DECODING DECEPTION AND COLLUSION: BEHAVIORAL ANALYSIS OF RELATIONAL MESSAGES AND INTERPERSONAL RELATIONSHIPS IN GROUP COMMUNICATION

By

Saiying Ge

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF MANAGEMENT INFORMATION SYSTEMS

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2024

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by: Saiying Ge
titled: DECODING DECEPTION AND COLLUSION: BEHAVIORAL ANALYSIS OF RELATIONAL MESSAGES AND INTERPERSONAL RELATIONSHIPS IN

and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Jay F. Nunamaker — Date: Jun 20, 2024

Judee Burgoon (Jun 26, 2024 21:07 EDT) — Date: Jun 26, 2024
Judee Burgoon

Yong Ge — Date: Jun 20, 2024
Yong Ge

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Jay F. Nunamaker — Date: Jun 20, 2024
Jay F. Nunamaker
MIS Dissertation Committee Chair
Management Information Systems

ACKNOWLEDGEMENTS

I extend my deepest gratitude to my advisor, Dr. Jay F. Nunamaker Jr., whose unwavering support and guidance have been instrumental in shaping this dissertation. As his last student, I am honored to be the 105th. I have learned more than just research from him; his mentorship has profoundly influenced my growth. My sincere thanks to my co-advisor, Dr. Judee K. Burgoon. Many aspects of my study build on her work, and her contributions in sharpening the theoretical aspects have been invaluable. I am also grateful to Dr. Yong Ge for his insights from the deep learning and machine learning perspective, which enriched my understanding and the quality of this work. To the MIS department and my cohort, your collective support and resources have been a cornerstone of my success. I am thankful for the nurturing environment that has fostered my development throughout this academic journey.

To my loving family, your unwavering belief in me has been a pillar of strength. I also extend my gratitude to my friends for their constant encouragement. Special thanks to my husband, Brendon, whose support helped me stay focused and overcome moments of procrastination, and to my sister, Kathleen, who has been there to laugh with me and at me when it was needed most. You both have been my sanity checks and the best cheerleaders I could ask for. A heartfelt thanks to my lab mate, study partner, collaborator, and maid of honor, Xinran (Rebecca) Wang, for the countless hours we spent delving into research and navigating the ups and downs of PhD life together. Your support and companionship have been invaluable.

Thank you all for being part of this journey.

DEDICATION

This dissertation is dedicated to my dad, Feng Ge, the first Dr. Ge in the family, whose hard work and dedication have always inspired me.

And to my mom, Yujiao Zhang, whose unwavering love and support have been the backbone for both Drs. Ge.

Thank you both for your endless encouragement and belief in me.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Collusion is prevalent and costly, particularly in fraud scenarios. Despite extensive research on deception, there are limited studies on collusive deception, where a group of deceivers with hidden objectives works together to undermine a larger group. This dissertation investigates how secretive collaboration and deception influence relational communication within groups. Two experiments were conducted to explore how collusive deception affects deceivers' behaviors and interactions with other group members. Analysis of recorded conversations revealed that deceivers often adopt "flight" strategies to evade detection, express less trust, and isolate themselves from other deceivers. Successful deception involves deceivers expressing more trust towards truth-tellers as a bonding tactic. Additionally, the increased overall level of distrust within the group gives deceivers an edge to obscure their actions and achieve their agenda. The second experiment analyzed nonverbal behaviors in a mock hiring scenario using automated tools for face and head feature analysis. Results indicated that deceivers exhibited reduced dominance and appeared less nervous emotionally, yet more nervous cognitively during interactions with fellow deceivers. This points to how different levels of appraisals during collusive interactions influence emotions and cognition. This study further develops deception detection models by incorporating network metrics and leveraging Graph Neural Networks (GNNs), which improve performance in deception detection. Additionally, the integration of advanced language technologies like GPT for automated context-level tagging of transcripts enhances scalability and cost-effectiveness.

The findings significantly enhance the understanding of how collusive deception influences verbal and nonverbal relational messages, providing crucial insights into deception research. The tested deception detection system demonstrates the potential for developing

comprehensive end-to-end solutions with high efficiency and performance in identifying deceptive behavior.

Chapter 1: Introduction

Collusion is recognized as a substantial problem in various domains. In the financial sector, it often involves managers, employees, and suppliers who collude to commit fraud, significantly harming shareholder interests (Islam et al., 2011). In the case of stock trading, collusion can take the form of traders working together to manipulate market prices for personal gain. This not only disrupts market integrity but also impacts other investors and the financial system as a whole (Palshikar and Apte, 2008). At an industry level, collusion can occur between firms conspiring to fix prices, which adversely affects consumer interests. In the gaming world, collusion undermines fair competition and can significantly detract from the gaming experience (Mazrooei et al., 2013). According to a report by the Association of Certified Fraud Examiners (ACFE), the median US firm loses approximately 5 percent of its annual revenue to fraud, with approximately 50 percent of fraud cases involving collusion schemes. Additionally, the report reveals that the median loss from collusive fraud is 259 percent higher than the median loss from fraud committed by a single individual (Dorris, 2018).

Despite extensive research on deception, studies focusing on collusive deception, where a group of deceivers with hidden objectives collectively act to undermine the interests of a larger group, are limited. This dissertation seeks to explore how secretive collaboration and deception influence relational communication within groups, addressing two primary research questions:

1. How does collusive deception affect the deceiver's behaviors within groups?
2. How does collusive deception affect a deceiver's interactive behaviors with other group members?

According to Interpersonal Deception Theory (IDT), deceivers engage in a blend of strategic and nonstrategic activities. Strategic efforts are directed at managing information, behaviors, and overall image to facilitate deception, while nonstrategic activities often inadvertently reveal signs of cognitive load, memory issues, speech disruptions, or failed attempts to suppress information leakage (Burgoon et al., 1994). By exploring both verbal and nonverbal communication, this research aims to outline the complex interplay between the strategies deceivers consciously and unconsciously use in the conversation and the unintentional signals they release in their face and head movement, thus offering a nuanced understanding of how collusive deception influences relationships within deceptive groups.

Verbal behaviors are hypothesized to be more aligned with strategic activities because they involve deliberate manipulation of language to influence perceptions and outcomes. Note that strategic activity may not always be highly conscious (Burgoon et al., 2021). This research focuses on the strategic use of relational messages regarding trust and distrust to manipulate group dynamics. By expressing trust and distrust, deceivers can manage information and impressions to facilitate deception. In addition to verbal cues, we also test the nonverbal behaviors that are generally less deliberately controlled and tend to reflect more nonstrategic activities. By examining nonverbal cues indicating dominance and nervousness, we aim to explore how collusive deception impacts authentic relationships between deceivers.

The results of the verbal study indicate that deceivers often adopt "flight" strategies to evade detection. Specifically, deceivers tend to express less trust and distrust toward their fellow colluders and generally reduce their overall expressions of trust. These behaviors indicate a tendency for deceivers to isolate themselves from both truth-tellers and fellow deceivers, perhaps in an attempt to “fly under the radar” and avoid detection. On the nonverbal front, the research

shows that deceivers exhibit reduced dominance and less emotional nervousness but simultaneously show heightened cognitive stress during interactions with other deceivers. This mixed result in nervousness underscores how collusive deception reduces negative emotions due to decreased threats from fellow deceivers yet increases the cognitive load due to the heightened demand to control the situation and manage the impressions.

This study is pioneering in its focus on collusive interactions within groups, expanding beyond the traditionally studied dyadic scenarios to include settings where multiple deceivers operate collaboratively. The findings of this study significantly enhance our understanding of how collusive deception influences verbal and nonverbal relational messages, contributing crucial insights to the field of deception research. By incorporating network metrics and leveraging Graph Neural Networks (GNNs), the study demonstrates a marked improvement in the performance of deception detection models. Furthermore, it explores the use of advanced language technologies like GPT for automated context-level tagging of transcripts, which enhances the scalability and cost-effectiveness of the detection process. These innovations not only hold potential for applications in security, law enforcement, and corporate compliance but also pave the way for developing comprehensive, end-to-end solutions capable of real-time fraud detection and rapid analysis of large datasets.

Additionally, this dissertation is part of the SCAN: Socio-Cultural Attitudinal Networks project, a component of the Multidisciplinary University Research Initiative (MURI, ONR-15-FOA-0011), which aims to enhance cross-cultural understanding of relationships among groups of people. The ultimate objective is to equip military officers with the insights needed to navigate complex social dynamics effectively during overseas operations. For military officers involved in international missions, effective communication and cooperation with local communities are crucial.

Understanding and strategically leveraging group interactions allow officers to quickly identify trustworthy individuals, leaders, friends, and foes, which is essential for the success of any operation. These officers must be adept at interpreting vertical relationships (e.g. dominance-deference, power hierarchy) amongst a group of people, as well as horizontal relationships (e.g. like-dislike, affection, trust). Misinterpretations or failure to recognize deceptive cues within these interactions can jeopardize mission outcomes. This research initiative included a large-scale international experiment utilizing the "Resistance" adversarial game, executed across seven locations in six regions, including the U.S.A., Hong Kong, Israel, Fiji, Zambia, and Singapore. A subsequent experiment, a mock hiring interview conducted via Zoom, incorporated an upgraded experimental design and enhanced technology to further this research. This dissertation employs computational models to analyze the dynamics of trust-distrust, dominance-submission, and nervousness-composure as they are manifested verbally and nonverbally in cooperative and adversarial contexts. This analytical approach enhances our understanding of the nuanced social cues and strategic communications that play a crucial role in group interaction.

The dissertation is structured into three main chapters following the introduction. Chapter 2 establishes the theoretical background, detailing the frameworks that inform our understanding of collusive deception and relational messages. Chapter 3 the verbal relational messages of trust and distrust, analyzing how deceivers manipulate these elements, while Chapter 4 examines the nonverbal cues associated with dominance and nervousness. Together, these chapters provide a comprehensive exploration of both the verbal and nonverbal aspects of collusive deception in group settings.

Chapter 2: Collusive Deception and Relational Message

2.1 Theoretical Backgrounds

2.1.1 Relational Message

Relational messages are intricate embroideries woven from the threads of both verbal and nonverbal communication, each thread contributing to the overall representation of the interpersonal dynamic. Bateson's mid-20th-century concept of metacommunication (1951,1958) laid the groundwork for our understanding of relational messages. He proposed that communication operates on two levels: the 'report' level, which is the content of the message, and the 'command' level, which contextualizes how the message should be interpreted based on the relationship between the sender and receiver. Watzlawick, Beavin, and Jackson (2011) expanded on this, emphasizing that relational information is inherent in all communicative acts across all contexts. They highlighted the importance of nonverbal information in conveying relational messages, which can often carry more weight than the words used. Building on these foundational ideas, Burgoon and Hale (1984, 1987) developed the "Topoi of Relational Communication." This concept categorizes the various dimensions through which relational messages are conveyed. These non-orthogonal dimensions include but are not limited to, dominance, affection, involvement, composure, similarity, and trust. Each dimension has distinct verbal and nonverbal signals that uniquely characterize the relationship, providing rich, complex meanings that underpin interpersonal communications. In this dissertation*, we adopt the definition of relational messages as nonverbal and verbal messages that signal how people regard one another and define their interpersonal relationships* (Burgoon, 2021).

Relational messages indeed encompass both verbal and nonverbal cues that together provide a multidimensional view of interpersonal dynamics. Verbal cues such as word choice,

tone, and speech patterns provide insights into the nature of the relationship. For example, Supportive and polite language (Lam, 2011) may also engender trust in the speaker. In contrast, swear words and language barriers impede trust-building (Feely & Harzing, 2003). Moreover, dominant persons talk more often, for a longer duration, and are more likely to contribute to the conversation (Zhou et al., 2004; Pentland et al., 2021). On the other hand, nonverbal behaviors such as body language and facial expressions are potent conveyors of relational messages. A smile or a nod can indicate trust and composure (Burgoon et al. 2021), while crossed arms or lack of eye contact might signal less nervousness and a no-smiling face associated with perceived dominance (Keating et al., 1981). These cues, when read in conjunction with verbal communication, provide a fuller understanding of how individuals relate to each other within various contexts.

