaboration could prompt users to treat agent teammates more like social entities. As shown in Chapter 5, when agents become more capable and more intelligent, catching up and even surpassing what human teammates could contribute to teamwork, users showed a greater tendency to socially compare themselves with the teamwork agents. Similarly, in Chapter 6, when agents were attached to gendered identities, their social presence was effective in supporting marginalized individuals in teams. In other words, though, by default, users no longer treat agents and humans alike, making these agents more anthropomorphic does encourage users to respond socially.

On one hand, this may indicate scholars could once again adopt social science theories to explain users' behaviors in human-AI interaction when highly anthropomorphic agents could perfectly simulate humans' behavioral, cognitive, and affective responses. But before such visions become widely feasible and accessible, it remains necessary to extend existing theories by investigating users' unique responses to intelligent agents. In the meantime, it is important to take into account how constant improvements in the technical affordance of AI may moderate users' expectations and experience accordingly. Finally, it is worth acknowledging that empirical studies in the present work did not directly manipulate the degree of agents' autonomy. Neither did I investigate other forms of interaction paradigm – for instance, beyond collaboration, users may as well experience competition with agents. Therefore, this again suggests avenues for future research to determine whether and how relevant theories from the other propositions (i.e., computers as cooperative tools and computers as autonomous agents) require updates, revision, or extension to capture the dynamics of human-AI collaboration.

7.2 Practical Implications

Findings from the present work also offered various practical implications for the design of teamwork agents and human-AI collaboration paradigms. To begin with, across the three studies, I found it is, in fact, beneficial to highlight the "bot-ness" of teamwork agents. While I have reviewed and assembled a number of design approaches to anthropomorphize agents in order to elicit social experiences in human-agent teamwork (see Section 3.2), applying these humanoid cues to agent design may not be necessarily helpful for users' performance. In particular, I observed this phenomenon across the empirical studies of this dissertation, during which participants were asked to prioritize their productivity and creativity in teamwork. However, this does not imply teamwork agents need to be presented as mechanical, invariable, and *robotic*. In fact, the agents used in the present work all adopted natural language for verbal and text communication, so that interactions with users during teamwork remained fluent; neither did participants need to apply specific commands or on-screen actions to interact with the agents. To emphasize the bot-ness of teamwork agents, I avoided establishing human identities and characteristics for these agents in the empirical studies, such that they would always refer to themselves as "AI agent" in all team conversations. I also did not add humanoid figures and assign names to these agents, and in fact, I intentionally minimized visual cues as much as possible when presenting the bot in all studies. Building on top of these attempts, I encourage future research to further examine whether there are specific sets of design approaches that could *de-anthropomorphize* teamwork agents to benefit human-AI teamwork.

As I advocate the criticality of addressing individual differences throughout the dissertation, I also saw findings from the series of works suggesting conceptualizing individual differences in human-AI teams may be more complex and rely more than on their demographic data. Instead, segmenting participants through their subjective experiences and self-report scales may be effective in the context of teamwork research. In particular, prioritizing the evaluation of participants' subjective experiences could already provide abundant insights into differences in team behaviors. For example, simply taking into account participants' self-reported anxiousness in group settings could consistently explain their behavioral differences in human-agent teamwork. Even more ideally, coupling subjective experiences and demographic profiles of participants could further reflect their uniqueness. Like what was conducted and reported in Chapter 6, grouping participants through their demographic data (i.e., gender) can be the first step. But when I further took into account participants' self-report marginalization experiences in the past, the interventions (i.e., the support from a same- or different-gender agent) then showed very different effects on each gender.

The present dissertation also shows the need to consider more sustainable team relationships between users and agents, as what benefits a single teamwork session (e.g., the Smart Bot that brought high-quality ideas to a brainstorming session) does not guarantee continuous cooperation and collaboration between humans and machine teammates. In this regard, I first propose examining team effectiveness beyond a single iteration as a crucial methodological practice for human-AI collaboration research. As executed in the current studies, triangulating across multiple measurements could enable a more comprehensive view of teamwork effectiveness, and it is particularly helpful for projecting and understanding users' behaviors beyond a single occurrence. While abundant teamwork research measures team performance, team experience, and perceptions toward teammates simultaneously (e.g., [67, 66, 220]), the present series of work further supports the benefit of this methodological approach and finds it particularly applicable to understanding the complex dynamics in human-AI collaboration.

7.3 Closing Thoughts

Synthesizing the literature review and empirical findings from multiple studies, I conclude that while the presence and participation of autonomous agents show the potential to bring out individuals' productivity and creativity, additional interventions to elicit social experiences in human-AI collaboration are not necessarily beneficial and may inhibit individuals' performance in specific circumstances. In other words, the mere social presence of autonomous agents could be sufficiently advantageous to collaboration settings, and such teamwork agents need not demonstrate personified behaviors nor simulate human-human interaction paradigms.

Findings from the present dissertation could provide important theoretical and practical implications for the future of work where the use of AI applications and the participation of intelligent agents at work serves as new norms. While much discussion centers around whether AI should serve as *either* a tool *or* as a human-like teammate as it becomes more pervasive in work practices of all fields, I posit to situate the use of AI in its unique position and move beyond this dichotomous view when addressing challenges around human-AI collaboration. Since its origin, AI has served unique functions unlike other tools and let alone humans. While its technical capabilities continue to grow, use cases of AI may expand accordingly, but the purpose of advancing human work does not necessarily shift. Indeed, these intelligent artifacts were not designed to replace humans or any specific tools in the first place, and up to the present, this original intention still persists.

REFERENCES

[1] Hussein Abbass, Axel Bender, Svetoslav Gaidow, and Paul Whitbread. 2011. Computational Red Teaming: Past, Present and Future. *IEEE Computational Intelligence Magazine* 6, 1 (2011), 30–42. https://doi.org/10.1109/MCI.2010.939578

[2] Gavin Abercrombie, Amanda Cercas Curry, Mugdha Pandya, and Verena Rieser. 2021. Alexa, Google, Siri: What are Your Pronouns? Gender and Anthropomorphism in the Design and Perception of Conversational Assistants. (June 2021). https://doi.org/10.48550/arXiv.2106.02578 arXiv:2106.02578 [cs].

[3] Ralph Adolphs. 1999. Social Cognition and the Human Brain. *Trends in Cognitive Sciences* 3, 12 (Dec. 1999), 469–479. https://doi.org/10.1016/S1364-6613(99)01399-6

[4] John R. Aiello and Elizabeth A. Douthitt. 2001. Social Facilitation from Triplett to Electronic Performance Monitoring. *Group Dynamics: Theory, Research, and Practice* 5, 3 (2001), 163–180. https://doi.org/10.1037/1089-2699.5.3.163

[5] Floyd H Allport. 1920. The Influence of the Group upon Association and Thought. *Journal of Experimental Psychology* 3, 3 (1920), 159.

[6] Gordon W. Allport. 1954. *The Nature of Prejudice*. Addison-Wesley, Reading, Massachusetts.

[7] Saleema Amershi, Maya Cakmak, William Bradley Knox, and Todd Kulesza. 2014. Power to the People: The Role of Humans in Interactive Machine Learning. *AI Magazine* 35, 4 (Dec. 2014), 105–120. https://doi.org/10.1609/aimag.v35i4.2513

[8] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300233

[9] Elizabeth Joan Aries. 1982. Verbal and Nonverbal Behavior in Single-Sex and Mixed-Sex Groups: Are Traditional Sex Roles Changing? *Psychological Reports* 51, 1 (Aug. 1982), 127–134. https://doi.org/10.2466/pr0.1982.51.1.127

[10] Zahra Ashktorab, Q. Vera Liao, Casey Dugan, James Johnson, Qian Pan, Wei Zhang, Sadhana Kumaravel, and Murray Campbell. 2020. Human-AI Collaboration in a Cooperative Game Setting: Measuring Social Perception and Outcomes. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW2, Article 96 (Oct. 2020), 20 pages. https://doi.org/10.1145/3415167

[11] Chen Avin, Barbara Keller, Zvi Lotker, Claire Mathieu, David Peleg, and Yvonne-Anne Pignolet. 2015. Homophily and the Glass Ceiling Effect in Social Networks. In *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science*. Association for Computing Machinery, Rehovot Israel, 41–50. https://doi.org/10.1145/2688073.2688097

[12] Soufyane Ayanouz, Boudhir Anouar Abdelhakim, and Mohammed Benhmed. 2020. A Smart Chatbot Architecture Based NLP and Machine Learning for Health Care Assistance. In *Proceedings of the 3rd International Conference on Networking, Information Systems amp; Security* (Marrakech, Morocco) *(NISS2020)*. Association for Computing Machinery, New York, NY, USA, Article 78, 6 pages. https://doi.org/10.1145/3386723.3387897

[13] Matthew P. Aylett, Selina Jeanne Sutton, and Yolanda Vazquez-Alvarez. 2019. The Right Kind of Unnatural: Designing a Robot Voice. In *Proceedings of the 1st International Conference on Conversational User Interfaces* (Dublin, Ireland) *(CUI '19)*. Association for Computing Machinery, New York, NY, USA, Article 25, 2 pages. https://doi.org/10.1145/3342775.3342806

[14] David Azul. 2015. Transmasculine People's Vocal Situations: A Critical Review of Gender-Related Discourses and Empirical Data. *International Journal of Language Communication Disorders* 50, 1 (Jan. 2015), 31–47. https://doi.org/10.1111/1460-6984.12121

[15] Cyril Joe Baby, Faizan Ayyub Khan, and J. N. Swathi. 2017. Home automation using IoT and a chatbot using natural language processing. In *2017 Innovations in Power and Advanced Computing Technologies (i-PACT)*. 1–6. https://doi.org/10.1109/IPACT.2017.8245185

[16] Renée Baillargeon, Rose M. Scott, and Zijing He. 2010. False-belief Understanding in Infants. *Trends in Cognitive Sciences* 14, 3 (March 2010), 110–118. https://doi.org/10.1016/j.tics.2009.12.006

[17] Daniel Balliet, Junhui Wu, and Carsten K. W. De Dreu. 2014. Ingroup Favoritism in Cooperation: A Meta-analysis. *Psychological Bulletin* 140, 6 (2014), 1556–1581. https://doi.org/10.1037/a0037737

[18] Gagan Bansal, Besmira Nushi, Ece Kamar, Eric Horvitz, and

Daniel S. Weld. 2021. Is the Most Accurate AI the Best Teammate? Optimizing AI for Teamwork. *Proceedings of the AAAI Conference on Artificial Intelligence* 35, 13 (May 2021), 11405–11414. https://doi.org/10.1609/aaai.v35i13.17359

[19] Gagan Bansal, Besmira Nushi, Ece Kamar, Daniel S. Weld, Walter S. Lasecki, and Eric Horvitz. 2019. Updates in Human-AI Teams: Understanding and Addressing the Performance/Compatibility Tradeoff. *Proceedings of the AAAI Conference on Artificial Intelligence* 33, 01 (July 2019), 2429–2437. https://doi.org/10.1609/aaai.v33i01.33012429

[20] Gagan Bansal, Tongshuang Wu, Joyce Zhou, Raymond Fok, Besmira Nushi, Ece Kamar, Marco Tulio Ribeiro, and Daniel Weld. 2021. Does the Whole Exceed its Parts? The Effect of AI Explanations on Complementary Team Performance. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Yokohama Japan, 1–16. https://doi.org/10.1145/3411764.3445717

[21] Robert S. Baron. 1986. *Distraction-Conflict Theory: Progress and Problems*. Vol. 19. Elsevier, 1–40. https://doi.org/10.1016/S0065-2601(08)60211-7

[22] Robert S Baron, Danny Moore, and Glenn S Sanders. 1978. Distraction as a Source of Drive in Social Facilitation Research. *Journal of Personality and Social Psychology* 36, 8 (1978), 816.

[23] Dale J. Barr and Mandana Seyfeddinipur. 2010. The Role of Fillers in Listener Attributions for Speaker Disfluency. *Language and Cognitive Processes* 25, 4 (May 2010), 441–455. https://doi.org/10.1080/01690960903047122

[24] John Barresi and Chris Moore. 1996. Intentional Relations and Social Understanding. *Behavioral and Brain Sciences* 19, 1 (March 1996), 107–122. https://doi.org/10.1017/S0140525X00041790

[25] Lisa Feldman Barrett, Ralph Adolphs, Stacy Marsella, Aleix M. Martinez, and Seth D. Pollak. 2019. Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. *Psychological Science in the Public Interest* 20, 1 (2019), 1–68.

[26] Katharine T. Bartlett. 2009. Making Good on Good Intentions: The Critical Role of Motivation in Reducing Implicit Workplace Discrimination. *Virginia Law Review* 95, 8 (2009), 1893–1972. http://www.jstor.org/stable/27759975 Publisher: Virginia Law Review.