2.1.2 Relational Message in Deceptive Interaction

In the context of group interactions, particularly when examining deception, relational messages serve as a critical tool for assessment. These messages, encoded in both what is said and what remains unspoken, offer a window into the perceptions held by group members about one another. They serve a dual function: firstly, to decode how individuals interpret and respond to the relational message sent by their peers, and secondly, to determine the potential for these signals to predict deceptive behavior (Burgoon, 2022).

The utilization of the Brunswikian lens model in deception detection research, as detailed by Burgoon et al. (2022), provides a framework for correlating observable behaviors with the internal perceptions that they evoke. As Figure 1. demonstrates, this model helps to dissect the complexity of relational communication by charting the path from observable, measurable verbal and nonverbal cues (distal indicators) to the subjective interpretations of these cues (perceptions of relational dimensions). From the application perspective, The Brunswikian lens model shows how the different aspects of the signaling process can be combined. The distal, objective signals

that can be measured and factored with automated computer tools can be linked to the psychological perceptual judgments that represent relational message themes. These subjective percepts in turn predict communicative outcomes such as successful identification of another's deception or credibility.

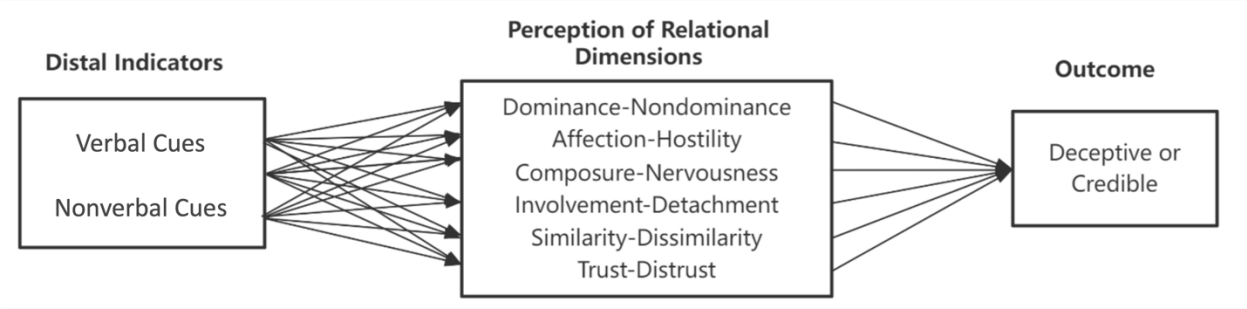


Figure 1. Brunswikian Lens Model of Relational Communication

For instance, in an early study involving a mock theft scenario, researchers used this model to analyze behaviors like tension, arousal, and involvement, finding that deceivers displayed higher levels of arousal and lower levels of involvement (Jensen et al., 2008). Substantial research has reported that deception is often characterized by less verbal or nonverbal dominance than truthful communication (Buller et al, 1994; Burgoon et al, 2001). Since dominance connotes credibility, panache, and interpersonal persuasiveness (Burgoon et al., 1998), it is often the case that liars who use dominant strategies are more likely to be believed. Following the same steam, when taking more granular, temporal measurements, researchers found that while deceivers decreased their perceived trust and dominance over time (Dunbar et al., 2021). Whereas deceivers maintained the same degree of nervousness they displayed in the interaction, truth-tellers became increasingly relaxed (Burgoon et al., 2021). Expanding on the study of relational communication and deception, Burgoon and the team (2022) advocated for the integration of automated technologies capable of detecting subtle kinesic and vocalic changes. These advancements in measurement techniques

stand to significantly refine the analysis of relational messages, enhancing our ability to discern truthfulness and deceit in group interactions.

2.1.3 Deception in Groups

Following the study of relational messages within the context of group interactions and their role in deception detection, it's important to understand the distinctive nature of deception in group settings as opposed to dyadic interactions.

In larger groups, the dynamics of communication and deception differ markedly from dyadic interactions. The exchange of information tends to be less intimate, with lower levels of affective ties (Moreland 2010), leading to a perceived distance among members (Cartwright and Zander, 1968; DeSanctis and Gallupe, 1987). This distance can result in cautiousness towards new messages and less tolerance for different viewpoints (Zhou and Zhang, 2006). Additionally, larger groups offer opportunities for free-riding (Valacich et al., 1992) and typically see a few individuals dominating the conversation (Báles et al., 1951). Such a lack of direct connection would further motivate the deceiver to choose the avoidance strategy where deceivers are likely to hide information as much as possible. This dynamic could also allow deceivers to avoid detection and strategically observe and adapt to others' behaviors (Giles and Baker, 2008).

While groups are generally less emotionally charged than dyads, the complex relationship and information flowing in the group could introduce more arousal and give cues to detect deception. The presence of multiple individuals can lead to increased arousal and behavioral leakage, as noted by McCullagh and Landers (1976). Moreover, the diversity of knowledge and perspectives within a group, accelerated by turn-taking and interaction patterns, places a higher cognitive load on deceivers as they monitor group members and gauge suspicion levels (Zhou and Zhang, 2006).

When examining the relational messages, the evidence supports that group size moderated the relational message conveyed in the interaction with the influence of deception. For example, contradicting from the dyad group that deceivers are less pleasant, but the group of more than two does not show the same effect (Zhou and Zhang 2006, Zhou et al., 2014). Interestingly, when comparing dominance in communication between deceivers in dyads and triads, the researchers found that deceivers exhibited more dominance in triads than in dyads (Zhou and Zhang, 2006). However, this pattern reverses in larger groups (six to eight members), where deceivers show less dominance than truth-tellers (Zhou et al.2014; Dunbar et al, 2021). Additionally, vocal pitch, a traditional indicator of nervousness, may not be as significant in group settings (Burgoon et al., 2021).

While research on group deception and the strategic alteration of relational messages has gained increasing interest among researchers, the primary focus has been on the behaviors of individuals (deceivers or truth-tellers). However, the exploration of how verbal and nonverbal relational messages manifest in interactions involving individuals with different roles, such as between a deceiver and a truth-teller, as well as among individuals with the same role, like deceiver-to-deceiver interactions, remains an area less explored in current research.

2.1.4 Collusion in Group Deception

In contrast to dyadic interactions, group dynamics introduce the complex element of deceiver-to-deceiver communication. Collusion, a prevalent issue across various sectors such as finance and gaming, has traditionally been examined primarily through the lenses of economics and accounting. These studies tend to focus on the financial and regulatory impacts of collusion (as summarized in Table 1). However, this approach overlooks a critical aspect: the interpersonal dynamics and communication strategies employed by deceivers within these groups.

This oversight highlights a significant gap in our understanding, emphasizing the need to explore collusion through an Information Systems (IS) lens. By adopting a behavioral analytics approach, it becomes possible to gain a deeper understanding of the underlying interpersonal mechanisms and strategies that facilitate collusion. This perspective not only enriches the theoretical framework but also enhances the practical applications for detecting and mitigating collusion in diverse settings.

Collusion Context	**Colluders**	**Detection/Mitigation Method**	**Collusion Definition**	**Reference**
Financial Fraud	Managers, Employees, Suppliers, etc.	Combining understanding of collusion with audit standards to address financial fraud.	Collusion is a secret agreement between two or more individuals for a deceitful or fraudulent purpose.	Islam et al., 2011; Rechtman, 2019
Stock Trading	Traders	Identifying collusion sets in trading via graph clustering.	Unscrupulous organizations and groups of individuals attempt to manipulate or influence activities on stock exchanges with the intention of making profits through illegal or unfair means.	Palshikar & Apte, 2008; Islam, Haque, Alam, & Tarikuzzaman, 2009; K, A., T, I., & W, M., 2022
Price Setting	Firms	Trade associations facilitating communication among firms.	Exchanged information about the gross list price of product, informed each other of their planned gross price increases, and agreed upon harmonizing the list prices over the period	Rahman, 2014; Sutar, Chaudhary, & Yadav, 2022; Boshoff & Paha, 2021
Online Multiplayer Gaming	Players	Analysis of social relationships and behavioral patterns in games.	Collusion in a competitive multi-agent game occurs when two or more agents cooperate covertly to the disadvantage of others.	Smed, Knuutila, & Hakonen, 2006; 2007; Bonjour, Aggarwal, &

				Bhargava, 2022
Card Games (e.g., Poker, Bridge)	Players	Rule-based models analyzing historical game records.	The practice of two or more parties deliberately cooperating to the detriment of the other parties.	Mazrooei, Archibald, & Bowling, 2013; Vallve-Guionnet, 2005; Yan, 2010

Table 1. Summary of the Current Collusion Studies

This study adopts a comprehensive definition of collusive deception as "*the collective action by a group of deceivers with hidden objectives aimed at undermining the interests of a larger group*." Throughout this dissertation, the terms 'colluder' and 'deceiver' are used interchangeably to denote individuals involved in collusive deception. The nature of collusive deception influences how relational messages are communicated, including three aspects: deception, cooperative solidarity, and strategic isolation.

Deception

Deception within groups involves the intricate coordination of false narratives, the manipulation of information, and the crafting of misleading scenarios that negatively impact the larger group's interests. Interpersonal Deception Theory (Buller and Burgoon, 1996) posits that both deceivers (senders) and the truth-teller (receivers) navigate complex relational dynamics. Specifically, deceivers must actively work to cultivate trust from those they intend to deceive. This often entails carefully monitoring the behavior of the truth-teller and adjusting their own actions to project an image of trustworthiness and likability. Consequently, deceivers might employ relational messages aimed at enhancing their perceived trustworthiness and appeal, leveraging emotional and social cues to construct a facade of integrity and benevolence. For instance, a

deceiver may demonstrate increased dominance to appear confident, poised, credible, and persuasive (Burgoon and Qin, 2006).

Cooperative Solidarity

In collusive deception, cooperative solidarity arises as colluders work closely together to pursue a shared deceptive objective that ensures consistency and reinforces the group's false narrative. This solidarity fosters an in-group effect (Tajfel and Turner, 2004), enhancing unity and loyalty among members. Such a strong group identity shifts the relational dynamics, leading to an us-versus-them mentality that fundamentally shapes members' interactions both within the group and with outsiders. Many scholarly articles and research studies explore how group dynamics influence individual behavior and communication. For Example, Research found that people prefer those in their ingroup and find them to be more trustworthy than people from outgroups (Hewstone et al., 2002). In the case of collusive deception, the cooperative solidarity process may result in the unintended expression of verbal and nonverbal cues that signal affirmations of trust, support, and agreement.

Strategic Isolation

The core of collusion is maintaining secrecy about the collaborators' real relationship to preserve the appearance of integrity while pursuing hidden agendas. To minimize suspicion and protect colluder group interests, colluders might strategically maintain a distance from one another. Research on group deception indicates that deceivers, concerned with being discovered, often adopt a submissive (flight) approach to hide their identity and “fly under the radar.” (Dunbar et al., 2021). Contrary to the cooperative solidarity that fosters a sense of bonding, strategic isolation can potentially undermine the interaction between colluders or even manifest negative interactions among them.

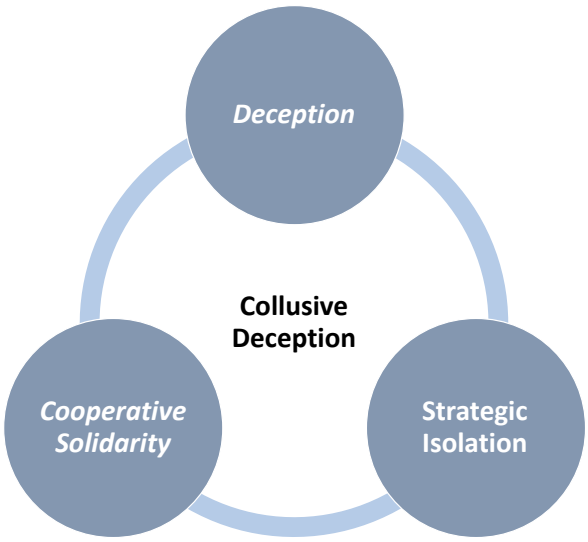


Figure 2. Factors of Collusive Deception

These three factors—deception, cooperative solidarity, and strategic isolation—highlight the complex interdependence at the core of collusive deception and underscore the importance of comprehensive research to understand these dynamics fully. This analysis informs our exploration of two critical research questions: RQ1, which examines how collusive deception influences the behaviors of deceivers within groups, and RQ2, which explores how it affects deceivers' interactive behaviors with other group members. The forthcoming chapters will further dissect these behaviors, with Chapter 3 focusing on verbal cues and Chapter 4 delving into non-verbal cues within the context of collusive deception.