[27] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*

67, 1 (2015), 1–48. https://doi.org/10.18637/jss.v067.i01

[28] Julia B Bear and Anita Williams Woolley. 2011. The Role of Gender in Team Collaboration and Performance. *Interdisciplinary Science Reviews* 36, 2 (June 2011), 146–153. https://doi.org/10.1179/030801811X13013181961473

[29] Behzad Beigpourian and Matthew Ohland. 2019. A Systematized Review: Gender and Race in Teamwork in Undergraduate Engineering Classrooms. In *2019 ASEE Annual Conference Exposition Proceedings*. ASEE Conferences, Tampa, Florida, 32011. https://doi.org/10.18260/1-2–32011

[30] Stefano R. Belli, Robert D. Rogers, and Jennifer Y.F. Lau. 2012. Adult and Adolescent Social Reciprocity: Experimental Data from the Trust Game. *Journal of Adolescence* 35, 5 (Oct. 2012), 1341–1349. https://doi.org/10.1016/j.adolescence.2012.05.004

[31] Jennifer L. Berdahl and Cameron Anderson. 2005. Men, Women, and Leadership Centralization in Groups Over Time. *Group Dynamics: Theory, Research, and Practice* 9, 1 (March 2005), 45–57. https://doi.org/10.1037/1089-2699.9.1.45

[32] Cindy L Bethel, Merijn Bruijnes, Malte Jung, Christoforos Mavrogiannis, Simon Parsons, Catherine Pelachaud, Rui Prada, Laurel Riek, Sarah Strohkorb Sebo, Julie Shah, et al. 2020. 4.4 Working Group on Social Cognition for Robots and Virtual Agents. *Dagstuhl Reports, Vol. 9, Issue 10 ISSN 2192-5283* (2020), 21.

[33] Reuben Binns, Max Van Kleek, Michael Veale, Ulrik Lyngs, Jun Zhao, and Nigel Shadbolt. 2018. 'It's Reducing a Human Being to a Percentage' Perceptions of Justice in Algorithmic Decisions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–14.

[34] Filip Bircanin, Bernd Ploderer, Laurianne Sitbon, Andrew A. Bayor, and Margot Brereton. 2019. Challenges and Opportunities in Using Augmentative and Alternative Communication (AAC) Technologies: Design Considerations for Adults with Severe Disabilities. In *Proceedings of the 31st Australian Conference on Human-Computer-Interaction* (Fremantle, WA, Australia) *(OZCHI'19)*. Association for Computing Machinery, New York, NY, USA, 184–196. https://doi.org/10.1145/3369457.3369473

[35] Eva AC Bittner, Sarah Oeste-Reiß, and Jan Marco Leimeister. 2019. Where is the bot in our team? Toward a taxonomy of design option combinations for conversational agents in collaborative work. In *Hawaii International Conference on System Sciences (HICSS)*.

[36] Jim Blascovich, Wendy Berry Mendes, Sarah B Hunter, and Kristen Salomon. 1999. Social "Facilitation" as Challenge and Threat. *Journal of Personality and Social Psychology* 77, 1 (1999), 68.

[37] Markus Blut, Cheng Wang, Nancy V. Wünderlich, and Christian Brock. 2021. Understanding anthropomorphism in service provision: a meta-analysis of physical robots, chatbots, and other AI. *Journal of the Academy of Marketing Science* 49, 4 (July 2021), 632–658. https://doi.org/10.1007/s11747-020-00762-y

[38] Jennifer Boldry, Wendy Wood, and Deborah A. Kashy. 2001. Gender Stereotypes and the Evaluation of Men and Women in Military Training. *Journal of Social Issues* 57, 4 (Jan. 2001), 689–705. https://doi.org/10.1111/0022-4537.00236

[39] Charles F Bond. 1982. Social Facilitation: A Self-presentational View. *Journal of Personality and Social Psychology* 42, 6 (1982), 1042.

[40] Manuele Brambilla, Eliseo Ferrante, Mauro Birattari, and Marco Dorigo. [n. d.]. Swarm robotics: a review from the swarm engineering perspective. 7, 1 ([n. d.]), 1–41. https://doi.org/10.1007/s11721-012-0075-2

[41] Michael T. Brannick, Eduardo Salas, and Carolyn Prince. 1997. *Team Performance Assessment and Measurement: Theory, Methods, and Applications*. Lawrence Erlbaum Associates, Mahwah, N.J.

[42] Cynthia Breazeal. [n. d.]. Toward sociable robots. 42, 3 ([n. d.]), 167–175. https://doi.org/10.1016/S0921-8890(02)00373-1

[43] Cynthia Breazeal. 2003. Emotion and sociable humanoid robots. *International journal of human-computer studies* 59, 1-2 (2003), 119–155.

[44] Cynthia L Breazeal. 2004. *Designing sociable robots*. MIT press.

[45] Marilynn B. Brewer. 1999. The Psychology of Prejudice: Ingroup Love and Outgroup Hate? *Journal of Social Issues* 55, 3 (Jan. 1999), 429–444. https://doi.org/10.1111/0022-4537.00126

[46] Alison Wood Brooks and Maurice E. Schweitzer. 2011. Can Nervous Nelly Negotiate? How Anxiety Causes Negotiators to Make Low First Offers, Exit Early, and Earn Less Profit. *Organizational Behavior and Human Decision Processes* 115, 1 (May 2011), 43–54. https://doi.org/10.1016/j.obhdp.2011.01.008

[47] Vicki Bruce and Andy Young. 1986. Understanding Face Recognition. *British Journal of Psychology* 77, 3 (Aug. 1986), 305–327.

https://doi.org/10.1111/j.2044-8295.1986.tb02199.x

[48] Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, et al. 2023. Sparks of Artificial General Intelligence: Early Experiments with GPT-4. *arXiv preprint arXiv:2303.12712* (2023).

[49] Ross Buck, Jeffrey I. Losow, Mark M. Murphy, and Paul Costanzo. 1992. Social Facilitation and Inhibition of Emotional Expression and Communication. *Journal of Personality and Social Psychology* 63, 6 (1992), 962–968. https://doi.org/10.1037/0022-3514.63.6.962

[50] J.K Burgoon, J.A Bonito, B Bengtsson, C Cederberg, M Lundeberg, and L Allspach. 2000. Interactivity in human–computer interaction: a study of credibility, understanding, and influence. 16, 6 (2000), 553–574. https://doi.org/10.1016/S0747-5632(00)00029-7

[51] Federico Cabitza, Andrea Campagner, and Valentina Cavosi. 2021. Assessing the Impact of Medical AI: A Survey of Physicians' Perceptions. In *Proceedings of the 5th International Conference on Medical and Health Informatics* (Kyoto, Japan) *(ICMHI '21)*. Association for Computing Machinery, New York, NY, USA, 225–231. https://doi.org/10.1145/3472813.3473195

[52] Carrie J. Cai, Samantha Winter, David Steiner, Lauren Wilcox, and Michael Terry. 2019. "Hello AI": Uncovering the Onboarding Needs of Medical Practitioners for Human-AI Collaborative Decision-Making. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 1–24. https://doi.org/10.1145/3359206

[53] John C. Calanni, Saba N. Siddiki, Christopher M. Weible, and William D. Leach. 2015. Explaining Coordination in Collaborative Partnerships and Clarifying the Scope of the Belief Homophily Hypothesis. *Journal of Public Administration Research and Theory* 25, 3 (July 2015), 901–927. https://doi.org/10.1093/jopart/mut080

[54] L. Mabel Camacho and Paul B. Paulus. 1995. The Role of Social Anxiousness in Group Brainstorming. *Journal of Personality and Social Psychology* 68, 6 (June 1995), 1071–1080. https://doi.org/10.1037/0022-3514.68.6.1071

[55] Julia Cambre, Jessica Colnago, Jim Maddock, Janice Tsai, and Jofish Kaye. 2020. *Choice of Voices: A Large-Scale Evaluation of Text-to-Speech Voice Quality for Long-Form Content*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376789

[56] Alessandro Canossa, Dmitry Salimov, Ahmad Azadvar, Casper Harteveld, and Georgios Yannakakis. 2021. For Honor, for Toxicity: Detecting Toxic Behavior through Gameplay. *Proceedings of the ACM on Human-Computer Interaction (PACM HCI)* 5, CHI PLAY, Article 253 (Oct. 2021), 29 pages. https://doi.org/10.1145/3474680

[57] Bolin Cao and Wan-Ying Lin. 2017. Revisiting the Contact Hypothesis: Effects of Different Modes of Computer-Mediated Communication on Intergroup Relationships. *International Journal of Intercultural Relations* 58 (May 2017), 23–30. https://doi.org/10.1016/j.ijintrel.2017.03.003

[58] John M. Carroll. 1997. Human–computer interaction: psychology as a science of design. 46, 4 (1997), 501–522. https://doi.org/10.1006/ijhc.1996.0101

[59] Charles S Carver and Michael F Scheier. 1981. The Self-attention-induced Feedback Loop and Social Facilitation. *Journal of Experimental Social Psychology* 17, 6 (1981), 545–568.

[60] Luigi Castelli, Luciana Carraro, Giulia Pavan, Elisa Murelli, and Alessia Carraro. 2012. The Power of the Unsaid: The Influence of Nonverbal Cues on Implicit Attitudes: Power of the Unsaid. *Journal of Applied Social Psychology* 42, 6 (June 2012), 1376–1393. https://doi.org/10.1111/j.1559-1816.2012.00903.x

[61] Luigi Castelli, Leyla De Amicis, and Steven J. Sherman. 2007. The Loyal Member Effect: On the Preference for Ingroup Members Who Engage in Exclusive Relations with the Ingroup. *Developmental Psychology* 43, 6 (Nov. 2007), 1347–1359. https://doi.org/10.1037/0012-1649.43.6.1347

[62] Linda G. Castillo, Collie W. Conoley, Daniel F. Brossart, and Alexander E. Quiros. 2007. Construction and Validation of the Intragroup Marginalization Inventory. *Cultural Diversity and Ethnic Minority Psychology* 13, 3 (July 2007), 232–240. https://doi.org/10.1037/1099-9809.13.3.232

[63] Jesse Chandler, Cheskie Rosenzweig, Aaron J. Moss, Jonathan Robinson, and Leib Litman. 2019. Online Panels in Social Science Research: Expanding Sampling Methods Beyond Mechanical Turk. *Behavior Research Methods* 51, 5 (Oct. 2019), 2022–2038. https://doi.org/10.3758/s13428-019-01273-7

[64] Min-Hua Chao, Amy J. C. Trappey, and Chun-Ting Wu. [n. d.]. Emerging Technologies of Natural Language-Enabled Chatbots: A Review and Trend Forecast Using Intelligent Ontology Extraction and Patent Analytics. 2021 ([n. d.]), 1–26. https://doi.org/10.1155/2021/5511866

[65] Anjan Chatterjee, Amy Thomas, Sabrina E. Smith, and Geoffrey K. Aguirre. 2009. The Neural Response to Facial Attractiveness. *Neuropsychology* 23, 2 (March 2009), 135–143. https://doi.org/10.1037/a0014430

[66] Gilad Chen and Ruth Kanfer. 2006. Toward a Systems Theory of Motivated Behavior in Work Teams. *Research in Organizational Behavior* 27 (Jan. 2006), 223–267. https://doi.org/10.1016/S0191-3085(06)27006-0

[67] Gilad Chen, Ruth Kanfer, Richard P DeShon, John E Mathieu, and Steve W. J. Kozlowski. 2009. The Motivating Potential of Teams: Test and Extension of Cross-Level Model of Motivation in Teams. *Organizational Behavior and Human Decision Processes* 110, 1 (2009), 45–55. Publisher: Elsevier.

[68] EunJeong Cheon, Cristina Zaga, Hee Rin Lee, Maria Luce Lupetti, Lynn Dombrowski, and Malte F. Jung. 2021. Human-Machine Partnerships in the Future of Work: Exploring the Role of Emerging Technologies in Future Workplaces. In *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing*. Association for Computing Machinery, Virtual Event USA, 323–326. https://doi.org/10.1145/3462204.3481726

[69] Erin K. Chiou and John D. Lee. [n. d.]. Cooperation in Human-Agent Systems to Support Resilience: A Microworld Experiment. 58, 6 ([n. d.]), 846–863. https://doi.org/10.1177/0018720816649094

[70] H Clark. 2002. Using Uh and Um in Spontaneous Speaking. *Cognition* 84, 1 (May 2002), 73–111. https://doi.org/10.1016/S0010-0277(02)00017-3

[71] A.A. Clarke and M.G.G. Smyth. 1993. A Co-operative Computer Based on the Principles of Human Co-operation. *International Journal of Man-Machine Studies* 38, 1 (Jan. 1993), 3–22. https://doi.org/10.1006/imms.1993.1002

[72] Peter Cole and Jerry L. Morgan (Eds.). 1975. *Logic and Conversation*. BRILL. 41–58 pages. https://doi.org/10.1163/9789004368811$_0$03

[73] Rebecca L. Collins. 1996. For Better or Worse: The Impact of Upward Social Comparison on Self-Evaluations. *Psychological Bulletin* 119, 1 (1996), 51–69. https://doi.org/10.1037/0033-2909.119.1.51

[74] Nancy J. Cooke, Eduardo Salas, Janis A. Cannon-Bowers, and Renée J. Stout. 2000. Measuring Team Knowledge. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 42, 1 (March 2000), 151–173. https://doi.org/10.1518/001872000779656561

[75] Jose Cordova, Virginia Eaton, and Kimberly Taylor. 2011. Experiences in Computer Science Wonderland: A Success Story with Alice. *Journal of Computing Sciences in Colleges* 26, 5 (May 2011), 16–22.