Chapter 3: Verbal Behaviors, Collusive Deception, and Interpersonal Relationships in Group Communications

3.1 Introduction

This chapter explores how collusive deception influences the verbal behaviors of deceivers in interactions with both fellow deceivers and truth-tellers, focusing on strategic shifts in relational messages of trust and distrust. The study conducted a large cross-culture experiment where colluders operated covertly among truth-tellers, with conversations recorded and trust-distrust messages annotated. Results indicate that deceivers often adopt "flight" strategies to evade detection. Specifically, while deceivers express trust less often and even less so towards fellow deceivers, typically isolating themselves from other people in the group. Additionally, the study demonstrates the potential of incorporating network metrics and leveraging Graph Neural Networks (GNNs) to enhance deception detection accuracy, along with exploring the use of advanced language models like GPT for automated transcript tagging, enhancing scalability and cost-effectiveness.

3.2 Literature Review and Theoretical Backgrounds

3.2.1 Verbal Relational Messages

The study of verbal communication is crucial in identifying and understanding deceptive behaviors. Verbal information provides the "ground truth" on whether a speaker is concealing, misleading, or lying about certain facts. Additionally, the process of deceiving often affects the speaker's physiological state, resulting in changes in arousal, affect, and cognitive load (Zuckerman and Driver, 2014). This in turn may lead to deceivers inadvertently displaying “leakage” behaviors that give away their deceptive intent (Ekman and Friesen, 1969). Research has identified distinct verbal patterns associated with deception, such as negative verbal statements,

irrelevant information, discrepant information, verbal uncertainty, repetition, sentence quantity, verbal negation, emotional messages, and the use of second and third-person pronouns, as well as motion verbs (Zuckerman et al., 1981; Zuckerman and Driver, 2014; DePaulo et al., 2003; Sporer and Schwandt, 2006; Hauch et al., 2015). These patterns serve as critical indicators of potential deceit.

However, focusing solely on individual verbal cues can overlook the broader, more complex relational dynamics within groups. Relational messages, which convey the nature of relationships and social dynamics among group members, provide significant insights into how trust, power, and intimacy are negotiated and represented. This dissertation adopts the term "Verbal Relational Messages" to describe the verbally conveyed relational messages, which signal how people regard one another and define their interpersonal relationships.

Verbal relational messages can be strategically utilized on two distinct levels to facilitate deception. At the individual level, deceivers may adapt and manipulate verbal relational messages to sustain their relationships with truth-tellers and maintain an appearance of innocence. This manipulation helps them blend in and avoid suspicion, crucial for the long-term success of their deceptive strategies. On the group level, deceivers often collaborate, using verbal relational messages to assist one another in managing their public personas. This collaborative effort is aimed at creating a cohesive group-level impression that supports their false narrative, making it more persuasive and difficult to question. By effectively managing how they are perceived collectively, deceivers can enhance the credibility of their deceptive narrative within the group and beyond. Analyzing verbal relational messages allows for a more profound insight into how deceivers construct and maintain their facades both individually and collectively.

3.2.2 Trust-Distrust in Collusive Deception

While Burgoon and Hale's (1984) relational topoi framework comprehensively categorizes various dimensions of interpersonal relationships, trust-distrust is particularly emphasized as one of the fundamental dimensions of how interpersonal relationships are understood and navigated. Trust is a key relational theme, like power or status, through which people evaluate and express their relationship's status both verbally and nonverbally (Burgoon, 2001). It frames how messages are interpreted within interactions. Trust-building messages contribute to a sense of security and confidence in the communicator, crucial for group functionality (Iacono and Weisband 1997). Conversely, expressions of distrust can foster skepticism and caution, which may weaken relational bonds (Tomlinson and Lewicki, 2003).

Within the concept of collusive deception, the trust-distrust continua becomes particularly relevant for several reasons. Firstly, deception is a breach of trust (Elangovan and Shapiro, 1998). Collusive deception involves crafting false narratives and manipulating information, which directly undermines the trustworthiness of the deceiver. Therefore, pretending to be trustworthy and honest is a common strategy of deception (Buller and Burgoon, 1996). Additionally, the presence of subgroups within a larger group introduces additional layers to the trust-distrust dynamic. Collusive deception divides the group into colluders and truth-tellers, each subgroup having its own team agenda and varying degrees of trust. Research suggests that in relationships marked by diversity or perceived adversarial intent, individuals tend to attribute more positive traits and intentions to ingroup members than to outgroup members (Foddy et al., 2009; Lewicki et al., 1998). Given the inherent breach of trust and the complex trust dynamics, it is crucial to further examine verbal relational messages to better understand how trust is communicated and perceived in collusive deception.

To understand the role of trust within collusive deception fully, we must first define what trust and distrust entail in interpersonal relations. Lewicki and his colleagues (1998) provide a comprehensive examination of trust and distrust, emphasizing that while they are separate constructs, they also function as conceptual opposites. Trust is characterized by having a confident positive outlook on another's behavior, whereas distrust is marked by a confident negative expectation of another's actions. Building on these notions, McKnight and Chervany (2001) offer a framework within the information systems (IS) field that aligns with communication scholars' perspectives, dividing trust at the interpersonal level into perceptual, intentional, and behavioral components. This framework posits that trust involves *positive beliefs in another's competence, benevolence, integrity, and the predictability of their actions, leading to a willingness to rely on them, which is reflected in cooperative behaviors*. Conversely, distrust stems from *skepticism about these qualities and results in reluctance to depend on others, often leading to defensive behaviors like information withholding and strict formal agreements.*

3.2.3 Deceivers' Perceived Trust and Distrust

Extensive research has demonstrated that truth-tellers can often sense deception. This detection is facilitated not only through direct evidence but also by subtle clues that deceivers may inadvertently reveal through changes in their physical and communicative behaviors. These cues critically inform perceptions of trust and distrust. The modified Brunswikian lens model supports the idea that observable behaviors serve as objective indicators that influence psychological judgments and can predict deception or credibility (Burgoon et al., 2021). Researchers have used this model to predict perceived relational messages from automatically measured kinesic, vocalic, and linguistic indicators, which in turn help to predict the veracity of group members (Burgoon et al., 2021). Additionally, group members act as 'human smart sensors' collaboratively attributing trustworthiness by analyzing the congruence of verbal and nonverbal communication (Ho and

Benbasat, 2014). Evidence from computer-mediated communication supports this collective capacity to discern deception (Ho et al., 2017).

The perception of being deceived significantly impacts truth-tellers' interpersonal relationships with deceivers, as the violation of expectations plays a crucial role in diminishing trust. Expectancy Violations Theory (Burgoon, 1978) posits that individuals have established expectations for how others should behave based on social norms and past interactions. When these expectations are violated, it results in negative evaluations and decreased trust. Derived from Expectancy Violations Theory and Interpersonal Deception Theory, the Integrated Spiral Model of Trust (Burgoon et al, 2021) provides a framework for understanding how trust can spiral upwards or downwards depending on the ongoing interactions and perception. In collusive deception, deceivers manipulate information and craft false narratives, directly undermining the expectations of honesty and integrity crucial for trust. The unmet expectation not only damages the immediate trust but also contributes to a continued decline in trust during interaction. An adversarial group experiment demonstrates that even without knowing the identities of the deceivers, truth-tellers' peer ratings showed they trusted the deceivers less, and those ratings continued to decline as the experiment progressed. Conversely, the ratings for truth-tellers remained relatively higher and showed an upswing over time (Burgoon, 2021).

Given that truth-tellers can sense deception, which significantly impacts perceived trust and distrust, the following hypotheses are proposed:

H1a: Deceivers are likely to receive fewer verbal messages that indicate trust compared to truth-tellers.

H2a: Deceivers are likely to receive more verbal messages that indicate distrust compared to truth-tellers.

3.2.4 Deception Outcomes and Perceived Trustworthiness

Deception is a complex social activity, requiring deceivers to mask their true feelings and intentions while striving to appear credible and responsive to suspicion from others (Vrij, 1993; Burgoon & Bacue, 2003). In the context of collusive deception, the outcome of deceptive efforts significantly influences perceptions of trustworthiness among colluders. In instances of successful deception, deceivers are paradoxically perceived as more trustworthy than one might expect given their deceptive intent.

This phenomenon can be explained by the deceivers' ability to convincingly project an image of trustworthiness, effectively masking their deceptive intentions. Deceivers leverage various interpersonal skills such as persuasion, social adeptness, and acting capabilities to enhance their credibility. As credibility is a pivotal factor in determining the success of deceptive acts (Bond and DePaulo, 2008), when the narrative crafted by colluders is believed, it enhances their perceived reliability and sincerity to such an extent that truth-tellers may rate them as comparably trustworthy as non-deceivers.

To explore this dynamic, the study proposes two specific hypotheses:

H1b: The deception outcome alters the relationship between deception and perceived trustworthiness, such that the gap in perceived trustworthiness between deceivers and truth-tellers is smaller in cases of deception success.

H2b: The deception outcome alters the relationship between deception and perceived untrustworthiness, such that the gap in perceived untrustworthiness between deceivers and truth-tellers is smaller in cases of deception success.

3.2.5 Fight or Flight

Dunbar et al. (2014) proposed that deceivers adopt one of two strategies in deceptive interactions: a defensive "flight" strategy, where the deceiver acts more submissively, or an

aggressive "fight" strategy aimed at persuading truth-tellers to trust them. Substantial research indicates that the submissive approach is more commonly chosen by deceivers, especially in complex group interactions. This tendency may be due to the challenges and time required to build rapport in these settings, prompting deceivers to prioritize self-protection before adopting more assertive tactics (Zhou et al., 2014). Additionally, direct connections between all members are less common in larger groups (Forsyth, 2010). This structural aspect of larger groups further motivates deceivers to adopt a submissive strategy when interacting with multiple people.

Empirical findings suggest that deception is typically less verbally or nonverbally dominant—more submissive—compared to truthful communication. Specifically, in adversarial group games, deceivers are perceived as less dominant (Dunbar et al., 2021). Moreover, research supports that deceivers minimize information disclosure by deviating from direct responses and avoiding sticking to the topic. Consequently, their responses tend to be less frequent, less detailed, and generally indicative of withholding information (DePaulo et al., 2003; Zhou et al., 2014).

3.2.6 Solidarity or Secrecy

In the complex setting of collusive deception, deceivers often face a critical strategic decision: whether to openly support their co-conspirators, reinforce cooperative solidarity, or maintain distance to avoid detection by opting for strategic isolation. The choice between solidarity and secrecy in collusive deception can be explained through the dual-goal theory of deception, which involves both persuading the truth-teller to accept deceptive information and protecting oneself from detection (Buller & Burgoon, 1996). Cooperative solidarity represents an aggressive strategy that strengthens group cohesion and enhances the collective's ability to manipulate and control narratives, thereby making their deception more convincing to outsiders. On the other hand, strategic isolation is a more submissive approach where deceivers aim to avoid drawing attention,

adopting a low-profile demeanor to minimize behaviors that might attract attention or leak deceptive intentions.

Research supports a general preference for the submissive strategy in deceptive contexts, especially when the group dynamics are complex. This preference for strategic isolation helps manage risks associated with collective exposure, allowing individual deniability and safeguarding other deceivers in the group from being jeopardized. For instance, if one member is detected, strategic isolation ensures that others can remain undetected, thus maintaining a layer of security within the group. Furthermore, strategic isolation affords greater flexibility to deceivers. It enables them to adapt their tactics as needed without strong commitments to other group members, which is crucial in dynamic and uncertain environments. Additionally, in group settings, each member of the group loses some control over what information is provided and how each member behaves (Vernham et al., 2016). Deceivers must not only regulate their own information but also manage the impressions of others, close bonds could lead to unintentional leaks of inconsistencies. These bonds necessitate additional efforts to synchronize deceitful narratives, increasing the risk of exposure. Although it is anticipated that deceivers would favor strategic isolation, driven by a preference for a submissive strategy, this choice is still shaped by a combination of situational and psychological factors. These factors significantly influence how deceivers manage their interactions and express trust and distrust within the group.

3.2.7 Deceivers' Verbally Expressed Trust

Verbal expressions of trust are a powerful tool for influencing others' perceptions within interpersonal relationships. Verbal expression of trust facilitates advocating credibility and trust-building. Deceivers can use them strategically to manipulate perceptions and boost their credibility.

Verbal expression of trust influences the perceived trustworthiness of the trustee. When individuals verbally express trust towards others, they are not just communicating their feelings

but also advocating for the credibility of those they endorse. this can significantly enhance the perceived integrity and reliability of the trustee, especially within group settings where such endorsements can influence collective opinion (Kramer, 1999). For instance, when deceiver A expresses trust towards deceiver B, this can elevate the trustee's status in the eyes of truth-tellers, enhancing deceiver B's perceived trustworthiness and reliability.

Moreover, verbal expressions of trust can profoundly impact the interpersonal relationship between the trustor and the trustee. Expressions of trust have been found to have a reciprocal effect on trust levels, meaning that a trustor's trusting behaviors can stimulate the demonstration of trust from the trustee (Serva et al., 2005). Deceivers might leverage this mechanism to express trust towards truth-tellers, thereby fostering an environment of trust-building that gains positive opinions.

Based on the object of expressed trust from deceivers and the intensity of trust, there are four potential scenarios:

Deceivers express more trust toward truth-tellers: Deceivers may strategically express a higher level of trust toward truth-tellers to facilitate trust-building and reduce suspicion. This approach manipulates the perceptions of their honesty and reliability, aiming to integrate into the group of truth-tellers. By appearing trustworthy, deceivers can influence group dynamics and decision-making, thereby protecting and advancing their deceptive agenda.

Deceivers express less trust toward truth-tellers: In this approach, deceivers deliberately minimize their expressions of trust toward truth-tellers to maintain a cautious distance. This can serve to avoid attention that might reveal deceptive motives or to maintain a facade that aligns with their deceptive narrative. Furthermore, advocating for truth-tellers does not benefit deceivers

as it strengthens the trust status of those outside their colluding group, which offers no advantage to the deceivers' team.

Deceivers express more trust toward other deceivers: In this scenario, deceivers express increased trust towards their fellow colluders to strengthen their group's trust status. This strategy is influenced by the concept of "cooperative solidarity" in collusive deception, where colluders work closely to maintain a consistent and unified false narrative. In a group context, deceivers not only manage their own impressions but also those of their fellow colluders (Vernham, 2016). By endorsing another deceiver, they enhance the perceived credibility of the trustee—who is also a deceiver in this scenario. This internal validation among colluders can make their deceitful narrative more convincing to outsiders or truth-tellers, thereby strengthening the overall deceptive strategy.

Deceivers express less trust toward other deceivers: Conversely, deceivers might choose to express less trust toward their fellow colluders to manage risks associated with collective exposure. This strategy aligns with “strategic isolation” in collusive deception, aiming to preserve the appearance of integrity while pursuing hidden agendas. By maintaining a level of distance, they can preserve individual deniability and avoid implicating the entire group if suspicion arises. This approach allows them to hide their identity and "fly under the radar."

Following the previous discussion, deceivers often favor a submissive strategy characterized by strategic isolation in verbal communications. This approach is primarily driven by the need to manage risks and maintain operational secrecy within complex group dynamics.

Deceivers typically exhibit a cautious approach by expressing less trust within group interactions. This restrained trust expression is a strategic measure to prevent drawing attention to themselves, thereby reducing the risk of detection. By maintaining a passive profile in their

communications, deceivers avoid aligning closely with any specific group members, which helps preserve ambiguity about their intentions and alignments. Additionally, deceivers strategically differentiate their trust expressions so that they express even less trust toward fellow deceivers than toward truth-tellers. This controlled trust expression is crucial for maintaining individual deniability and minimizing the risks associated with collective exposure. Expressing some level of trust towards truth-tellers may serve as part of their deceptive strategy to integrate into and manipulate the group. However, displaying similar levels of trust toward fellow deceivers can increase the risk of exposing the entire group if any member becomes suspect or if their deception is uncovered. Additionally, openly bonding with another colluder reduces flexibility and necessitates extra effort to synchronize deceitful narratives, which could further jeopardize their position if inconsistencies are detected.

Given these dynamics, the following hypotheses are proposed:

H3.1: Compared with truth-tellers, deceivers are less likely to verbally express trust toward other group members.

H3.2: Deceivers are less likely to verbally express trust toward fellow deceivers than toward truth-tellers

3.2.8 Deceivers' Verbally Expressed Distrust

Just as expressed trust, verbal expressions of distrust are strategically employed in deception to influence perceptions within interpersonal relationships. While trust can build credibility and foster positive relations, distrust leads to distance, disengagement, or suspicion (Lewichi et al, 1998). Deceivers can strategically employ expressions of distrust to manipulate perceptions, aligning them with their goals of obfuscation and misdirection.

When deceivers express distrust, it strategically influences perceptions within the group. Firstly, expressing distrust can act as a protective mechanism, shifting suspicion towards others.

For instance, if deceiver A shows distrust towards truth-teller B, it might sow doubt among other group members about B's motives, benefiting deceiver A by maintaining an image of innocence. Secondly, by expressing distrust against someone perceived as suspicious by the group, the deceiver can appear as an ally, enhancing their perceived trustworthiness. This tactic, often summarized by the term "the enemy of my enemy is my friend," boosts the deceiver's trust status within the group by aligning with the group's attitudes.

These distrust expressions can also create barriers, reduce the reciprocity of trust, and foster an environment of general suspicion which can be advantageous for deceivers. By promoting distrust, deceivers might prevent the formation of strong bonds between group members, which can hammer collaborative efforts to uncover deception.

Based on the object of expressed distrust from deceivers and the intensity of such expressions, the following four scenarios emerge:

Deceivers express more distrust towards truth-tellers: By directing suspicion toward truth-tellers, deceivers can prevent them from gaining unanimous support and keep the group divided, thus easing their manipulative efforts. However, overly expressing distrust could backfire, undermining the deceivers' attempts to appear trustworthy and reliable, potentially raising alarm among other group members about their deceitful intentions.

Deceivers express less distrust towards truth-tellers: Conversely, minimizing expressions of distrust towards truth-tellers might be used to blend in or align with group norms if excessive suspicion might seem out of place. This can be a calculated attempt to avoid raising alarms while maintaining their deceptive stance uncovered.

Deceivers express more distrust towards other deceivers: In this scenario, deceivers express increased distrust towards their fellow colluders to position themselves as independent and

unbiased, which helps in maintaining strategic isolation. This can enhance their perceived trustworthiness among truth-tellers by appearing less collusive. However, it may jeopardize the deceiver's strategic position, threatening the cohesion and effectiveness of their collaborative deceptive efforts.

Deceivers express less distrust towards other deceivers: Conversely, by refraining from showing distrust toward fellow deceivers. This approach minimizes the risk of drawing attention that could reveal their deceit.

In discussions about strategic behavior among deceivers, expressing less distrust aligns with a preference for a submissive strategy. Deceivers aiming to maintain a low profile within complex group dynamics opt for this approach to reduce the risk of detection and avoid attracting attention.

We previously discussed why deceivers favor strategic isolation over a collective solidarity strategy; however, this may not hold for expressing distrust. Although trust and distrust are conceptual opposites, they function as separate mechanisms within interpersonal relationships (Lewicki et al., 1998). Consequently, expressing more distrust does not equate to the effects of expressing less trust toward other deceivers. Distrust carries a negative affection that can create emotional tension and often a frantic demeanor, distinguishing it sharply from trust (Luhmann, 1979). Expressing trust aligns with social norms and is generally seen as positive and cooperative, making it less likely to arouse suspicion within a group. In contrast, expressing distrust or skepticism is often counter-normative and can signal conflict or division, which might draw unwanted attention and go against the submissive strategy. Moreover, blaming a fellow deceiver to boost one's own trust status requires not only a deep understanding of group interactions but also sophisticated manipulation skills. Thus, the potential risk of exposure and the high demand

for social skills make expressing more distrust toward fellow deceivers a less favored strategy among deceivers who prefer to maintain a more reserved and less confrontational facade.

Given these dynamics, the following hypotheses are proposed:

H4.1: Compared with truth-tellers, deceivers are less likely to verbally express distrust toward other group members.

H4.2: Deceivers are less likely to verbally express distrust toward fellow deceivers than toward truth-tellers.

3.2.9 Social Networks in Collusive Deception

Social network analysis serves as a robust framework for dissecting complex social interactions and phenomena. This analytical approach is premised on the idea that relationships among entities can provide deeper insights than individual characteristics alone. A social network comprises two types of key components – actors(nodes) and ties(links). Actors, represented as nodes, can be individuals, groups, or organizations, while ties, shown as links between these nodes, represent interactions, flows of information, affiliations, or social relations (Wasserman and Faust, 1994). Network analysis employs a suite of techniques to map these interactions and examine the structures that emerge from them. A fundamental premise of social network theory is that the structure of these relationships significantly influences social outcomes (Borgatti et al., 2009).

In the context of collusive deception, network analysis becomes a potent tool for unraveling the complex web of interactions and influences. Deception involves dynamic social interactions where deceivers and truth-tellers influence each other, either directly or indirectly. By employing network analysis, researchers can identify patterns of deception by examining how deceivers interact with, influence, and are influenced by others within the network (Pak and Zhou, 2014). In this study, deceivers and truth-tellers are treated as social actors (nodes), with their verbal

expressions of trust and distrust conceptualized as ties (links) within the network. These relational messages reflect the ongoing iterative processes of cognitive and behavioral adjustments among group members. As discussed previously, each verbal relational message has the potential to alter perceptions of trustworthiness and credibility among group members, thereby influencing the social network structure and trust dynamics within the group.

In this study of collective deception, we will employ a heterogeneous signed directed network to model the interactions and dynamics within the group. This network includes two distinct types of nodes—deceivers and truth-tellers—allowing for detailed analysis of interactions both within and across these groups. The links between nodes are signed to indicate the nature of the relationships, with positive and negative signs reflecting expressions of trust or distrust, respectively. This signed network feature is crucial for capturing the nuanced contrasting sentiments in communications. Additionally, the links are directed, clearly indicating the direction of communication from the sender to the receiver, which illustrates the flow of influence and information within the network.

Social network analysis provides a robust framework for studying the patterns of relationships among nodes at various levels—individual nodes, communities, and the entire network. By examining these interactions, it becomes possible to trace how collusive deception impacts the interactive behaviors of group members, observing the strategies that colluders may employ.

3.2.10 Node-Level Network Metrics

The degree of a node in a network indicates the total number of connections it has. Within our study's context, we distinguish node degrees into four specific types based on the direction of links (inbound and outbound) and the nature of the relationships they represent (trust or distrust).

Positive In-Degree measures the number of trust expressions directed at a node from other nodes in the network. This metric reflects perceived trustworthiness; a higher positive in-degree indicates that the actor is widely trusted by other group members. According to our H1a, we anticipate that deceivers will typically exhibit a lower positive in-degree than truth-tellers. When deception is successful, the gap in positive in-degree between deceivers and truth-tellers is expected to narrow, as deceivers successfully project an image of trustworthiness that convinces other group members (H1b).