[76] Martin Corley, Lucy J. MacGregor, and David I. Donaldson. 2007. It's The Way That You, er, Say It: Hesitations in Speech Affect Language Comprehension. *Cognition* 105, 3 (Dec. 2007), 658–668. https://doi.org/10.1016/j.cognition.2006.10.010

[77] Jane M. Craig and Carolyn W. Sherif. 1986. The Effectiveness of Men and Women in Problem-Solving Groups as a Function of Group Gender Composition. *Sex Roles* 14, 7–8 (April 1986), 453–466. https://doi.org/10.1007/BF00288427

[78] Tom Crick, Tom Prickett, Jill Bradnum, and Alan Godfrey. 2022. Gender Parity in Peer Assessment of Team Software Development Projects. In *Proceedings of 6th Conference on Computing Education Practice* (Durham, United Kingdom) *(CEP '22)*. Association for Computing Machinery, New York, NY, USA, 9–12. https://doi.org/10.1145/3498343.3498346

[79] Emily S. Cross and Richard Ramsey. 2021. Mind Meets Machine: Towards a Cognitive Science of Human–Machine Interactions. *Trends in Cognitive Sciences* 25, 3 (March 2021), 200–212. https://doi.org/10.1016/j.tics.2020.11.009

[80] Joe Crumpton and Cindy L. Bethel. 2016. A Survey of Using Vocal Prosody to Convey Emotion in Robot Speech. *International Journal of Social Robotics* 8, 2 (April 2016), 271–285. https://doi.org/10.1007/s12369-015-0329-4

[81] Arturo Cruz-Maya, François Ferland, and Adriana Tapus. 2015. Social facilitation in a game-like human-robot interaction using synthesized emotions and episodic memory. In *International Conference on Social Robotics*. Springer, 164–173.

[82] David Crystal. 1980. *A First Dictionary of Linguistics and Phonetics*. A. Deutsch, London.

[83] Haydee M. Cuevas, Stephen M. Fiore, Barrett S. Caldwell, and Laura Strater. 2007. Augmenting Team Cognition in Human-Automation Teams Performing in Complex Operational Environments. 78, 5 (2007), B63–B70.

[84] Alba Curry and Amanda Cercas Curry. 2022. Computer Says "No": The Case Against Empathetic Conversational AI. (Dec. 2022). https://doi.org/10.48550/arXiv.2212.10983 arXiv:2212.10983 [cs].

[85] Nicole B. Damen and Christine A. Toh. 2018. Implicit and Explicit Trust Behavior: Does Stereotype Congruence Affect User Trust in a Home Automation Device?. In *Volume 7: 30th International Conference on Design Theory and Methodology*. American Society of Mechanical Engineers, Quebec City, Quebec, Canada, V007T06A033. https://doi.org/10.1115/DETC2018-86091

[86] Luisa Damiano and Paul Dumouchel. 2018. Anthropomorphism in Human–Robot Co-evolution. 9 (2018), 468. https://doi.org/10.3389/fpsyg.2018.00468

[87] Valdemar Danry, Pat Pataranutaporn, Adam Haar Horowitz, Paul Strohmeier, Josh Andres, Rakesh Patibanda, Zhuying Li, Takuto Nakamura, Jun Nishida, Pedro Lopes, Felipe León, Andrea Stevenson Won, Dag Svanæs, Florian Floyd Mueller, Pattie Maes, Sang-won Leigh, and Nathan Semertzidis. 2021. Do Cyborgs dream of Electric Limbs? Experiential Factors in Human-Computer Integration Design and Evaluation. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Yokohama Japan, 1–6. https://doi.org/10.1145/3411763.3441355

[88] Ewart J. de Visser, Samuel S. Monfort, Ryan McKendrick, Melissa A. B. Smith, Patrick E. McKnight, Frank Krueger, and Raja Parasuraman. [n. d.]. Almost human: Anthropomorphism increases trust resilience in cognitive agents. 22, 3 ([n. d.]), 331–349. https://doi.org/10.1037/xap0000092

[89] Mustafa Demir, Nathan J. McNeese, and Nancy J. Cooke. [n. d.]. Team communication behaviors of the human-automation teaming. In *2016 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)* (San Diego, CA, USA, 2016-03). IEEE, 28–34. https://doi.org/10.1109/COGSIMA.2016.7497782

[90] A Dennis and M Williams. 2003. Electronic Brainstorming. *Group Creativity: Innovation through Collaboration* (2003), 160–178.

[91] Pieternel Dijkstra and Bram P Buunk. 1998. Jealousy as a Function of Rival Characteristics: An Evolutionary Perspective. *Personality and Social Psychology Bulletin* 24, 11 (1998), 1158–1166. Publisher: Sage Publications Sage CA: Thousand Oaks, CA.

[92] Shena J. Dillon, Whitney Kleinmann, Angela Seasely, Rebecca Ames, Phyllis Dyess-Nugent, Donald D. McIntire, Ellen Suen, and David B. Nelson. 2021. How Personality Affects Teamwork: A Study in Multidisciplinary Obstetrical Simulation. *American Journal of Obstetrics Gynecology*

3, 2 (March 2021), 100303. https://doi.org/10.1016/j.ajogmf.2020.100303

[93] Tanvi Dinkar, Chloé Clavel, and Ioana Vasilescu. 2023. Fillers in Spoken Language Understanding: Computational and Psycholinguistic Perspectives. (March 2023). http://arxiv.org/abs/2301.10761 arXiv:2301.10761 [cs].

[94] Karen M. Douglas and Craig McGarty. 2001. Identifiability and Self-Presentation: Computer-Mediated Communication and Intergroup Interaction. *British Journal of Social Psychology* 40, 3 (Sept. 2001), 399–416. https://doi.org/10.1348/014466601164894

[95] John F. Dovidio, Clifford E. Brown, Karen Heltman, Steve L. Ellyson, and Caroline F. Keating. 1988. Power Displays between Women and Men in Discussions of Gender-Linked Tasks: A Multichannel Study. *Journal of Personality and Social Psychology* 55, 4 (Oct. 1988), 580–587. https://doi.org/10.1037/0022-3514.55.4.580

[96] John F. Dovidio, Samuel L. Gaertner, and Tamar Saguy. 2007. Another View of "We": Majority and Minority Group Perspectives on a Common Ingroup Identity. *European Review of Social Psychology* 18, 1 (Nov. 2007), 296–330. https://doi.org/10.1080/10463280701726132

[97] James E. Driskell, Gerald F. Goodwin, Eduardo Salas, and Patrick Gavan O'Shea. 2006. What Makes a Good Team Player? Personality and Team Effectiveness. *Group Dynamics: Theory, Research, and Practice* 10, 4 (Dec. 2006), 249–271. https://doi.org/10.1037/1089-2699.10.4.249

[98] James E. Driskell, Eduardo Salas, and Joan Johnston. 1999. Does stress lead to a loss of team perspective? *Group Dynamics: Theory, Research, and Practice* 3, 4 (Dec. 1999), 291–302. https://doi.org/10.1037/1089-2699.3.4.291

[99] Wen Duan, Naomi Yamashita, and Susan R. Fussell. 2019. Increasing Native Speakers' Awareness of the Need to Slow Down in Multilingual Conversations Using a Real-Time Speech Speedometer. *Proceedings of the ACM on Human-Computer Interaction (PACM HCI)* 3, CSCW, Article 171 (Nov. 2019), 25 pages. https://doi.org/10.1145/3359273

[100] Brian R. Duffy. 2003. Anthropomorphism and the social robot. 42, 3 (2003), 177–190. https://doi.org/10.1016/S0921-8890(02)00374-3

[101] Elizabeth Duffy. 1962. *Activation and Behavior*. Wiley, New York.

[102] Cydney H. Dupree and Susan T. Fiske. 2019. Self-Presentation in In-

terracial Settings: The Competence Downshift by White Liberals. *Journal of Personality and Social Psychology* 117, 3 (Sept. 2019), 579–604. https://doi.org/10.1037/pspi0000166

[103] Alice H. Eagly and Blair T. Johnson. 1990. Gender and Leadership Style: A Meta-analysis. *Psychological Bulletin* 108, 2 (Sept. 1990), 233–256. https://doi.org/10.1037/0033-2909.108.2.233

[104] Chad Edwards, Autumn Edwards, Brett Stoll, Xialing Lin, and Noelle Massey. 2019. Evaluations of an Artificial Intelligence Instructor's Voice: Social Identity Theory in Human-Robot Interactions. *Computers in Human Behavior* 90 (Jan. 2019), 357–362. https://doi.org/10.1016/j.chb.2018.08.027

[105] Paul Ekman. 1973. *Darwin and Facial Expression : A Century of Research in Review*. Academic Press, New York.

[106] Edona Elshan and Philipp Ebel. [n. d.]. Let's team up: Designing conversational agents as teammates. In *International Conference on Information Systems (ICIS)* (2020).

[107] Nicholas Epley, Adam Waytz, and John T. Cacioppo. 2007. On Seeing Human: A Three-Factor Theory of Anthropomorphism. *Psychological Review* 114, 4 (2007), 864–886. https://doi.org/10.1037/0033-295X.114.4.864

[108] Gokhan Ertug, Julia Brennecke, Balázs Kovács, and Tengjian Zou. 2022. What Does Homophily Do? A Review of the Consequences of Homophily. *Academy of Management Annals* 16, 1 (Jan. 2022), 38–69. https://doi.org/10.5465/annals.2020.0230

[109] Cole Etherington, Simon Kitto, Joseph K. Burns, Tracey L. Adams, Arija Birze, Meghan Britton, Sukhbir Singh, and Sylvain Boet. 2021. How Gender Shapes Interprofessional Teamwork in The Operating Room: A Qualitative Secondary Analysis. *BMC Health Services Research* 21, 1 (Dec. 2021), 1357. https://doi.org/10.1186/s12913-021-07403-2

[110] Anthony M. Evans and Joachim I. Krueger. 2011. Elements of Trust: Risk and Perspective-taking. *Journal of Experimental Social Psychology* 47, 1 (Jan. 2011), 171–177. https://doi.org/10.1016/j.jesp.2010.08.007

[111] Friederike Eyssel, Dieta Kuchenbrandt, and Simon Bobinger. 2011. Effects of Anticipated Human-Robot Interaction and Predictability of Robot Behavior on Perceptions of Anthropomorphism. In *Proceedings of the 6th International Conference on Human-Robot Interaction*. 61–68.

[112] Liz W. Faber. 2020. *The Computer's Voice: From Star Trek to Siri*. University of Minnesota Press. https://doi.org/10.5749/j.ctv1bzfnsv

[113] G. T. Fairhurst and B. K. Snavely. 1983. A Test of the Social Isolation of Male Tokens. *Academy of Management Journal* 26, 2 (June 1983), 353–361. https://doi.org/10.2307/255983

[114] Teresa Farroni, Mark H. Johnson, Enrica Menon, Luisa Zulian, Dino Faraguna, and Gergely Csibra. 2005. Newborns' Preference for Face-relevant Stimuli: Effects of Contrast Polarity. *Proceedings of the National Academy of Sciences* 102, 47 (Nov. 2005), 17245–17250. https://doi.org/10.1073/pnas.0502205102

[115] L. Fernandez-Sanz and Sanjay Misra. 2012. Analysis of Cultural and Gender Influences on Teamwork Performance for Software Requirements Analysis in Multinational Environments. *IET Software* 6, 3 (2012), 167. https://doi.org/10.1049/iet-sen.2011.0070

[116] Leon Festinger. 1954. A Theory of Social Comparison Processes. *Human Relations* 7, 2 (1954), 117–140. Publisher: Sage Publications Sage CA: Thousand Oaks, CA.

[117] Anne-Kathrin J. Fett, Wolfgang Viechtbauer, Maria-de-Gracia Dominguez, David L. Penn, Jim Van Os, and Lydia Krabbendam. 2011. The Relationship between Neurocognition and Social Cognition with Functional Outcomes in Schizophrenia: A Meta-analysis. *Neuroscience Biobehavioral Reviews* 35, 3 (Jan. 2011), 573–588. https://doi.org/10.1016/j.neubiorev.2010.07.001

[118] Julia Fink. 2012. Anthropomorphism and human likeness in the design of robots and human-robot interaction. In *International Conference on Social Robotics*. Springer, 199–208.