Negative In-Degree counts the expressions of distrust directed at a node. It signifies perceived distrust and quantifies how often an actor is openly doubted by their peers. For deceivers, we expect a higher negative in-degree compared to truth-tellers, which aligns with hypothesis H2a. This suggests that deceivers are more frequently suspected of deceptive behavior by other group members. Additionally, under hypothesis H2b, successful deception leads to a smaller gap in negative in-degrees. This is because if the deception goes undetected, fewer group members suspect the deceivers, leading to fewer expressions of distrust directed at them.

Positive Out-Degree represents the number of trust expressions an actor directs toward others. This metric can indicate an actor's strategy to build alliances or persuade others by verbally addressing trust. According to our hypothesis H3.1, we hypothesize that deceivers, due to their preference for a submissive strategy, may exhibit a lower positive out-degree compared to truth-tellers. This could reflect their cautious approach in not overtly expressing trust that might expose their deceitful intentions.

Negative Out-Degree involves the number of distrust expressions an actor communicates towards others. This measure is critical for understanding how actors may engage in creating divisions or maintaining a façade of skepticism to disguise their true intentions. In line with

hypothesis H4.1, we propose that deceivers will exhibit a lower negative out-degree compared to truth-tellers. This is because expressing high levels of distrust could draw unwanted attention and scrutiny towards the deceivers, going against their need to maintain a low profile and avoid detection.

The network degree metrics described above are summarized in Table 2. These measures provide a quantitative framework to assess and compare the relational dynamics of trust and distrust among deceivers and truth-tellers within the network.

Network Degrees		Relational Messages	Network Notion	Related Hypotheses
The number of incoming connections a node has	Positive in-degree	Perceived Trust	IN+	H1a, H1b
	Negative in-degree	Perceived Distrust	IN-	H2a, H2b
The number of outgoing connections a node has	Positive out-degree	Expressed Trust	OUT+	H3.1
	Negative out-degree	Expressed Distrust	OUT-	H4.1

Table 2. Network Degree Metrics in Collusive Deception Network

3.2.11 Community-Level Network Metrics

Community-level network metrics play a crucial role in understanding the broader patterns of interaction that occur within and between distinct groups or communities in a network (Kadushin, 2004). These metrics shed light on the density of connections both within a community (intra-community) and between different communities (inter-community), offering significant insights into the dynamics of collusive deception.

Intra-community connections refer to the interactions and relationships that occur within a specific group or community, such as among members of the same team or organization. A high density of these connections among deceivers could suggest a strong sense of cooperative solidarity, characterized by tightly knit interactions that support and reinforce each other's

deceptive narratives. Conversely, a lower density might indicate strategic isolation, suggesting that deceivers are employing tactics to maintain secrecy and minimize direct associations within their group.

Inter-community connections, on the other hand, involve interactions between members of different groups or communities. Analyzing the interactions that occur between deceivers and truth-tellers can provide insights into how deceivers bridge the gap between communities to spread their influence and deception. Deceivers, often guided by a preference for a submissive strategy, might strategically limit their connections toward truth-tellers to reduce their visibility and susceptibility to detection.

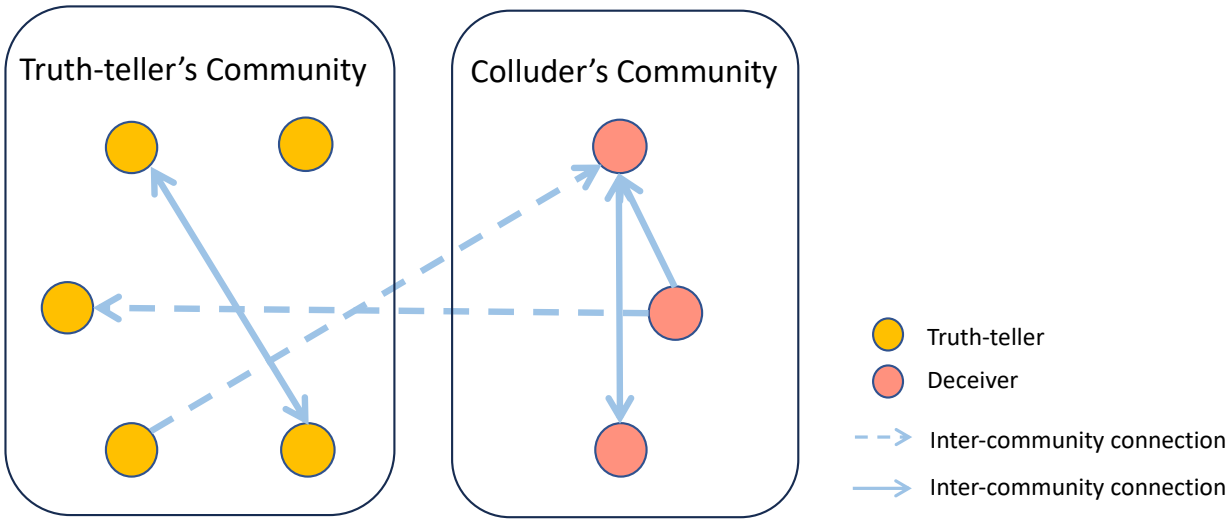


Figure 3. Community-Level Connections in Collusive Deception Networks

Figure 3. illustrates the intra-community and inter-community connections within the collusive deception network. These community-level metrics are instrumental in testing hypotheses related to how deceivers manage their expressions of trust and distrust across different communities.

H3.2 states that deceivers are less likely to verbally express trust toward fellow deceivers than toward truth-tellers. This indicates that deceivers exhibit less trust within their own

community compared to interactions with members outside their community, emphasizing a strategic preference to maintain distance within the deceptive community. Consequently, it is anticipated that deceivers' intra-community trust connections will be less than their inter-community trust connections with truth-tellers. H4.2 suggests that deceivers are less likely to verbally express distrust toward fellow deceivers than toward truth-tellers, reflecting a strategic preference to avoid conflict and minimize risks of exposure. Therefore, it is anticipated that deceivers' intra-community distrust connections will be less than their inter-community distrust connections.

3.2.12 Social Network and Deception Detection

Network analysis of trust has been popularly employed by the computer science community in detecting malicious individuals or behaviors within a large network. Utilizing digital footprints such as votes, reviews, and social tags, these methods construct predictive networks to differentiate trustworthy from untrustworthy individuals. Early studies by Kamvar et al. (2003) pioneered algorithms that computed user trustworthiness based on peer voting, effectively isolating malicious users in peer-to-peer file-sharing networks. Guha et al. (2004) further advanced these algorithms to consider both trust and distrust ratings, successfully identifying deceptive patterns in e-commerce consumer review datasets. Ortega et al. (2012) extended this work by constructing social networks from online discussion platforms, identifying users disseminating false content to manipulate or provoke readers.

Current network applications for detecting malicious intentions primarily focus on large online groups with a vast number of nodes and links, often numbering in the thousands. However, this approach may not be suitable for small group contexts, where interactions are much denser. Understanding the networks of small groups is important as it closely mirrors real-life scenarios of conversations and organizational decision-making processes. In smaller networks, such as those

found in tightly-knit communities or organizational teams, every interaction and relationship holds greater significance due to the closer proximity and frequency of engagement. This level of intimacy can amplify the impact of trust or distrust dynamics, making it essential to study how these networks operate and how deceptive behaviors manifest within them. Additionally, the literature in the computer science domain has predominantly emphasized improving computational performance rather than hypothesis testing and theory development (Tang et al., 2016). Therefore, there exists a research gap in applying network analysis techniques to small group settings for the purpose of understanding and detecting malicious behaviors.

Study	**Dataset**	**Verbal Feature**	**Main Classification Model**	**Reported Accuracy**
Newman et al., 2003	Five topic discussion in face-to-face communication transcribed as text data	Linguistic features	Logistic Regression	48.0% - 67.0%
Zhou et al., 2004	Two experimental studies (dyadic online chatting)	Linguistic features	Discriminant analysis, Logistic regression, Decision tree, ANN	70.5% - 79.2%
Fuller et al., 2011	Statements officially reported and written by a subject or witness in an official investigation	Linguistic features	KNN, SVM, NB	65.9% - 73.9%
Wu and Liu, 2020	Online adversarial group game (Chinese)	Discourse features (Human coded)	SVM, NB, C4.5, ANNs	61.84% - 80.89%
Chen, 2023	In-person adversarial group game	Content-based linguistic features (Human coded)	Logistic Regression, Random Forest, Gradient Boosting	68.6% - 69.6%

Note: ANN: Artificial neural networks; KNN: K-Nearest Neighbors; SVM: Support vector machine; NB: Naive Bayes

Table 3. Deception Detection Classification Models with Verbal Features

Despite the widespread application of network analysis and network models across various domains in computer science, their utilization for studying interactions within small groups, particularly for deception detection, remains notably rare. This gap is partially addressed by Bai et

al. (2021), who innovatively apply the seminal network analysis algorithm, PageRank, to nonverbal communication data from the Resistance Game. Bai et al. developed the DominanceRank algorithm, which effectively sorts group members based on the nonverbal signals they transmit and receive, demonstrating high accuracy in identifying the most dominant group member. Building on this, Kumar et al. (2021) achieved exceptional accuracy in detecting spies within the game, with eye contact and speaking activities serving as critical nonverbal edges within their network models.

Regarding verbal cues, extensive research has employed traditional machine learning models—such as Logistic Regression, SVM, ANN, KNN, and Decision Tree—to analyze verbal features for deception detection. Each of these models offers unique insights into detecting deceptive behavior, as detailed in Table 3. However, the integration of verbal features into network models remains notably limited. Notably, Wu and colleagues (2015) stand out by extending the discourse analysis from an online Mafia game into a temporal social network, termed DOBNet. Their research analyzes the impacts of deception intent on various network structures and utilizes these network features to predict deception (Wu and Liu, 2020). However, their method constructs networks based solely on individual's discourse, with the discourse as nodes and interactions or links created based on temporal sequences. It does not account for interactions between group members, nor does it leverage more advanced models such as Graph Neural Networks (GNN), which could potentially exploit the full capabilities of network models in detecting deception.

Graph Neural Networks (GNNs) present a transformative approach in network modeling, which is particularly beneficial in the domain of deception detection within social interactions. GNNs excel in encapsulating relational dependencies inherent in communication data, which is pivotal in scenarios involving deceptive behavior. Unlike conventional machine learning models,

GNNs effectively leverage the interconnectivity between nodes in a network, thus offering a more intricate comprehension of social dynamics (Zhou et al., 2020). This is particularly valuable when analyzing group interactions in deceptive contexts. The implementation of GNNs holds the promise of markedly enhancing the performance of deception detection, overcoming the limitations of prior deception detection classification models that use verbal features.

3.2.13 Advanced Language Models

The study of verbal relational messages demands a nuanced level of analysis that surpasses the capabilities of traditional natural language processing (NLP) techniques. Specifically, understanding these messages requires a pragmatic or content-level approach that considers the context, implied meanings, and relational messages conveyed through language. Traditional NLP methods, which often focus on syntax and surface-level semantics, fail to capture the full depth of interpersonal communications where much of the meaning is context-dependent and not explicitly stated.

Previously, the analysis of pragmatic or content-level in verbal communication has relied heavily on human annotation (Zhou et al, 2014; Chen 2013). This process requires annotators to interpret and code messages based on their understanding of the context and the subtle cues within the communication. While this method allows for a detailed understanding of the pragmatic layers of communication, it is labor-intensive, subject to human bias, and not scalable.

The emergence of advanced language models, such as GPT (Generative Pre-trained Transformer), offers a transformative approach to this challenge. GPT's deep learning foundation, trained on extensive text corpora, equips it with an almost human-like understanding of language nuances, context, and implicit meanings (Vaswani et al., 2017). This advancement enables contextual understanding, where GPT excels at interpreting relational dynamics from communicative cues missed by traditional models. Scalability is another key advantage, as GPT

can efficiently process large-scale datasets and support real-time applications. Additionally, the diverse linguistic training of GPT contributes to reduced bias, decreasing the common biases associated with manual annotation, thus enhancing the reliability and objectivity of the analysis.

Incorporating GPT into the analysis of verbal relational messages represents a significant advancement over traditional NLP methods. By using this technology, researchers and practitioners can gain insights into the complex dynamics of interpersonal interactions more accurately and cost-effectively, thus opening new avenues for understanding relational communication.

3.3 Methods

3.3.1 The Resistance Experiment

To explore the dynamics of collusive deception within group interactions, we initiated a large-scale, international experiment, engaging participants in an adversarial group game based on the board game "Resistance." This format, along with similar settings like the Pirate Game, Mafia Game, and Werewolf, has been widely used to study deception in group interactions (Zhou, 2013) and to analyze relational messages (Burgoon et al., 2021; Spitzley et al., 2022). These game-based frameworks offer a rich context for examining complex group dynamics and the role of deception in interpersonal relations.

Participants were grouped into sessions, each with 5 to 8 players. Roles were assigned at random, dividing players into spies and villagers, who were considered as deceivers or truth-tellers, respectively. The game involved several rounds, each centered around a mission. These missions required the election of a leader and the formation of a team, decided through a group vote. The outcome of each mission hinged on secret votes from team members, with deceivers having the option to vote against the mission's success. Following each mission, players were given a chance

to discuss why they thought the mission resulted in the way it did and to guess other players' identities. Villagers won the round by figuring out who the spies were and excluding them from the mission teams. Spies won by causing mission failures. The game's ultimate winners (spies or villagers) were determined by which team won the most rounds. A detailed description of the experiment protocols can be found in Dorn et al. (2021). The game procedure is shown in Figure 3.2 For 5 or 6-player games, we randomly assigned 2 spies; for 7 or 8-player games, 3 players were randomly assigned as spies. The dataset includes 46 games collected at 7 locations across six regions, including the U.S.A, Hong Kong, Israel, Fiji, Zambia, and Singapore.

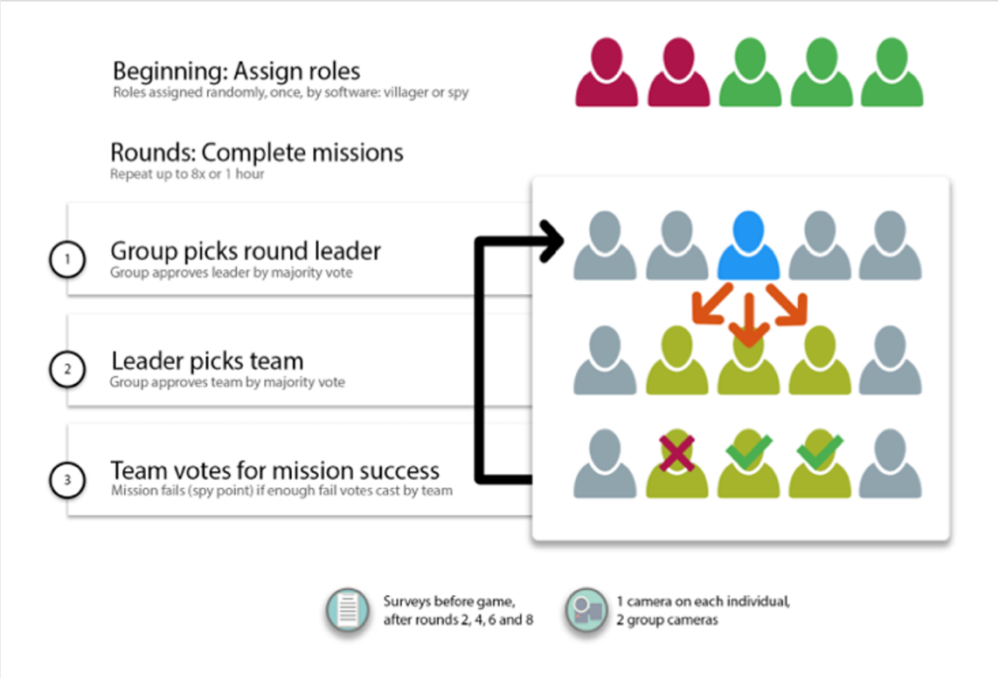


Figure 4. Resistance Game Procedure

To capture the nuances of group dynamics more accurately, the experimental setup was designed to collect extensive data, including audio, video, and participants' responses to surveys. Figure 5. demonstrates the arrangement of the experiment where participants were seated in a circle, each with a tablet in front of them. These tablets were used to collect survey responses and record the participants' audiovisual signals. The experiment began with a pre-survey and a practice

round to familiarize participants with the rules. Subsequent in-game surveys were conducted every two rounds to capture participants' assessments of others' trustworthiness, likability, dominance, and nervousness, allowing a comprehensive understanding of the interpersonal dynamics in the game.

Figure 5. Resistance Game Experiment Setting

The experimental scenario is instrumental in examining collusion and relational messages within group dynamics, providing a structured environment that closely mimics strategic contexts found in military, government, and corporate settings. This design is particularly effective in studying the subtle nuances of collusive behavior and communication.

First, it creates an environment suitable for collusion, with multiple deceivers strategically hiding in the group. They engage in covert deception for shared gain, operating under secrecy to sabotage the goals of the larger group. This satisfies the conditions necessary to explore the concept of collusion as defined by our research questions. Second, the design of the experiment promotes extensive interaction, which is important for analyzing relational messages. Beyond the structured discussions at the end of each round, players have the autonomy to express their opinions

continuously throughout the game, thereby providing a multitude of verbal and nonverbal cues for analysis. Third, the careful monitoring of in-game decisions and outcomes affords a granular view of how collusion affects individual choices and group decisions. This careful observation helps capture the essence of collusive interactions and their impact on the group, enhancing our understanding of relational communication.

3.3.2 Utterance Coding Taxonomy

Lower-level verbal features such as syntax and surface-level semantics often fail to capture the full depth of interpersonal communications, which are crucial for understanding relational messages. To address this limitation, human coders were employed to annotate the data, allowing for a richer analysis that takes into account the nuances of verbal cues.

In this study, we utilized a refined coding system informed by methodologies from seminal works by Chen (2023) and Zhou's team (2014). These studies have proven the robustness of the taxonomy in content-based language analysis within deception research, utilizing online chat rooms and face-to-face interactions across different languages.

Drawing on Chen's approach, the objective was to identify discourse features that explicate the trust-distrust relationships between speakers. Two coders were independently selected to analyze the features listed in the existing taxonomy (Appendix B). Discourse elements such as Agreeing, Disagreeing, Trusting, Distrusting, Accusation, Nomination, Question, and Justification from the original framework were selected. These discourse elements inherently have an object (from one player to another) and represent sentiments of trust-distrust. After a thorough evaluation of the transcripts, the categories of Question and Justification were narrowed down to Challenging and Defense, respectively, to represent the dynamics of trust and distrust more accurately. Details of this coding system are outlined in Table 4.

Relational Message Type	Label	Description/Definition	Illustrative Examples
Trust	Trusting Statement	Utterances that express trust towards another participant or a mission team	Yeah, I trust you. He seems trustworthy. I think she's a villager.
	Nomination Statement	Utterances that propose or endorse someone to be a team leader	I nominate player one. I want you to be the leader.
	Defense Statement	Utterances that support or justify someone's actions in response to doubts, distrust, or accusations	she's just an ordinary citizen so she can work to succeed the mission. He also didn't know…
	Agreeing Statement	Utterances that express agreement, acceptance, compliment, or willingness to cooperate	Yeah, I agreed to that… That's also what I, I was goanna say that. True.
Distrust	Distrusting Statement	Expressing suspicion about a person's motives or honesty, implying a belief they may be misleading or deceptive.	You're acting sketchy… I don't trust… You just like to watch the world burn Julie, you're the odd girl out pretending.
	Accusation Statement	Utterances of direct allegations made against a participant, accusing them of deception, being a deceiver or being guilty of a failed mission explicitly	Your both deceivers That makes it more sense that you are a deceiver
	Challenging Statement	Directly confronting the logic or evidence behind a statement or action, indicating skepticism.	I don't get it if you think I am a spy why would you take me with you on the mission. Now you're talking a lot, in the beginning you didn't talk too much. He just didn't give a good enough reason…
	Disagreeing Statement	Utterances that express disagreement or unwillingness to cooperate	No, I'm against … I don't agree… No, you can't. You can't do that.

Table 4. Utterance Coding Taxonomy

The tagging approach at the utterance level, as used in previous work, was also adopted. Text segments are bounded by shifts in speakers and are most suitably analyzed at the utterance level (Brown and Yule, 1983). Given that an utterance can convey multiple meanings or sentiments, it can be tagged with multiple labels.

This independent coding was conducted to fortify the integrity of the analysis. Post-coding, the annotations were cross-examined and merged to compile a dataset for analysis. This coding system is particularly effective in highlighting the subjects discussed and the nature of the relationships between speakers, which are crucial to the investigation of relational communication.

3.3.3 Network Construction

In continuation from the previous section, a heterogeneous directed unweighted signed social network was constructed for each game to analyze the verbally expressed trust and distrust among players. Each node within the network represents a player, and the nodes are heterogeneous, embodying two distinct classes: spies and villagers, or deceivers and truth-tellers. These are visually distinguished by red and yellow colors in Figure 6., respectively. Links within the network are created from the utterance coding labels; a trust link, depicted in gray, indicates a positive relationship and occurs when a player verbally expresses a trust statement, which includes trusting, nomination, defense, and agreeing. Conversely, a distrust link, represented in red, is established from one player to another when expressions of distrusting, accusation, challenge, or disagreeing are made. For instance, in the Israel Game number 009, when player 4 nominated player 1 as the leader (as noted in transcript line 1 of Figure 6.), a trust link was connected from player 4 to player 1. Similarly, when player 2 openly accused player 5 of being a spy (transcript line 4), a distrust link was drawn from player 2 to player 5.

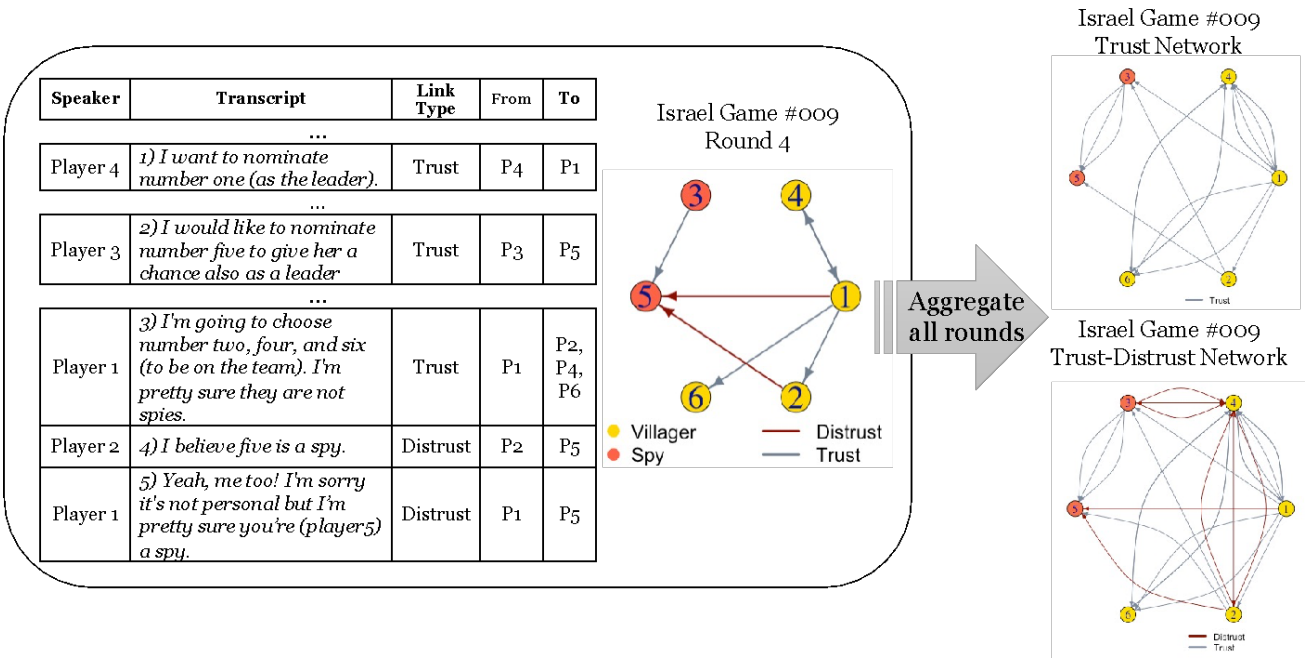


Figure 6. Network Visualizations

Given the subjective nature and challenge in quantifying the intensity of each trust and distrust utterance, the model simplifies by assigning a positive sign to all trust statements and a negative sign to all distrust statements, without differentiating by weight. To avoid redundancy, multiple utterances from the same player expressing the same sentiment towards another player were consolidated, and duplicate links within the same round were removed. Trust and distrust links are documented at the round level, while game-level links are created by aggregating observations across all rounds. Self-referential links were also omitted from the network.