[119] Samantha L. Finkelstein, Eve Powell, Andrew Hicks, Katelyn Doran, Sandhya Rani Charugulla, and Tiffany Barnes. 2010. SNAG: Using Social Networking Games to Increase Student Retention in Computer Science. In *Proceedings of the Fifteenth Annual Conference on Innovation and Technology in Computer Science Education* (Bilkent, Ankara, Turkey) *(ITiCSE '10)*. Association for Computing Machinery, New York, NY, USA, 142–146. https://doi.org/10.1145/1822090.1822131

[120] Jesse Fox and Bree McEwan. 2017. Distinguishing Technologies for Social Interaction: The Perceived Social Affordances of Communication Channels Scale. *Communication Monographs* 84, 3 (2017), 298–318.

[121] Scott H. Fraundorf, Jennifer Arnold, and Valerie J. Langlois. 2014. Disfluency. *Oxford Bibliographies* (2014). https://doi.org/10.1093/obo/9780199772810-0189

[122] Terri A. Fredrick. 2008. Facilitating Better Teamwork: Analyzing the Challenges and Strategies of Classroom-Based Collaboration. *Business Communication Quarterly* 71, 4 (Dec. 2008), 439–455. https://doi.org/10.1177/1080569908325860

[123] Robert W. Frick. 1985. Communicating Emotion: The Role of Prosodic Features. *Psychological Bulletin* 97, 3 (May 1985), 412–429. https://doi.org/10.1037/0033-2909.97.3.412

[124] Chris D Frith. 2008. Social cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363, 1499 (June 2008), 2033–2039. https://doi.org/10.1098/rstb.2008.0005

[125] Chris D. Frith and Uta Frith. 2007. Social Cognition in Humans. *Current Biology* 17, 16 (Aug. 2007), R724–R732. https://doi.org/10.1016/j.cub.2007.05.068

[126] Desire Furnes, Hege Berg, Rachel M. Mitchell, and Silke Paulmann. 2019. Exploring the Effects of Personality Traits on the Perception of Emotions From Prosody. *Frontiers in Psychology* 10 (Feb. 2019), 184. https://doi.org/10.3389/fpsyg.2019.00184

[127] Sarah E. Gaither and Samuel R. Sommers. 2013. Living with an Other-Race Roommate Shapes Whites' Behavior in Subsequent Diverse Settings. *Journal of Experimental Social Psychology* 49, 2 (March 2013), 272–276. https://doi.org/10.1016/j.jesp.2012.10.020

[128] Andrew Gambino, Jesse Fox, and Rabindra Ratan. 2020. Building a stronger CASA: Extending the computers are social actors paradigm. *Human machine communication journal (Print)* 1 (2020), 71–85.

[129] Ge Gao, Naomi Yamashita, Ari MJ Hautasaari, Andy Echenique, and Susan R. Fussell. 2014. Effects of Public vs. Private Automated Transcripts on Multiparty Communication between Native and Non-Native English Speakers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 843–852. https://doi.org/10.1145/2556288.2557303

[130] Ge Gao, Naomi Yamashita, Ari M.J. Hautasaari, and Susan R. Fussell. 2015. Improving Multilingual Collaboration by Displaying How Non-

Native Speakers Use Automated Transcripts and Bilingual Dictionaries. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 3463–3472. https://doi.org/10.1145/2702123.2702498

[131] Rita Garcia, Chieh-Ju Liao, Ariane Pearce, and Christoph Treude. 2022. Gender Influence on Communication Initiated within Student Teams. In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education - Volume 1* (Providence, RI, USA) *(SIGCSE 2022)*. Association for Computing Machinery, New York, NY, USA, 432–438. https://doi.org/10.1145/3478431.3499279

[132] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. 2015. A Neural Algorithm of Artistic Style. (2015). https://doi.org/10.48550/ARXIV.1508.06576 Publisher: arXiv Version Number: 2.

[133] Andrzej T. Gałecki and Tomasz Burzykowski. 2013. *Linear Mixed-Effects Models Using R: A Step-by-Step Approach*. Springer, New York Heidelberg Dordrecht London.

[134] Harold B. Gerard and Michael F. Hoyt. 1974. Distinctiveness of Social Categorization and Attitude toward Ingroup Members. *Journal of Personality and Social Psychology* 29, 6 (June 1974), 836–842. https://doi.org/10.1037/h0036204

[135] Katy Ilonka Gero, Vivian Liu, and Lydia Chilton. 2022. Sparks: Inspiration for Science Writing Using Language Models. In *Designing Interactive Systems Conference* (Virtual Event, Australia) *(DIS '22)*. Association for Computing Machinery, New York, NY, USA, 1002–1019. https://doi.org/10.1145/3532106.3533533

[136] Supratip Ghose and Jagat Joyti Barua. 2013. Toward the implementation of a topic specific dialogue based natural language chatbot as an undergraduate advisor. In *2013 International Conference on Informatics, Electronics and Vision (ICIEV)*. 1–5. https://doi.org/10.1109/ICIEV.2013.6572650

[137] Howard Giles and Cynthia Gallois (Eds.). 2012. *The Handbook of Intergroup Communication*. Routledge, New York, NY.

[138] Gabriela Goldschmidt. 2014. *Linkography: Unfolding the Design Process*. MIT Press.

[139] Kate Grandprey-Shores, Yilin He, Kristina L. Swanenburg, Robert Kraut,

and John Riedl. 2014. The Identification of Deviance and Its Impact on Retention in a Multiplayer Game. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work Social Computing* (Baltimore, Maryland, USA) *(CSCW '14)*. Association for Computing Machinery, New York, NY, USA, 1356–1365. https://doi.org/10.1145/2531602.2531724

[140] Kurt Gray and Daniel M. Wegner. 2012. Feeling Robots and Human Zombies: Mind Perception and the Uncanny Valley. *Cognition* 125, 1 (2012), 125–130. https://doi.org/10.1016/j.cognition.2012.06.007

[141] Jennifer M. Grossman and Michelle V. Porche. 2014. Perceived Gender and Racial/Ethnic Barriers to STEM Success. *Urban Education* 49, 6 (Sept. 2014), 698–727. https://doi.org/10.1177/0042085913481364

[142] William B Gudykunst. 1993. Toward a Theory of Effective Interpersonal and Intergroup Communication: An Anxiety/Uncertainty Management (AUM) Perspective. *International and Intercultural Communication Annual* 17 (1993), 33.

[143] Sandra G. Hart and Lowell E. Staveland. 1988. *Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research*. Vol. 52. Elsevier, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

[144] Fritz Heider and Marianne Simmel. 1944. An Experimental Study of Apparent Behavior. *The American Journal of Psychology* 57, 2 (April 1944), 243. https://doi.org/10.2307/1416950

[145] Beth A Hennessey. 2003. Is the Social Psychology of Creativity Really Social. *Group Creativity* (2003), 181–201.

[146] Nicholas Hertz and Eva Wiese. 2017. Social Facilitation with Non-human Agents: Possible or Not?. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 61. SAGE Publications Sage CA: Los Angeles, CA, 222–225.

[147] E. Tory Higgins and Arie W. Kruglanski (Eds.). 1996. *Social Psychology: Handbook of Basic Principles*. Guilford Press, New York.

[148] Andrew Hodges and Douglas R. Hofstadter. 2014. *Alan Turing: the Enigma: the Book that Inspired the Film The Imitation Game*. Princeton University Press, Princeton, New Jersey.

[149] Michael A. Hogg, Deborah J. Terry, and Katherine M. White. 1995. A Tale of Two Theories: A Critical Comparison of Identity Theory with Social Identity Theory. *Social Psychology Quarterly* 58, 4 (Dec. 1995), 255.

https://doi.org/10.2307/2787127

[150] Jess Hohenstein and Malte Jung. 2018. AI-supported messaging: An investigation of human-human text conversation with AI support. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–6.

[151] Jess Hohenstein and Malte Jung. 2020. AI as a Moral Crumple Zone: The Effects of AI-Mediated Communication on Attribution and Trust. *Computers in Human Behavior* 106 (2020), 106190. Publisher: Elsevier.

[152] Aike C Horstmann and Nicole C Krämer. 2019. Great Expectations? Relation of Previous Experiences With Social Robots in Real Life or in the Media and Expectancies Based on Qualitative and Quantitative Assessment. *Frontiers in Psychology* 10 (2019), 939–939.

[153] Yoyo Tsung-Yu Hou and Malte F. Jung. 2021. Who is the Expert? Reconciling Algorithm Aversion and Algorithm Appreciation in AI-Supported Decision Making. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct. 2021), 1–25. https://doi.org/10.1145/3479864

[154] Ayanna Howard. 2020. Are We Trusting AI Too Much? Examining Human-Robot Interactions in the Real World. In *2020 15th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 1–1.

[155] John Howard. 2019. Artificial Intelligence: Implications for the Future of Work. *American Journal of Industrial Medicine* 62, 11 (Nov. 2019), 917–926. https://doi.org/10.1002/ajim.23037

[156] Crystal L Hoyt, Jim Blascovich, and Kimberly R Swinth. 2003. Social inhibition in immersive virtual environments. *Presence: Teleoperators & Virtual Environments* 12, 2 (2003), 183–195.

[157] Karen Huang, Michael Yeomans, Alison Wood Brooks, Julia Minson, and Francesca Gino. 2017. It Doesn't Hurt to Ask: Question-Asking Increases Liking. *Journal of Personality and Social Psychology* 113, 3 (Sept. 2017), 430–452. https://doi.org/10.1037/pspi0000097

[158] Lorraine Hudson, Clement Amponsah, Josephine Ohenewa Bampoe, Julie Marshall, Nana Akua Victoria Owusu, Khalid Hussein, Jess Linington, Zoe Banks Gross, Jane Stokes, and Róisín McNaney. 2020. Co-Designing Digital Tools to Enhance Speech and Language Therapy Training in Ghana. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13.

https://doi.org/10.1145/3313831.3376474

[159] David A. Huffaker and Sandra L. Calvert. 2006. Gender, Identity, and Language Use in Teenage Blogs. *Journal of Computer-Mediated Communication* 10, 2 (June 2006). https://doi.org/10.1111/j.1083-6101.2005.tb00238.x

[160] Pascal Huguet, Marie P. Galvaing, Jean M. Monteil, and Florence Dumas. 1999. Social Presence Effects in the Stroop Task: Further Evidence for an Attentional View of Social Facilitation. *Journal of Personality and Social Psychology* 77, 5 (1999), 1011–1025. https://doi.org/10.1037/0022-3514.77.5.1011

[161] Angel Hsing-Chi Hwang and Andrea Stevenson Won. 2021. IdeaBot: Investigating Social Facilitation in Human-Machine Team Creativity. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–16. https://doi.org/10.1145/3411764.3445270

[162] Gilhwan Hwang, Jeewon Lee, Cindy Yoonjung Oh, and Joonhwan Lee. 2019. It Sounds Like A Woman: Exploring Gender Stereotypes in South Korean Voice Assistants. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3312915

[163] Mandy Hütter and Michael Diehl. 2011. Motivation Losses in Teamwork: The Effects of Team Diversity and Equity Sensitivity on Reactions to Free-riding. *Group Processes Intergroup Relations* 14, 6 (Nov. 2011), 845–856. https://doi.org/10.1177/1368430211402405

[164] Rachael E. Jack and Philippe G. Schyns. 2015. The Human Face as a Dynamic Tool for Social Communication. *Current Biology* 25, 14 (July 2015), R621–R634. https://doi.org/10.1016/j.cub.2015.05.052

[165] Mohammad Hossein Jarrahi. 2018. Artificial Intelligence and the Future of Work: Human-AI Symbiosis in Organizational Decision Making. *Business Horizons* 61, 4 (July 2018), 577–586. https://doi.org/10.1016/j.bushor.2018.03.007

[166] Theodore Jensen, Mohammad Maifi Hasan Khan, Md Abdullah Al Fahim, and Yusuf Albayram. 2021. Trust and Anthropomorphism in Tandem: The Interrelated Nature of Automated Agent Appearance and Reliability in Trustworthiness Perceptions. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference* (Virtual Event, USA) *(DIS '21)*. Association for Computing Machinery, New York, NY, USA, 1470–1480. https://doi.org/10.1145/3461778.3462102

[167] Liwei Jiang, Jena D. Hwang, Chandra Bhagavatula, Ronan Le Bras, Jenny Liang, Jesse Dodge, Keisuke Sakaguchi, Maxwell Forbes, Jon Borchardt, Saadia Gabriel, Yulia Tsvetkov, Oren Etzioni, Maarten Sap, Regina Rini, and Yejin Choi. 2022. Can Machines Learn Morality? The Delphi Experiment. (July 2022). https://doi.org/10.48550/arXiv.2110.07574 arXiv:2110.07574 [cs].

[168] Susan C. Johnson. 2003. Detecting Agents. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 358, 1431 (March 2003), 549–559. https://doi.org/10.1098/rstb.2002.1237

[169] Malte F Jung, Nikolas Martelaro, and Pamela J Hinds. 2015. Using robots to moderate team conflict: the case of repairing violations. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. 229–236.