3.3.4 Regression Configurations

Regression analysis was conducted using a multivariate fixed-effect model to test hypotheses H1a and b, H2a and b, H3.1, and H4.1, which study perceived trust and expressed trust within the group. As discussed in 3.2.10, perceived trust was operationalized as node's positive in-degree, and perceived as negative in-degree; similarly, express trust was measured by positive

out-degree, and expressed distrust by negative out-degree. These measures were set as dependent variables.

The independent variables included the player's role and the game outcome. The player's role is a binary variable, denoted as "spy" or "villager." The group of spies represents the colluders, engaging in collective action to sabotage the interests of the larger group. A participant was coded as Role = 1 if they were a spy, and Role = 0 if they were a villager. Deception success, drawing from prior studies (Zhou et al. 2013), was operationalized based on whether the spy team successfully deceived others and won the game. Thus, the game outcome was a binary variable indicating if the spies or the villagers won the most rounds within the designated time frame.

Control variables included participants' gender, language (whether English as the native language), and prior experience with similar games. These variables were included as potential influencers of trust dynamics among players. Gender perceptions are particularly noted; female are often perceived as less deceptive than male, and male are generally considered less trustworthy (Boltz et al. 2010). And male are sometimes viewed as more likely to express trust than female (Buchan et al., 2008). Additionally, familiarity with the game was considered. As the game setting is complex, prior experience could provide players with a significant advantage in understanding the rules and nuances, potentially affecting in-game strategies and trust dynamics. Past research indicates that experience with deception can improve the success rate of deceivers due to learning abilities (Zhou et al. 2013). Language is also a significant factor, as non-native speakers are often perceived more negatively compared to native speakers, which could affect trust dynamics (Evans and Michael 2014).

Fixed effects for each game were included in the model to account for unobserved heterogeneity that could influence the analyses. The model specification is summarized as follows:

$$\text{Node degrees}_i = Role_i\ (\text{Spy} = 1) + \text{Game Outcome}_i\ (\text{SpyWin} = 1) + Role_i * \text{Game Outcome}_i + Gender_i\ (\text{Male} = 1) + Native\ English\ Speaker_i\ (\text{Yes} = 1) + \text{Game}\ Experience_i\ (\text{Yes} = 1) + \varepsilon_i$$

3.3.5 Community Connection Score Comparison

To assess the differential trust and distrust interactions within and between defined communities, inter-community and intra-community connection scores were computed at the network level for each game. The scores were normalized to account for the varying numbers of spies and villagers in each group. The formulas for the scores are as follows:

$$\text{Inter} - \text{community connection score} = \frac{Number\ of\ connections\ from\ Community\ A\ to\ Community\ B}{\text{Number of members in community B}}$$

$$\text{Intra} - \text{community connection score} = \frac{Number\ of\ connections\ within\ Community\ A}{\text{Number of members in community A}}$$

A t-test was then used to compare the difference between the connection scores for the following:

H3.2: Spy inter-community trust connection score vs. Spy intra-community trust connection score

H4.2: Spy inter-community distrust connection score vs. Spy intra-community distrust connection score

3.3.6 Predictive Models

This study explored the effectiveness of predictive models in distinguishing deceivers within groups by analyzing their verbal behaviors and interaction patterns. To assess the predictive power of network features, two distinct sets of features were utilized: Feature Set A with Player features only and Feature Set B with Player features and Network features.

Feature Set A included basic demographic and experiential attributes such as gender, whether the player is a native English speaker, player's game experience, and the cultural homogeneity of the group (homogeneous or heterogeneous). Feature Set B combined player

attributes from Feature Set A with additional network metrics, such as positive and negative trust-based connections (both inbound and outbound), along with network rankings like PageRank, hits, prestige, and received trust. Previous research (Ge, 2022) has indicated that network ranking scores can significantly differentiate truth-tellers from deceivers. The equation for calculating network ranking scores is detailed in the Appendix C and Appendix D.

The models evaluated in this research included Logistic Regression, Gradient Boosting Machine, Neural Network, and Graph Neural Network (GNN). The GNN is particularly noteworthy for its use of SignedGCNLayer components, which are designed to separately process positive and negative trust-based links before integrating them. This sophisticated architecture enables the GNN to effectively capture the complex interdependencies within the network data, significantly enhancing its capability to identify subtle patterns of deceptive interaction. The structure of the GNN model is detailed in Appendix E. Each model was assessed using both feature sets under a 5-fold cross-validation scheme to ensure that each data point contributed to validating the models' efficacy. The performance metrics included accuracy, precision, recall, and F1 score.

3.3.7 Automated Utterance Coding Using GPT Models

To enhance the efficiency of predictive models concerning trust and distrust in verbal communications, our study employed GPT models to automate the utterance coding process. This approach aimed not only to evaluate the impact of GPT-generated labels on model performance but also to explore the feasibility of developing an end-to-end predictive model that improves efficiency.

To accurately capture expressions of trust and distrust in verbal communications, we developed and refined GPT prompts, ensuring sensitivity to the nuances of trust-related sentiments. The study evaluated several variants of Generative Pre-trained Transformer (GPT) models for sentiment analysis, including GPT-4, GPT-4-1106-preview, GPT-4-0125-preview, GPT-3.5-

turbo-instruct, GPT-3.5-turbo-0125, and Davinci-002. Performance was primarily measured by the percentage overlap with manual tags, which served as a benchmark for comparing automated and human-generated coding accuracy.

Cost-effectiveness was also considered, with costs calculated based on the tokens used and their associated prices. The comparative analysis highlighted GPT-4 as the most accurate model, followed by its subsequent versions. However, for a balance between cost and performance, GPT-3.5-turbo-instruct emerged as the preferred model, with 1000 input tokens and 1500 output tokens. Details on the prompt configuration are available in Appendix F.

The models achieved an overlap of approximately 75.6% with manual labels, indicating significant congruence and affirming the effectiveness of the tuned GPT prompts in replicating human-like coding accuracy. Moreover, the total cost for tagging transcripts of 46 games was under $3, and the process took less than three hours—a stark contrast to the weeks required for manual transcription and annotation. This dramatic increase in efficiency underscores the potential of GPT models to streamline complex linguistic data processing in social science research, paving the way for more sophisticated and expedient predictive modeling.

3.4 Results

3.4.1 Summary Statistics

The dataset documents 46 experiments involving 324 participants with a mean age of 22.47. Among them, 127 were assigned as deceivers and 197 as truth-tellers, with males representing 60%. Half had experience with similar games (54%), and 53% were native English speakers. Spy won in 20 of the 46 games. Games lasted 3 to 8 rounds, with a mean of 5.43 rounds. The typical game had 5 to 8 players, a mean of 7.

The mean positive in-degree was 4.257 with a standard deviation of 2.836, while the mean negative in-degree was 4.713 with a standard deviation of 3.613. The positive out-degree had the same mean as the positive in-degree, and the negative out-degree matched the negative in-degree, with standard deviations of 3.550 and 4.148, respectively.

The 47 experiments generated a total of 16,851 utterances, with participants contributing 13,794 and facilitators adding 3,057. Every participant spoke at least once. However, a portion of the data was rendered unintelligible due to factors such as crosstalk, whispers, pauses, topic changes, and ambiguous word connections. These challenging utterances were omitted from the analysis, as agreed upon by two coders. From the intelligible subset, 3,601 utterances were labeled. Manual labeling efforts resulted in the identification of 4,607 links, comprising 2,100 trust-related and 2,507 distrust-related connections.

Moreover, GPT analysis initially flagged 7,195 utterances with raw labels indicating trust or distrust sentiments. Subsequent data cleansing to remove inaccuracies left 5,825 valid links, with 2,851 signifying trust and 2,974 indicating distrust.

3.4.2 Perceived Trust

Table 5 column 1 presents the regression results regarding perceived trust within the game settings. The findings indicate that spies generally receive fewer trust verbal messages from other players, thus supporting Hypothesis 1a. This result aligns with the expectation that deceivers (spies) are less trusted verbally by their peers compared to truth-tellers (villagers).

Additionally, the results also lend support to Hypothesis 1.b, which posits that the gap in perceived trustworthiness between deceivers and truth-tellers narrows in cases of deception success. Specifically, while spies typically receive significantly less verbal trust compared to villagers (-1.878, $p<0.001$), this gap narrows in games where spies win. The combined effect of being a spy and winning reduces the trust deficit to approximately -0.822, as the interaction term

(1.428, p<0.001), offsets the negative trust associated with being a spy. This suggests that successfully manipulating trust dynamics within the group may provide deceivers with an advantage in achieving their agenda.

	Perceived Trust	Perceived Distrust	Expressed Trust	Expressed Distrust
	(1)	(2)	(3)	(4)
Spy	**-1.878*** **	0.33	**-1.803*** **	-0.66
	(0.311)	(0.397)	(0.463)	(0.514)
SpyWin	-0.372	1.467*	-0.566	**1.599** **
	(0.612)	(0.752)	(0.640)	(0.783)
Spy:SpyWin	**1.428*** **	-0.189	**1.761** **	-0.387
	(0.476)	(0.608)	(0.708)	(0.786)
Male	0.267	1.190***	1.436***	1.259***
	(0.253)	(0.324)	(0.372)	(0.415)
NativeEngSpeaker	1.054***	1.510***	1.466***	1.442***
	(0.299)	(0.383)	(0.417)	(0.478)
GameExperience	0.990***	0.159	0.329	0.701
	(0.286)	(0.365)	(0.401)	(0.457)
Group Effect	0.765***	0.687***	0.254***	0.361***
	(0.209)	(0.192)	(0.091)	(0.116)
Intercept	3.710***	2.649***	3.456***	2.748***
	(0.476)	(0.593)	(0.552)	(0.659)
Hypothesis	H1a&b	H2a&b	H3	H4
Support or Not	Yes	No	Yes	No
Observations	324	324	324	324
Residual Std. Error	1.928 (df=317)	2.469 (df=317)	2.922 (df=317)	3.223 (df=317)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 5. Regression Results of Node Degrees

3.4.3 Perceived Distrust

The results from the analysis do not support hypotheses H2a and H2b (Table 5, column 2). Contrary to expectations, neither the act of deception nor the interaction between deception and its outcomes significantly alters the perceived untrustworthiness derived from verbal relational messages among players. Interestingly, the deception outcome stand-alone is significantly related to perceived distrust in the group. In games where the spies win, there is a notable increase in the verbally conveyed distrust among players.