[170] Katarzyna Kabacińska, Tony J. Prescott, and Julie M. Robillard. 2021. Socially Assistive Robots as Mental Health Interventions for Children: A Scoping Review. *International Journal of Social Robotics* 13, 5 (Aug. 2021), 919–935. https://doi.org/10.1007/s12369-020-00679-0

[171] Knut K. W. Kampe, Chris D. Frith, Raymond J. Dolan, and Uta Frith. 2001. Reward Value of Attractiveness and Gaze. *Nature* 413, 6856 (Oct. 2001), 589–589. https://doi.org/10.1038/35098149

[172] Rosabeth Moss Kanter. 1977. Some Effects of Proportions on Group Life: Skewed Sex Ratios and Responses to Token Women. *Amer. J. Sociology* 82, 5 (March 1977), 965–990. https://doi.org/10.1086/226425

[173] Rosabeth Moss Kanter. 2010. *Men and Women of the Corporation*. Basic Books, New York, NY.

[174] James C Kaufman and Robert J Sternberg. 2010. *The Cambridge handbook of creativity*. Cambridge University Press.

[175] Caroline F. Keating. 2002. *Facial Attractiveness: Evolutionary, Cognitive, and Social Perspectives*. Ablex Publishing, Chapter Charismatic Faces: Social Status Cues Put Face Appeal in Context, 153–192.

[176] Caroline F. Keating and James Doyle. 2002. The Faces of Desirable Mates and Dates Contain Mixed Social Status Cues. *Journal of Experimental Social Psychology* 38, 4 (July 2002), 414–424. https://doi.org/10.1016/S0022-1031(02)00007-0

[177] Alice Kerlyl, Phil Hall, and Susan Bull. [n. d.]. Bringing Chatbots into ed-

ucation: Towards Natural Language Negotiation of Open Learner Models. In *Applications and Innovations in Intelligent Systems XIV*, Richard Ellis, Tony Allen, and Andrew Tuson (Eds.). Springer London, 179–192. https://doi.org/10.1007/978-1-84628-666-7$_1$4

[178] Pranav Khadpe, Ranjay Krishna, Li Fei-Fei, Jeffrey T. Hancock, and Michael S. Bernstein. 2020. Conceptual Metaphors Impact Perceptions of Human-AI Collaboration. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW2, Article 163 (Oct. 2020), 26 pages. https://doi.org/10.1145/3415234

[179] Zahra Rezaei Khavas, S Reza Ahmadzadeh, and Paul Robinette. 2020. Modeling Trust in Human-Robot Interaction: A Survey. In *International Conference on Social Robotics*. Springer, 529–541.

[180] Sara Kiesler, Aaron Powers, Susan R Fussell, and Cristen Torrey. 2008. Anthropomorphic Interactions with a Robot and Robot-like Agent. *Social Cognition* 26, 2 (2008), 169–181.

[181] Emma J. Kilford, Emily Garrett, and Sarah-Jayne Blakemore. 2016. The Development of Social Cognition in Adolescence: An Integrated Perspective. *Neuroscience Biobehavioral Reviews* 70 (Nov. 2016), 106–120. https://doi.org/10.1016/j.neubiorev.2016.08.016

[182] Lawrence H. Kim, Daniel S. Drew, Veronika Domova, and Sean Follmer. 2020. User-Defined Swarm Robot Control. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376814

[183] Youjeong Kim and S. Shyam Sundar. 2012. Anthropomorphism of Computers: Is It Mindful or Mindless? *Computers in Human Behavior* 28, 1 (Jan. 2012), 241–250. https://doi.org/10.1016/j.chb.2011.09.006

[184] Ambika Kirkland, Harm Lameris, Eva Szekely, and Joakim Gustafson. 2022. Where's the Uh, Hesitation? The Interplay Between Filled Pause Location, Speech Rate and Fundamental Frequency in Perception of Confidence. In *Interspeech 2022*. ISCA, 4990–4994. https://doi.org/10.21437/Interspeech.2022-10973

[185] Ute-Christine Klehe, Neil Anderson, and Esther A. Hoefnagels. 2007. Social Facilitation and Inhibition During Maximum Versus Typical Performance Situations. *Human Performance* 20, 3 (May 2007), 223–239. https://doi.org/10.1080/08959280701333040

[186] Gary Klein, David D. Woods, Jeffrey M. Bradshaw, Robert R. Hoffman,

and Paul J. Feltovich. 2004. Ten Challenges for Making Automation a "Team Player" in Joint Human-Agent Activity. 19, 6 (2004), 91–95. https://doi.org/10.1109/MIS.2004.74

[187] Tua Korhonen. 2019. *Anthropomorphism and the Aesopic Animal Fables*. Springer Fachmedien Wiesbaden, Wiesbaden, 211–231. https://doi.org/10.1007/978-3-658-24388-3$_1$0

[188] Steve W. J. Kozlowski and Daniel R. Ilgen. 2006. Enhancing the Effectiveness of Work Groups and Teams. 7, 3 (2006), 77–124. https://doi.org/10.1111/j.1529-1006.2006.00030.x

[189] Robert M Krauss, Robin Freyberg, and Ezequiel Morsella. 2002. Inferring Speakers' Physical Attributes from Their Voices. *Journal of Experimental Social Psychology* 38, 6 (Nov. 2002), 618–625. https://doi.org/10.1016/S0022-1031(02)00510-3

[190] Alap Kshirsagar, Bnaya Dreyfuss, Guy Ishai, Ori Heffetz, and Guy Hoffman. 2019. Monetary-Incentive Competition between Humans and Robots: Experimental Results. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 95–103.

[191] Robert Kurzban, John Tooby, and Leda Cosmides. 2001. Can Race be Erased? Coalitional Computation and Social Categorization. *Proceedings of the National Academy of Sciences* 98, 26 (Dec. 2001), 15387–15392. https://doi.org/10.1073/pnas.251541498

[192] Marek Kwiek and Wojciech Roszka. 2021. Gender-based Homophily in Research: A Large-scale Study of Man-Woman Collaboration. *Journal of Informetrics* 15, 3 (Aug. 2021), 101171. https://doi.org/10.1016/j.joi.2021.101171

[193] Christos Kyrlitsias and Despina Michael. 2016. Influence by Others' Opinions: Social Pressure from Agents in Immersive Virtual Environments. In *2016 IEEE Virtual Reality (VR)*. IEEE, 213–214.

[194] Vivian Lai, Chacha Chen, Alison Smith-Renner, Q. Vera Liao, and Chenhao Tan. 2023. Towards a Science of Human-AI Decision Making: An Overview of Design Space in Empirical Human-Subject Studies. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (Chicago, IL, USA) *(FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 1369–1385. https://doi.org/10.1145/3593013.3594087

[195] Yi Lai, Atreyi Kankanhalli, and Desmond Ong. 2021. *Human-*

AI Collaboration in Healthcare: A Review and Research Agenda. http://hdl.handle.net/10125/70657

[196] Tarun Lalwani, Shashank Bhalotia, Ashish Pal, Vasundhara Rathod, and Shreya Bisen. [n. d.]. Implementation of a Chatbot System using AI and NLP. ([n. d.]). https://doi.org/10.2139/ssrn.3531782

[197] John D. Lee and Katrina A. See. 2004. Trust in Automation: Designing for Appropriate Reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 46, 1 (2004), 50–80. https://doi.org/10.1518/hfes.46.1.50$_3$0392

[198] Sanguk Lee, Rabindra Ratan, and Taiwoo Park. 2019. The Voice Makes the Car: Enhancing Autonomous Vehicle Perceptions and Adoption Intention through Voice Agent Gender and Style. *Multimodal Technologies and Interaction* 3, 1 (March 2019), 20. https://doi.org/10.3390/mti3010020

[199] Gilly Leshed, Jeffrey T. Hancock, Dan Cosley, Poppy L. McLeod, and Geri Gay. 2007. Feedback for Guiding Reflection on Teamwork Practices. In *Proceedings of the 2007 ACM International Conference on Supporting Group Work* (Sanibel Island, Florida, USA) *(GROUP '07)*. Association for Computing Machinery, New York, NY, USA, 217–220. https://doi.org/10.1145/1316624.1316655

[200] Gilly Leshed, Diego Perez, Jeffrey T. Hancock, Dan Cosley, Jeremy Birnholtz, Soyoung Lee, Poppy L. McLeod, and Geri Gay. 2009. Visualizing Real-Time Language-Based Feedback on Teamwork Behavior in Computer-Mediated Groups. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 537–546. https://doi.org/10.1145/1518701.1518784

[201] Wai Sze Leung. 2017. Bad Blood: Managing Toxic Relationships through Belbin Roles for First Year Software Engineering Students. In *Proceedings of the 3rd International Conference on Communication and Information Processing* (Tokyo, Japan) *(ICCIP '17)*. Association for Computing Machinery, New York, NY, USA, 82–86. https://doi.org/10.1145/3162957.3163010

[202] Karen Levy. 2022. RoboTruckers: The Double Threat of AI for Low-Wage Work. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society* (Oxford, United Kingdom) *(AIES '22)*. Association for Computing Machinery, New York, NY, USA, 3. https://doi.org/10.1145/3514094.3539569

[203] Michael Lewis, Katia Sycara, and Phillip Walker. 2018. The Role of Trust

in Human-Robot Interaction. In *Foundations of trusted autonomy*. Springer, Cham, 135–159.

[204] Xinge Li and Yongjun Sung. 2021. Anthropomorphism brings us closer: The mediating role of psychological distance in User–AI assistant interactions. *Computers in Human Behavior* 118 (May 2021), 106680. https://doi.org/10.1016/j.chb.2021.106680

[205] Xiaoyan Li, Naomi Yamashita, Wen Duan, Yoshinari Shirai, and Susan R. Fussell. 2022. Improving Non-Native Speakers' Participation with an Automatic Agent in Multilingual Groups. *Proceedings of the ACM on Human-Computer Interaction (PACM HCI)* 7, GROUP, Article 12 (Dec. 2022), 28 pages. https://doi.org/10.1145/3567562

[206] Claire Liang, Julia Proft, Erik Andersen, and Ross A. Knepper. 2019. Implicit Communication of Actionable Information in Human-AI teams. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow Scotland Uk). ACM, 1–13. https://doi.org/10.1145/3290605.3300325

[207] Mengqi Liao and S. Shyam Sundar. 2021. How Should AI Systems Talk to Users when Collecting their Personal Information? Effects of Role Framing and Self-Referencing on Human-AI Interaction. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Yokohama Japan, 1–14. https://doi.org/10.1145/3411764.3445415

[208] P Light, K Littleton, S Bale, R Joiner, and D Messer. 2000. Gender and Social Comparison Effects in Computer-Based Problem Solving. *Learning and Instruction* 10, 6 (Dec. 2000), 483–496. https://doi.org/10.1016/S0959-4752(00)00010-4

[209] Maria R. Lima, Maitreyee Wairagkar, Nirupama Natarajan, Sridhar Vaitheswaran, and Ravi Vaidyanathan. 2021. Robotic Telemedicine for Mental Health: A Multimodal Approach to Improve Human-Robot Engagement. *Frontiers in Robotics and AI* 8 (March 2021), 618866. https://doi.org/10.3389/frobt.2021.618866

[210] Bingjie Liu. 2021. In AI We Trust? Effects of Agency Locus and Transparency on Uncertainty Reduction in Human–AI Interaction. *Journal of Computer-Mediated Communication* 26, 6 (Nov. 2021), 384–402. https://doi.org/10.1093/jcmc/zmab013

[211] Vivian Liu, Han Qiao, and Lydia Chilton. 2022. Opal: Multimodal Image Generation for News Illustration. In *Proceedings of the 35th Annual*

ACM Symposium on User Interface Software and Technology (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 73, 17 pages. https://doi.org/10.1145/3526113.3545621

[212] Patricia L. Lockwood, Matthew A.J. Apps, and Steve W.C. Chang. 2020. Is There a 'Social' Brain? Implementations and Algorithms. *Trends in Cognitive Sciences* 24, 10 (Oct. 2020), 802–813. https://doi.org/10.1016/j.tics.2020.06.011

[213] Jay William Lorsch (Ed.). 1987. *Handbook of organizational behavior*. Prentice-Hall.

[214] Max M. Louwerse, Arthur C. Graesser, Shulan Lu, and Heather H. Mitchell. 2005. Social Cues in Animated Conversational Agents. *Applied Cognitive Psychology* 19, 6 (Sept. 2005), 693–704. https://doi.org/10.1002/acp.1117

[215] Yuyan Luo and Renée Baillargeon. 2005. Can a Self-Propelled Box Have a Goal?: Psychological Reasoning in 5-Month-Old Infants. *Psychological Science* 16, 8 (2005), 601–608. https://doi.org/10.1111/j.1467-9280.2005.01582.x arXiv:https://doi.org/10.1111/j.1467-9280.2005.01582.x PMID: 16102062.

[216] Bertram F Malle and Daniel Ullman. 2021. A Multidimensional Conception and Measure of Human-Robot Trust. In *Trust in Human-Robot Interaction*. Elsevier, 3–25.

[217] Marianne Schmid Mast. 2001. Gender Differences and Similarities in Dominance Hierarchies in Same-Gender Groups Based on Speaking Time. *Sex Roles* 44, 9 (May 2001), 537–556. https://doi.org/10.1023/A:1012239024732

[218] Jasmine K. McBeath, Richard P. Durán, and Danielle B. Harlow. 2017. Not My Gumdrop Buttons! Youth Tool Use in Designing an Electronic Shrek-Themed Bean Bag Toss. In *Proceedings of the 2017 Conference on Interaction Design and Children* (Stanford, California, USA) *(IDC '17)*. Association for Computing Machinery, New York, NY, USA, 61–72. https://doi.org/10.1145/3078072.3079721

[219] Nathan J. McNeese, Mustafa Demir, Erin K. Chiou, and Nancy J. Cooke. 2021. Trust and Team Performance in Human–Autonomy Teaming. *International Journal of Electronic Commerce* 25, 1 (Jan. 2021), 51–72. https://doi.org/10.1080/10864415.2021.1846854

[220] Nathan J. McNeese, Mustafa Demir, Nancy J. Cooke, and Christo-

pher Myers. 2018. Teaming With a Synthetic Teammate: Insights into Human-Autonomy Teaming. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 60, 2 (2018), 262–273. https://doi.org/10.1177/0018720817743223

[221] Miller McPherson, Lynn Smith-Lovin, and James M Cook. 2001. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology* 27, 1 (Aug. 2001), 415–444. https://doi.org/10.1146/annurev.soc.27.1.415

[222] Beverly Metcalfe and Alison Linstead. 2003. Gendering Teamwork: Re-Writing the Feminine. *Gender, Work Organization* 10, 1 (Jan. 2003), 94–119. https://doi.org/10.1111/1468-0432.00005

[223] Katherine L. Milkman, Modupe Akinola, and Dolly Chugh. 2015. What Happens Before? A Field Experiment Exploring How Pay and Representation Differentially Shape Bias on the Pathway into Organizations. *Journal of Applied Psychology* 100, 6 (2015), 1678–1712. https://doi.org/10.1037/apl0000022

[224] Frances J. Milliken and Luis L. Martins. 1996. Searching for Common Threads: Understanding the Multiple Effects of Diversity in Organizational Groups. *The Academy of Management Review* 21, 2 (April 1996), 402. https://doi.org/10.2307/258667

[225] Marvin Minsky. 2007. *The Emotion Machine: Commensense Thinking, Artificial Intelligence, and the Future of the Human Mind*. Simon Schuster, New York.

[226] Maali Mnasri. 2019. Recent advances in conversational NLP: Towards the standardization of Chatbot building. (2019). https://doi.org/10.48550/ARXIV.1903.09025

[227] Kelly A. Mollica, Barbara Gray, and Linda K. Treviño. 2003. Racial Homophily and Its Persistence in Newcomers' Social Networks. *Organization Science* 14, 2 (April 2003), 123–136. https://doi.org/10.1287/orsc.14.2.123.14994

[228] Catherine J. Mondloch, Daphne Maurer, and Sara Ahola. 2006. Becoming a Face Expert. *Psychological Science* 17, 11 (Nov. 2006), 930–934. https://doi.org/10.1111/j.1467-9280.2006.01806.x

[229] Masahiro Mori, Karl F. MacDorman, and Norri Kageki. 2012. The Uncanny Valley [From the Field]. *IEEE Robotics Automation Magazine* 19, 2 (2012), 98–100. https://doi.org/10.1109/MRA.2012.2192811

[230] Elizabeth Wolef Morrison and Frances J Milliken. 2000. Organizational silence: A barrier to change and development in a pluralistic world. *Academy of Management review* 25, 4 (2000), 706–725.

[231] Jonathan Mumm and Bilge Mutlu. 2011. Designing Motivational Agents: The Role of Praise, Social Comparison, and Embodiment in Computer Feedback. *Computers in Human Behavior* 27, 5 (Sept. 2011), 1643–1650. https://doi.org/10.1016/j.chb.2011.02.002

[232] Thomas Mussweiler. 2003. Comparison Processes in Social Judgment: Mechanisms and Consequences. *Psychological Review* 110, 3 (2003), 472–489. https://doi.org/10.1037/0033-295X.110.3.472

[233] Thomas Mussweiler and Fritz Strack. 2000. Consequences of Social Comparison. In *Handbook of Social Comparison*, Jerry Suls and Ladd Wheeler (Eds.). Springer US, Boston, MA, 253–270. https://doi.org/10.1007/978-1-4615-4237-7$_1$3

[234] Larissa Myaskovsky, Emily Unikel, and Mary Amanda Dew. 2005. Effects of Gender Diversity on Performance and Interpersonal Behavior in Small Work Groups. *Sex Roles* 52, 9–10 (May 2005), 645–657. https://doi.org/10.1007/s11199-005-3732-8

[235] Clifford Nass, B.J. Fogg, and Youngme Moon. 1996. Can Computers be Teammates? *International Journal of Human-Computer Studies* 45, 6 (Dec. 1996), 669–678. https://doi.org/10.1006/ijhc.1996.0073

[236] Clifford Nass and Youngme Moon. 2000. Machines and Mindlessness: Social Responses to Computers. 56, 1 (2000), 81–103. https://doi.org/10.1111/0022-4537.00153

[237] Clifford Nass, Jonathan Steuer, and Ellen R Tauber. 1994. Computers Are Social Actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 72–78.

[238] Clifford Ivar Nass and Scott Brave. 2005. *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*. MIT Press, Cambridge, Mass.

[239] Manisha Natarajan and Matthew Gombolay. 2020. Effects of Anthropomorphism and Accountability on Trust in Human Robot Interaction. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (Cambridge, United Kingdom) *(HRI '20)*. Association for Computing Machinery, New York, NY, USA, 33–42. https://doi.org/10.1145/3319502.3374839

[240] Charlan Jeanne Nemeth and Alexander O'Connor. 2019. Better than Individuals? Dissent and Group Creativity. In *The Oxford Handbook of Group Creativity and Innovation*, P. B. Paulus and B. A. Nijstad (Eds.). Oxford University Press, 73–83.

[241] Harold W. Noonan. 2010. The Thinking Animal Problem and Personal Pronoun Revisionism. *Analysis* 70, 1 (Jan. 2010), 93–98. https://doi.org/10.1093/analys/anp137

[242] Changhoon Oh, Jungwoo Song, Jinhan Choi, Seonghyeon Kim, Sungwoo Lee, and Bongwon Suh. 2018. I Lead, You Help but Only with Enough Details: Understanding User Experience of Co-Creation with Artificial Intelligence. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Montreal QC Canada, 1–13. https://doi.org/10.1145/3173574.3174223

[243] Kazuo Okamura and Seiji Yamada. 2020. Adaptive Trust Calibration for Human-AI Collaboration. *PLOS ONE* 15, 2 (Feb. 2020), e0229132. https://doi.org/10.1371/journal.pone.0229132

[244] Eric T. Olson. 2002. Thinking Animals and the Reference of "I". *Philosophical Topics* 30, 1 (2002), 189–207. https://www.jstor.org/stable/43154385

[245] Lisa Orii, Diana Tosca, Andrew L Kun, and Orit Shaer. 2021. Perceptions on the Future of Automation in r/Truckers. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 311, 6 pages. https://doi.org/10.1145//3411763.3451637

[246] E. Marlies Ott. 1989. Effects of the Male-Female Ratio at Work: Policewomen and Male Nurses. *Psychology of Women Quarterly* 13, 1 (March 1989), 41–57. https://doi.org/10.1111/j.1471-6402.1989.tb00984.x

[247] Thomas O'Neill, Nathan McNeese, Amy Barron, and Beau Schelble. 2022. Human–Autonomy Teaming: A Review and Analysis of the Empirical Literature. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 64, 5 (2022), 904–938. https://doi.org/10.1177/0018720820960865

[248] Stefan Palan and Christian Schitter. 2018. Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance* 17 (March 2018), 22–27. https://doi.org/10.1016/j.jbef.2017.12.004

[249] Nicholas A. Palomares. 2009. Women Are Sort of More Tentative Than Men, Aren't They?: How Men and Women Use Tentative Lan-

guage Differently, Similarly, and Counterstereotypically as a Function of Gender Salience. *Communication Research* 36, 4 (Aug. 2009), 538–560. https://doi.org/10.1177/0093650209333034

[250] Sung Park and Richard Catrambone. 2007. Social Facilitation Effects of Virtual Humans. *Human Factors* 49, 6 (2007), 1054–1060.

[251] Sun Young Park, Pei-Yi Kuo, Andrea Barbarin, Elizabeth Kaziunas, Astrid Chow, Karandeep Singh, Lauren Wilcox, and Walter S. Lasecki. 2019. Identifying Challenges and Opportunities in Human-AI Collaboration in Healthcare. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing* (Austin, TX, USA) *(CSCW '19)*. Association for Computing Machinery, New York, NY, USA, 506–510. https://doi.org/10.1145/3311957.3359433

[252] Urja Pawar, Donna O'Shea, Susan Rea, and Ruairi O'Reilly. 2020. Explainable AI in Healthcare. In *2020 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA)*. IEEE, Dublin, Ireland, 1–2. https://doi.org/10.1109/CyberSA49311.2020.9139655

[253] Hannah R. M. Pelikan and Malte F. Jung. 2023. Designing Robot Sound-In-Interaction: The Case of Autonomous Public Transport Shuttle Buses. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction* (Stockholm, Sweden) *(HRI '23)*. Association for Computing Machinery, New York, NY, USA, 172–182. https://doi.org/10.1145/3568162.3576979

[254] Kevin A. Pelphrey and Elizabeth J. Carter. 2008. Charting the Typical and Atypical Development of the Social Brain. *Development and Psychopathology* 20, 4 (2008), 1081–1102. https://doi.org/10.1017/S0954579408000515

[255] Tamyra Pierce. 2009. Social Anxiety and Technology: Face-to-face Communication versus Technological Communication among Teens. *Computers in Human Behavior* 25, 6 (2009), 1367–1372.

[256] Aaron Powers, Sara Kiesler, Susan Fussell, and Cristen Torrey. 2007. Comparing a Computer Agent with a Humanoid Robot. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction*. 145–152.

[257] David Premack. 1990. The Infant's Theory of Self-propelled Objects. *Cognition* 36, 1 (1990), 1–16. https://doi.org/10.1016/0010-0277(90)90051-K

[258] Jesse J Prinz. 2004. *Gut Reactions: A Perceptual Theory of Emotion*. Oxford University Press.

[259] Han Qiao, Vivian Liu, and Lydia Chilton. 2022. Initial Images: Using Image Prompts to Improve Subject Representation in Multimodal AI Generated Art. In *Creativity and Cognition* (Venice, Italy) *(C&C '22)*. Association for Computing Machinery, New York, NY, USA, 15–28. https://doi.org/10.1145/3527927.3532792

[260] Lingyun Qiu and Izak Benbasat. 2005. An Investigation into the Effects of Text-To-Speech Voice and 3D Avatars on the Perception of Presence and Flow of Live Help in Electronic Commerce. *ACM Trans. Comput.-Hum. Interact.* 12, 4 (Dec. 2005), 329–355. https://doi.org/10.1145/1121112.1121113

[261] R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

[262] Sonja Radoš, Marija Zdraveva, and Iris Žeželj. 2019. Status Dynamics in the Classroom: Roma Children's Implicit and Explicit Preference for Majority Children across Age Groups. *Journal of Cross-Cultural Psychology* 50, 4 (May 2019), 577–593. https://doi.org/10.1177/0022022119828498

[263] A M Rahman, Abdullah Al Mamun, and Alma Islam. 2017. Programming challenges of chatbot: Current and future prospective. In *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*. 75–78. https://doi.org/10.1109/R10-HTC.2017.8288910

[264] Vilayanur S Ramachandran and Baland Jalal. 2017. The Evolutionary Psychology of Envy and Jealousy. *Frontiers in Psychology* 8 (2017), 1619. Publisher: Frontiers.

[265] Byron Reeves and Clifford Ivar Nass. 1996. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Center for the Study of Language and Information; Cambridge University Press, Stanford, California.

[266] Forbes Insights Reports. 2022. *Forbes Magazine* (2022). https://doi.org/connect/forbes-insights-reports/

[267] Gillian Rhodes and Leslie A. Zebrowitz. 2002. *Facial Attractiveness: Evolutionary, Cognitive, and Social Perspectives*. Ablex, Westport, Connecticut.

[268] Raoul Rickenberg and Byron Reeves. 2000. The Effects of Animated Characters on Anxiety, Task Performance, and Evaluations of User Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 49–56.

[269] Nina Riether, Frank Hegel, Britta Wrede, and Gernot Horstmann. 2012. Social Facilitation with Social Robots?. In *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 41–47.

[270] Ashleigh Shelby Rosette, Geoffrey J. Leonardelli, and Katherine W. Phillips. 2008. The White Standard: Racial Bias in Leader Categorization. *Journal of Applied Psychology* 93, 4 (July 2008), 758–777. https://doi.org/10.1037/0021-9010.93.4.758

[271] Janet B. Ruscher. 2017. Prejudiced Communication. In *Oxford Research Encyclopedia of Communication*. Oxford University Press. https://doi.org/10.1093/acrefore/9780190228613.013.419

[272] Paul R. Sackett, Cathy L. DuBois, and Ann W. Noe. 1991. Tokenism in Performance Evaluation: The Effects of Work Group Representation on Male-Female and White-Black Differences in Performance Ratings. *Journal of Applied Psychology* 76, 2 (April 1991), 263–267. https://doi.org/10.1037/0021-9010.76.2.263

[273] Glenn S Sanders, Robert Steven Baron, and Danny L Moore. 1978. Distraction and Social Comparison as Mediators of Social Facilitation Effects. *Journal of Experimental Social Psychology* 14, 3 (1978), 291–303.

[274] R. Keith Sawyer. 2012. *Explaining Creativity: The Science of Human Innovation* (2nd ed.). Oxford University Press, New York, New York.

[275] Brian J Scholl and Patrice D Tremoulet. 2000. Perceptual Causality and Animacy. *Trends in Cognitive Sciences* 4, 8 (Aug. 2000), 299–309. https://doi.org/10.1016/S1364-6613(00)01506-0

[276] Matthias Schurz, Joaquim Radua, Matthias G. Tholen, Lara Maliske, Daniel S. Margulies, Rogier B. Mars, Jerome Sallet, and Philipp Kanske. 2021. Toward a Hierarchical Model of Social Cognition: A Neuroimaging Meta-analysis and Integrative Review of Empathy and Theory of Mind. *Psychological Bulletin* 147, 3 (March 2021), 293–327. https://doi.org/10.1037/bul0000303

[277] Arielle Aj Scoglio, Erin D Reilly, Jay A Gorman, and Charles E Drebing. 2019. Use of Social Robots in Mental Health and Well-Being Research: Systematic Review. *Journal of Medical Internet Research* 21, 7 (July 2019), e13322. https://doi.org/10.2196/13322

[278] Roger Scruton. 2017. *On Human Nature*. Princeton University Press, Princeton.

[279] Sarah Sebo, Brett Stoll, Brian Scassellati, and Malte F. Jung. 2020. Robots in Groups and Teams: A Literature Review. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (Oct. 2020), 1–36. https://doi.org/10.1145/3415247

[280] Keisuke Seta, Masanori Yokoyama, Shigeo Yoshida, Takuji Narumi, Tomohiro Tanikawa, and Michitaka Hirose. 2018. Divided Presence: Improving Group Decision-Making via Pseudo-Population Increase. In *Proceedings of the 6th International Conference on Human-Agent Interaction* (Southampton, United Kingdom) *(HAI '18)*. Association for Computing Machinery, New York, NY, USA, 260–268. https://doi.org/10.1145/3284432.3284443

[281] William Seymour and Max Van Kleek. 2021. Exploring Interactions Between Trust, Anthropomorphism, and Relationship Development in Voice Assistants. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 371 (Oct. 2021), 16 pages. https://doi.org/10.1145/3479515

[282] Ameneh Shamekhi, Q Vera Liao, Dakuo Wang, Rachel KE Bellamy, and Thomas Erickson. 2018. Face Value? Exploring the Effects of Embodiment for a Group Facilitation Agent. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.

[283] Donghee Shin. 2021. The Effects of Explainability and Causability on Perception, Trust, and Acceptance: Implications for Explainable AI. *International Journal of Human-Computer Studies* 146 (2021), 102551. Publisher: Elsevier.

[284] Gabriel Skantze, Martin Johansson, and Jonas Beskow. 2015. Exploring Turn-Taking Cues in Multi-Party Human-Robot Discussions about Objects. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction* (Seattle, Washington, USA) *(ICMI '15)*. Association for Computing Machinery, New York, NY, USA, 67–74. https://doi.org/10.1145/2818346.2820749

[285] Vicki L. Smith and Herbert H. Clark. 1993. On the Course of Answering Questions. *Journal of Memory and Language* 32, 1 (Feb. 1993), 25–38. https://doi.org/10.1006/jmla.1993.1002

[286] Hai Son Nguyen, Andreas Mladenow, Christine Strauss, and Katharina Auer-Srnka. 2022. Voice Commerce: Anthropomorphism Using Voice Assistants. In *The 23rd International Conference on Information Integration and Web Intelligence* (Linz, Austria) *(iiWAS2021)*. Association for Computing Machinery, New York, NY, USA, 434–442.

https://doi.org/10.1145/3487664.3487724

[287] Yao Song and Yan Luximon. 2020. Trust in AI Agent: A Systematic Review of Facial Anthropomorphic Trustworthiness for Social Robot Design. *Sensors* 20, 18 (2020). https://doi.org/10.3390/s20185087

[288] Sabine Sonnentag and Judith Volmer. 2009. Individual-Level Predictors of Task-Related Teamwork Processes: The Role of Expertise and Self-Efficacy in Team Meetings. *Group Organization Management* 34, 1 (2009), 37–66. Publisher: Sage Publications Sage CA: Los Angeles, CA.

[289] Konrad Sowa, Aleksandra Przegalinska, and Leon Ciechanowski. 2021. Cobots in Knowledge Work. *Journal of Business Research* 125 (March 2021), 135–142. https://doi.org/10.1016/j.jbusres.2020.11.038

[290] Kenneth W Spence, IE Farber, and HH McFann. 1956. The Relation of Anxiety (Drive) Level to Performance in Competitional and Noncompetitional Paired-associates Learning. *Journal of Experimental Psychology* 52, 5 (1956), 296.

[291] Anna Stenzel, Eris Chinellato, Maria A. Tirado Bou, Ángel P. del Pobil, Markus Lappe, and Roman Liepelt. 2012. When humanoid robots become human-like interaction partners: Corepresentation of robotic actions. *Journal of Experimental Psychology: Human Perception and Performance* 38, 5 (2012), 1073–1077. https://doi.org/10.1037/a0029493

[292] Timothy D Stephen and Teresa M Harrison. 1986. Assessing Communication Style: A New Measure. *American Journal of Family Therapy* 14, 3 (1986), 213–234.

[293] Julia Stern, Christoph Schild, Benedict C. Jones, Lisa M. DeBruine, Amanda Hahn, David A. Puts, Ingo Zettler, Tobias L. Kordsmeyer, David Feinberg, Dan Zamfir, Lars Penke, and Ruben C. Arslan. 2021. Do Voices Carry Valid Information about a Speaker's Personality? *Journal of Research in Personality* 92 (June 2021), 104092. https://doi.org/10.1016/j.jrp.2021.104092

[294] Robert J. Sternberg. 1988. *The Nature of Creativity : Contemporary Psychological Perspectives*. Cambridge University Press, Cambridge ; New York.

[295] Brett Stoll, Malte F Jung, and Susan R Fussell. 2018. Keeping it light: perceptions of humor styles in robot-mediated conflict. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. 247–248.

[296] Terrell L. Strayhorn. 2010. Work in Progress - Social Barriers and Supports to Underrepresented Minorities' Success in STEM Fields. In *2010 IEEE Frontiers in Education Conference (FIE)*. IEEE, Arlington, VA, USA, S1H–1–S1H–5. https://doi.org/10.1109/FIE.2010.5673227

[297] S Subhash, Prajwal N Srivatsa, S Siddesh, A Ullas, and B Santhosh. 2020. Artificial Intelligence-based Voice Assistant. In *2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)*. 593–596. https://doi.org/10.1109/WorldS450073.2020.9210344

[298] Minhyang Suh, Emily Youngblom, Michael Terry, and Carrie J Cai. 2021. AI as Social Glue: Uncovering the Roles of Deep Generative AI during Social Music Composition. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–11.

[299] W.G. Sumner. 1906. *Folkways: A Study of the Sociological Importance of Usages, Manners, Customs, Mores, and Morals*. Ginn.

[300] Tingyu Sun. 2022. Research on the Influence of Artificial Intelligence on Female Labor Employment. In *2021 3rd International Conference on Artificial Intelligence and Advanced Manufacture* (Manchester, United Kingdom) *(AIAM2021)*. Association for Computing Machinery, New York, NY, USA, 2805–2809. https://doi.org/10.1145/3495018.3501186

[301] Yilu Sun, Omar Shaikh, and Andrea Stevenson Won. 2019. Nonverbal synchrony in virtual reality. *PLOS ONE* 14, 9 (Sept. 2019), e0221803. https://doi.org/10.1371/journal.pone.0221803

[302] S Shyam Sundar. 2020. Rise of Machine Agency: A Framework for Studying the Psychology of Human–AI Interaction (HAII). *Journal of Computer-Mediated Communication* 25, 1 (March 2020), 74–88. https://doi.org/10.1093/jcmc/zmz026

[303] S. Shyam Sundar and Clifford Nass. 2000. Source Orientation in Human-Computer Interaction: Programmer, Networker, or Independent Social Actor. 27, 6 (2000), 683–703. https://doi.org/10.1177/009365000027006001

[304] Deniz Susar and Vincenzo Aquaro. 2019. Artificial Intelligence: Opportunities and Challenges for the Public Sector. In *Proceedings of the 12th International Conference on Theory and Practice of Electronic Governance* (Melbourne, VIC, Australia) *(ICEGOV '19)*. Association for Computing Machinery, New York, NY, USA, 418–426. https://doi.org/10.1145/3326365.3326420

[305] Selina Jeanne Sutton. 2020. Gender Ambiguous, Not Genderless: Designing Gender in Voice User Interfaces (VUIs) with Sensitivity. In *Proceedings of the 2nd Conference on Conversational User Interfaces* (Bilbao, Spain) *(CUI '20)*. Association for Computing Machinery, New York, NY, USA, Article 11, 8 pages. https://doi.org/10.1145/3405755.3406123

[306] Ekaterina Svikhnushina, Iuliana Voinea, Anuradha Welivita, and Pearl Pu. 2022. A Taxonomy of Empathetic Questions in Social Dialogs. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Dublin, Ireland, 2952–2973. https://doi.org/10.18653/v1/2022.acl-long.211

[307] Robert W. Swezey and Eduardo Salas (Eds.). 1992. *Teams: Their Training and Performance*. Ablex Pub. Corp.

[308] Barbara G Tabachnick, Linda S Fidell, and Jodie B Ullman. 2007. *Using Multivariate Statistics*. Vol. 5. Pearson Boston, MA.

[309] Henri Tajfel. 1978. *Differentiation between Social Groups: Studies in the Social Psychology of Intergroup Relations*. European Association of Experimental Social Psychology by Academic Press, London ; New York.

[310] Henri Tajfel, M. G. Billig, R. P. Bundy, and Claude Flament. 1971. Social Categorization and Intergroup Behaviour. *European Journal of Social Psychology* 1, 2 (1971), 149–178. https://doi.org/10.1002/ejsp.2420010202

[311] Zeerak Talat, Hagen Blix, Josef Valvoda, Maya Indira Ganesh, Ryan Cotterell, and Adina Williams. 2022. On the Machine Learning of Ethical Judgments from Natural Language. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, Seattle, United States, 769–779. https://doi.org/10.18653/v1/2022.naacl-main.56

[312] Diana I. Tamir and Jason P. Mitchell. 2010. Neural Correlates of Anchoring-and-Adjustment during Mentalizing. *Proceedings of the National Academy of Sciences* 107, 24 (June 2010), 10827–10832. https://doi.org/10.1073/pnas.1003242107

[313] Kevin Tasa, Greg J. Sears, and Aaron C. H. Schat. 2011. Personality and Teamwork Behavior in Context: The Cross-Level Moderating Role of Collective Efficacy: Personality in Context. *Journal of Organizational Behavior* 32, 1 (Jan. 2011), 65–85. https://doi.org/10.1002/job.680

[314] Abraham Tesser. 1988. Toward a Self-Evaluation Maintenance Model of Social Behavior. In *Advances in Experimental Social Psychology*. Vol. 21. Elsevier, 181–227. https://doi.org/10.1016/S0065-2601(08)60227-0

[315] Susan L Thomas, Linda J Skitka, Stacy Christen, and Mark Jurgena. 2002. Social Facilitation and Impression Formation. *Basic and Applied Social Psychology* 24, 1 (2002), 67–70.

[316] Dorothy A Thornton and A John Arrowood. 1966. Self-evaluation, self-enhancement, and the locus of social comparison. *Journal of Experimental Social Psychology* 1 (1966), 40–48.

[317] Pamela Tierney and Steven M Farmer. 2002. Creative Self-Efficacy: Its Potential Antecedents and Relationship to Creative Performance. *Academy of Management Journal* 45, 6 (2002), 1137–1148.

[318] Ilaria Torre and Sébastien Le Maguer. 2020. Should Robots Have Accents?. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. 208–214. https://doi.org/10.1109/RO-MAN47096.2020.9223599

[319] Margaret L Traeger, Sarah Strohkorb Sebo, Malte Jung, Brian Scassellati, and Nicholas A Christakis. 2020. Vulnerable robots positively shape human conversational dynamics in a human–robot team. *Proceedings of the National Academy of Sciences* 117, 12 (2020), 6370–6375.

[320] Lee Edward Travis. 1925. The Effect of a Small Audience upon Eye-hand Coordination. *The Journal of Abnormal and Social Psychology* 20, 2 (1925), 142.

[321] Norman Triplett. 1898. The Dynamogenic Factors in Pacemaking and Competition. *The American Journal of Psychology* 9, 4 (1898), 507–533.

[322] Alan M. Turing. 1950. Computing Machinery and Intelligence. *Mind* LIX, 236 (Oct. 1950), 433–460. https://doi.org/10.1093/mind/LIX.236.433

[323] John C Turner, Michael A Hogg, Penelope J Oakes, Stephen D Reicher, and Margaret S Wetherell. 1987. *Rediscovering the Social Group: A Self-Categorization Theory*. Basil Blackwell.

[324] John C. Turner and Katherine J. Reynolds. 2012. *Handbook of Theories of Social Psychology: Volume 2*. SAGE Publications Ltd, London, 399–417. https://doi.org/10.4135/9781446249222

[325] Ertugrul Uysal, Sascha Alavi, and Valéry Bezençon. 2022. Trojan horse or useful helper? A relationship perspective on artificial intelligence assis-

tants with humanlike features. *Journal of the Academy of Marketing Science* 50, 6 (Nov. 2022), 1153–1175. https://doi.org/10.1007/s11747-022-00856-9

[326] Ertugrul Uysal, Sascha Alavi, and Valéry Bezençon. 2023. *Anthropomorphism in Artificial Intelligence: A Review of Empirical Work Across Domains and Insights for Future Research*. Emerald Publishing Limited, 273–308. https://doi.org/10.1108/S1548-643520230000020015

[327] Joanne V. Lloyd, Justine Schneider, Kezia Scales, Simon Bailey, and Rob Jones. 2011. Ingroup Identity as an Obstacle to Effective Multiprofessional and Interprofessional Teamwork: Findings from an Ethnographic Study of Healthcare Assistants in Dementia Care. *Journal of Interprofessional Care* 25, 5 (Sept. 2011), 345–351. https://doi.org/10.3109/13561820.2011.567381

[328] Wouter Van Den Bos, Eric Van Dijk, Michiel Westenberg, Serge A.R.B. Rombouts, and Eveline A. Crone. 2011. Changing Brains, Changing Perspectives: The Neurocognitive Development of Reciprocity. *Psychological Science* 22, 1 (Jan. 2011), 60–70. https://doi.org/10.1177/0956797610391102

[329] Wouter Van Den Bos, Michiel Westenberg, Eric Van Dijk, and Eveline A. Crone. 2010. Development of Trust and Reciprocity in Adolescence. *Cognitive Development* 25, 1 (Jan. 2010), 90–102. https://doi.org/10.1016/j.cogdev.2009.07.004

[330] Ruth Wageman, Heidi Gardner, and Mark Mortensen. 2012. The Changing Ecology of Teams: New Directions for Teams Research. *Journal of Organizational Behavior* 33, 3 (2012), 301–315. https://doi.org/10.1002/job.1775

[331] James C. Walliser, Patrick R. Mead, and Tyler H. Shaw. 2017. The Perception of Teamwork With an Autonomous Agent Enhances Affect and Performance Outcomes. 61, 1 (2017), 231–235. https://doi.org/10.1177/1541931213601541

[332] Joseph B. Walther. 2009. Computer-Mediated Communication and Virtual Groups: Applications to Interethnic Conflict. *Journal of Applied Communication Research* 37, 3 (Aug. 2009), 225–238. https://doi.org/10.1080/00909880903025937

[333] Dakuo Wang, Elizabeth Churchill, Pattie Maes, Xiangmin Fan, Ben Shneiderman, Yuanchun Shi, and Qianying Wang. 2020. From Human-Human Collaboration to Human-AI Collaboration: Designing AI Systems That

Can Work Together with People. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3334480.3381069

[334] Dakuo Wang, Justin D. Weisz, Michael Muller, Parikshit Ram, Werner Geyer, Casey Dugan, Yla Tausczik, Horst Samulowitz, and Alexander Gray. 2019. Human-AI Collaboration in Data Science: Exploring Data Scientists' Perceptions of Automated AI. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 211 (Nov. 2019), 24 pages. https://doi.org/10.1145/3359313

[335] Danding Wang, Qian Yang, Ashraf Abdul, and Brian Y Lim. 2019. Designing Theory-Driven User-Centric Explainable AI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–15.

[336] Ming-Te Wang and Jessica Degol. 2013. Motivational Pathways to STEM Career Choices: Using Expectancy–Value Perspective to Understand Individual and Gender Differences in STEM Fields. *Developmental Review* 33, 4 (Dec. 2013), 304–340. https://doi.org/10.1016/j.dr.2013.08.001

[337] Qiaosi Wang, Ida Camacho, Shan Jing, and Ashok K. Goel. 2022. Understanding the Design Space of AI-Mediated Social Interaction in Online Learning: Challenges and Opportunities. *Proceedings of the ACM on Human-Computer Interaction (PACM HCI)* 6, CSCW1, Article 130 (April 2022), 26 pages. https://doi.org/10.1145/3512977

[338] Weiyu Wang and Keng Siau. 2019. Artificial Intelligence, Machine Learning, Automation, Robotics, Future of Work and Future of Humanity: A Review and Research Agenda. *Journal of Database Management* 30, 1 (Jan. 2019), 61–79. https://doi.org/10.4018/JDM.2019010104

[339] Zuoming Wang, Joseph B. Walther, and Jeffrey T. Hancock. 2009. Social Identification and Interpersonal Communication in Computer-Mediated Communication: What You Do Versus Who You Are in Virtual Groups. *Human Communication Research* 35, 1 (Jan. 2009), 59–85. https://doi.org/10.1111/j.1468-2958.2008.01338.x

[340] Juergen Wegge. 2000. Participation in Group Goal Setting: Some Novel Findings and a Comprehensive Model as a New Ending to an Old Story. *Applied Psychology* 49, 3 (July 2000), 498–516. https://doi.org/10.1111/1464-0597.00028

[341] Katharina Weitz, Dominik Schiller, Ruben Schlagowski, Tobias Huber, and Elisabeth André. 2019. "Do You Trust Me?": Increasing User-

Trust by Integrating Virtual Agents in Explainable AI Interaction Design. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*. Association for Computing Machinery, Paris France, 7–9. https://doi.org/10.1145/3308532.3329441

[342] Darrell M. West. 2018. *The Future of Work: Robots, AI, and Automation*. Brookings Institution Press, Washington, D.C.

[343] Joel Wester, Minha Lee, and Niels Van Berkel. 2023. Moral Transparency as a Mitigator of Moral Bias in Conversational User Interfaces. In *Proceedings of the 5th International Conference on Conversational User Interfaces*. ACM, Eindhoven Netherlands, 1–6. https://doi.org/10.1145/3571884.3603752

[344] Ladd Wheeler. 1966. Motivation as a determinant of upward comparison. *Journal of Experimental Social Psychology* 1 (1966), 27–31.

[345] Ladd Wheeler and Jerry Suls. 2007. Assimilation in social comparison: Can we agree on what it is? *Revue internationale de psychologie sociale* 20, 1 (2007), 31–51.

[346] Thomas A. Wills. 1981. Downward Comparison Principles in Social Psychology. *Psychological Bulletin* 90, 2 (1981), 245–271. https://doi.org/10.1037/0033-2909.90.2.245

[347] Sarah Wilson and Roger K Moore. 2017. Robot, Alien and Cartoon Voices: Implications for Speech-enabled Systems. In *1st Int. Workshop on Vocal Interactivity in-and-between Humans, Animals and Robots (VIHAR-2017)*. 40–44.

[348] Katie Winkle, Praminda Caleb-Solly, Ute Leonards, Ailie Turton, and Paul Bremner. 2021. Assessing and Addressing Ethical Risk from Anthropomorphism and Deception in Socially Assistive Robots. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction* (Boulder, CO, USA) *(HRI '21)*. Association for Computing Machinery, New York, NY, USA, 101–109. https://doi.org/10.1145/3434073.3444666

[349] Elizabeth Baily Wolf, Jooa Julia Lee, Sunita Sah, and Alison Wood Brooks. 2016. Managing Perceptions of Distress at Work: Reframing Emotion as Passion. *Organizational Behavior and Human Decision Processes* 137 (Nov. 2016), 1–12. https://doi.org/10.1016/j.obhdp.2016.07.003

[350] Joanna Wolfe and Elizabeth Powell. 2009. Biases in Interpersonal Communication: How Engineering Students Perceive Gender Typical Speech Acts in Teamwork. *Journal of Engineering Education* 98, 1 (Jan. 2009), 5–16.

https://doi.org/10.1002/j.2168-9830.2009.tb01001.x

[351] Charlotte Wollermann, Eva Lasarcyk, Ulrich Schade, and Bernhard Schröder. 2013. Disfluencies and Uncertainty Perception-Evidence from a Human-Machine Scenario. In *The 6th Workshop on Disfluency in Spontaneous Speech (DiSS)*. 73–76.

[352] Sarah Woods, Kerstin Dautenhahn, and Christina Kaouri. 2005. Is Someone Watching Me? Consideration of Social Facilitation Effects in Human-Robot Interaction Experiments. In *2005 International Symposium on Computational Intelligence in Robotics and Automation*. IEEE, 53–60.

[353] Anita Williams Woolley, Christopher F. Chabris, Alex Pentland, Nada Hashmi, and Thomas W. Malone. 2010. Evidence for a Collective Intelligence Factor in the Performance of Human Groups. *Science* 330, 6004 (Oct. 2010), 686–688. https://doi.org/10.1126/science.1193147

[354] Meng-Hsin Wu, Su-Fang Yeh, XiJing Chang, and Yung-Ju Chang. 2021. Exploring Users' Preferences for Chatbot's Guidance Type and Timing. In *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing*. 191–194.

[355] Qian Yang, Aaron Steinfeld, Carolyn Rosé, and John Zimmerman. 2020. Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Honolulu HI USA, 1–13. https://doi.org/10.1145/3313831.3376301

[356] Janice D. Yoder. 1994. Looking Beyond Numbers: The Effects of Gender Status, Job Prestige, and Occupational Gender-Typing on Tokenism Processes. *Social Psychology Quarterly* 57, 2 (June 1994), 150. https://doi.org/10.2307/2786708

[357] Janice D. Yoder, Patricia Aniakudo, and Lynne Berendsen. 1996. Looking beyond Gender: The Effects of Racial Differences on Tokenism Perceptions of Women. *Sex Roles* 35, 7–8 (Oct. 1996), 389–400. https://doi.org/10.1007/BF01544128

[358] Janice D. Yoder and Laura M. Sinnett. 1985. Is it All in the Numbers?: A Case Study of Tokenism. *Psychology of Women Quarterly* 9, 3 (Sept. 1985), 413–418. https://doi.org/10.1111/j.1471-6402.1985.tb00890.x

[359] Robert B Zajonc. 1965. Social Facilitation. *Science* 149, 3681 (1965), 269–274.

[360] Catherine Amine Zanbaka, Amy Catherine Ulinski, Paula Goolkasian, and Larry F Hodges. 2007. Social Responses to Virtual Humans: Implications for Future Interface Design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1561–1570.

[361] Ali Zarifhonarvar. 2023. Economics of ChatGPT: A Labor Market View on the Occupational Impact of Artificial Intelligence. (Feb. 2023). https://doi.org/10.2139/ssrn.4350925

[362] Ling.Yu Zhu, Zhengkun Zhang, Jun Wang, Hongbin Wang, Haiying Wu, and Zhenglu Yang. 2022. Multi-Party Empathetic Dialogue Generation: A New Task for Dialog Systems. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Dublin, Ireland, 298–307. https://doi.org/10.18653/v1/2022.acl-long.24

[363] Jonathan C. Ziegert and Paul J. Hanges. 2005. Employment Discrimination: The Role of Implicit Attitudes, Motivation, and a Climate for Racial Bias. *Journal of Applied Psychology* 90, 3 (May 2005), 553–562. https://doi.org/10.1037/0021-9010.90.3.553

[364] Caleb Ziems, Jane A. Yu, Yi-Chia Wang, Alon Halevy, and Diyi Yang. 2022. The Moral Integrity Corpus: A Benchmark for Ethical Dialogue Systems. (April 2022). https://doi.org/10.48550/arXiv.2204.03021 arXiv:2204.03021 [cs].

[365] Jakub Złotowski, Diane Proudfoot, Kumar Yogeeswaran, and Christoph Bartneck. 2015. Anthropomorphism: Opportunities and Challenges in Human-Robot Interaction. *International Journal of Social Robotics* 7, 3 (2015), 347–360.

[366] Jakub Złotowski, Ewald Strasser, and Christoph Bartneck. 2014. Dimensions of Anthropomorphism: From Humanness to Humanlikeness. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction* (Bielefeld, Germany) *(HRI '14)*. Association for Computing Machinery, New York, NY, USA, 66–73. https://doi.org/10.1145/2559636.2559679