3.4.4 Verbally Expressed Trust

The results presented in Table 5 column 3 support Hypothesis H3.1, suggesting that spies are less likely to verbally express trust towards other players than truth-tellers, indicating that deceivers might adopt a submissive strategy in their trust behaviors. Additionally, the analysis of the spy's intra-community trust connection score (mean 0.09) compared to their inter-community trust connection score (mean 1.64), as illustrated in Figure 7., and the subsequent pairwise t-test ($p<0.01$) support Hypothesis H3.2. This result implies that deceivers express trust towards fellow deceivers less than towards truth-tellers which suggests that deceivers prefer strategic isolation over cooperative solidarity when using relational messages to convey trust. Additionally, the villager intra-community trust connection score, which is significantly higher than that of the spies (mean 3.15), serves as a benchmark for authentic trust interactions. This disparity further indicates that the deceiver community is significantly more loosely connected compared to the truth-tellers' group. This suggests a strategic positioning where deceivers maintain isolation both internally and from those they seek to deceive, aiding in masking their deceptive activities and maintaining low-key in their social interactions.

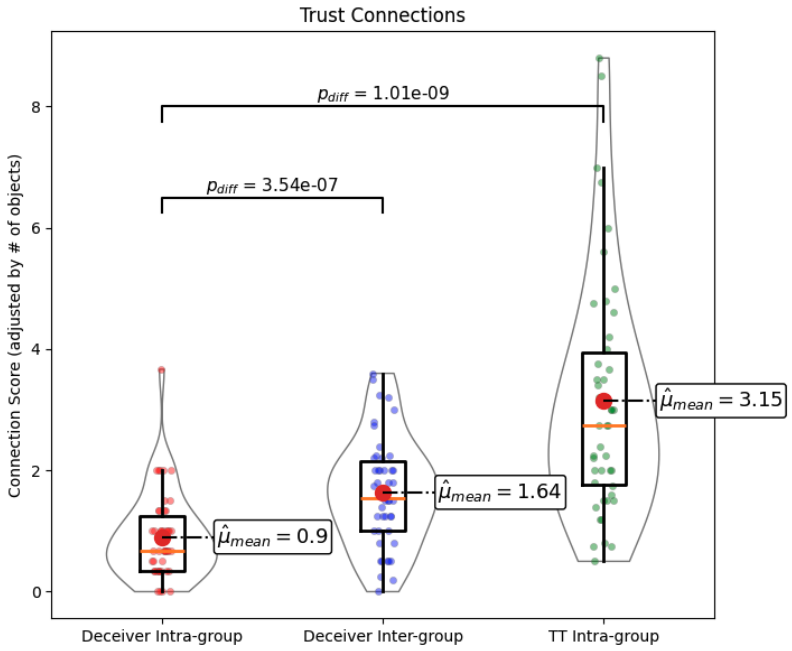


Figure 7. Comparison of Community Trust Connection Score

The role and deception outcome interactions reported in Table 5. are statistically significant, indicating that in games where spies win, they express trust more frequently than usual. Further analysis into expressed trust towards villagers and spies, detailed in Table 6., reveals distinct patterns. When examining trust expressions towards villagers (Column 1), the significant positive interaction effect suggests that although the trust level remains lower than the trust expressed by villagers, winning spies express more trust towards villagers compared to when they lose. Meanwhile, the interaction term for trust towards spies (Column 2) is not significant. This could potentially be explained by expressing trust to villagers may confer a tactical advantage for spies, possibly to blend in and reduce suspicion, thereby increasing their chances of winning the game.

	Trust Villager	Trust Spy
	(1)	(2)
Spy	-0.257***	-0.236***
	(0.079)	(0.084)

SpyWin	-0.276**	0.228**
	(0.132)	(0.105)
Spy:SpyWin	**0.410*** **	-0.074
	-0.121	-0.129
Male	0.275***	0.137**
	(0.064)	(0.067)
NativeEngSpeaker	0.261***	0.206***
	(0.075)	(0.073)
GameExperience	0.075	0.085
	-0.071	-0.071
Group Effect	0.479***	0.159**
	(0.149)	(0.075)
Intercept	0.580***	0.321***
	(0.108)	(0.094)
Observations	324	324
Residual Std. Error	0.495 (df=317)	0.537 (df=317)
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 6. Regression Results of Node Out-Degrees

3.4.5 Verbally Expressed Distrust

Hypothesis H4.1, which proposed deceivers are less likely to verbally express distrust toward other group members, is not supported (column 4 in Table 5.). There is no difference. The results show a similar pattern with perceived distrust (column 2), with only the game outcome having a significant effect. This phenomenon may be explained by the increased prevalence of verbal relational messages that convey distrust within the group during games where spies win. This atmosphere of heightened distrust creates "muddied waters," providing deceivers with a strategic advantage. By taking advantage of the existing confusion and skepticism, deceivers can manipulate the group dynamics more effectively, obscuring their actions and intentions to enhance their likelihood of success in the game.

On the other hand, Hypothesis H3.2 is supported. The data indicates that deceivers are less likely to verbally express distrust towards their fellow deceivers compared to truth-tellers. This is

evidenced by the spy's intra-community distrust connection score (mean 0.94) being lower than their inter-community distrust connection score (mean 2.13), as shown in Figure 8. The significant results from the pairwise t-test ($p<0.01$) indicate that deceivers employ collective solidarity to protect fellow deceivers, preferring this approach over strategic isolation when expressing distrust. This aligns with a submissive strategy, where deceivers opt to maintain a more reserved and less confrontational facade while simultaneously working to maximize the group's interests.

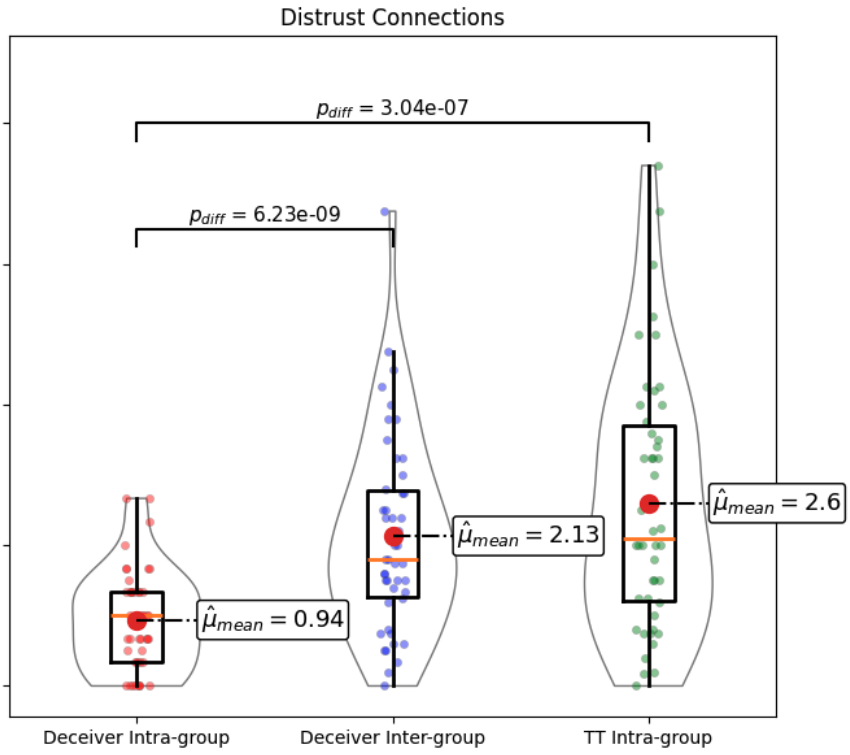


Figure 8. Comparison of Community Distrust Connection Score

3.4.6 Predictive Model

The table below summarizes the performance of each model under the 5-fold cross-validation scheme:

Model	Data Source	Accuracy	Precision	Recall	F1 Score
Logistic Regression	Feature set A	0.599	0.47	0.126	0.197
	Feature set B + Manual label	0.62	0.547	0.348	0.414

	Feature set B + GPT label	0.623	0.529	0.385	0.445
Gradient Boosting Machine	Feature set A	0.611	0.55	0.047	0.085
	Feature set B + Manual label	0.639	0.586	0.3	0.393
	Feature set B + GPT label	0.636	0.569	0.291	0.383
Neural Network	Feature set A	0.614	0.532	0.118	0.191
	Feature set B + Manual label	0.636	0.549	0.395	0.453
	Feature set B + GPT label	0.617	0.515	0.417	0.46
Graph Neural Network	Feature set B + Manual label	0.**653**	0.643	**0.591**	**0.615**
	Feature set B + GPT label	0.625	**0.671**	0.509	0.525
Feature set A: Player and game features only; Feature set B: set A + Network features					

Table 7. Performances of Predictive Models

Incorporating network features (Feature Set B) consistently improved model performance across all metrics compared to using player features alone (Feature Set A). This enhancement highlights the significant advantages of integrating network features into predictive models. Particularly, the Graph Neural Network (GNN) demonstrated superior performance, achieving an accuracy of 0.653 and an F1 score of 0.615 with manual labeling. This model adeptly utilized the network metrics to unravel the complex dynamics within the group, enhancing its capability to provide a deeper understanding of deceptive interactions.

Additionally, an interesting finding from the study was the comparable performance levels of models using manual labels versus those using GPT-generated labels. This similarity indicates that GPT-based labeling can effectively match the quality of manual coding when used as input for predictive models, presenting a scalable and cost-effective alternative for generating training data. The equivalence in performance between manual and GPT labeling underscores the potential of advanced language models to support robust predictive modeling in deception detection and other areas requiring nuanced, context-level data interpretation. This opens the possibility for

developing an end-to-end product that translates audio conversations directly into predictive results.

3.5 Discussion

3.5.1 Major findings and alternative explanations

The study explores the verbal interactive behaviors exhibited by deceivers during collective deception scenarios. Deceivers are typically less likely to express trust toward others, often adopting a "flight" strategy to blend in and avoid detection. They prefer strategic isolation when sending trust relational messages, tending to express trust towards truth-tellers rather than fellow deceivers to minimize collective suspicion. However, in expressing distrust, deceivers demonstrate a pattern of collective solidarity, especially directing less distrust toward colluders than toward truth-tellers. This switch between "Secrecy" and "Solidarity" strategies in dealing with fellow colluders highlights a submissive strategy that supervises both, suggesting deceivers will employ whichever strategy helps them to remain undetected and fly under the radar. In terms of trust, showing less trust helps them distance themselves from suspicious fellow deceivers while expressing distrust risks exposure and requires sophisticated deception skills and is strategically minimized to avoid drawing attention.

The relationship between deception outcomes and trust-distrust interaction is also significant. When deceivers successfully manipulate outcomes, they tend to express more trust toward truth-tellers compared to when they fail, possibly as a tactic to blend in and appear innocent. Additionally, successful deception is correlated with an increase in distrust expressed within group conversations, which could foster group division and provide deceivers an advantage to achieve their agendas.

Despite these findings, Hypotheses H2a, H2b, and H4.1 are not supported by the analysis, which reveals no significant difference in expressed or received distrust between deceivers and truth-tellers. This result underscores a clear distinction between trust and distrust behaviors. Although distrust, like trust, involves intense emotions, it may not correlate directly with the affective components of trust. Distrust is often associated with emotions such as wariness, caution, cynicism, defensiveness, anger, fear, and betrayal, while trust is linked with feelings of hope, safety, assurance, and confidence (Lewicki et al, 1998). The emotional responses tied to distrust and trust may be more orthogonal than merely at different ends of the same scale from each other. (McKnight and Chervany, 2001). Trust and distrust, therefore, are distinct mechanisms within interpersonal relationships; the presence of one does not necessitate the absence of the other, and expressing distrust is not merely the opposite of expressing trust.

3.5.2 Research Implications

The findings of this study significantly broaden our understanding of the impact of deceptive intent on verbal behavior, marking notable advancements in deception research. First, to the best of our knowledge, this study is pioneering in its investigation of collusion interactions within the realm of deception. Moving beyond commonly studied dyadic interactions, this research explores group settings where multiple deceivers collaborate. By examining the trust and distrust behaviors of deceivers, the study illuminates the strategies they may employ using relational messages to conceal their identities and intentions. Second, this research offers a novel perspective by considering the social structure of deceptive behaviors, providing empirical support for a social network analysis approach to deception in group communication. Deceptive interactions within a group are not merely the sum of individual tactics; they are shaped by the social hierarchy, roles, and relationships within the group. The data gathered provide a detailed view of how social

structures, influenced by the nature of deception, highlight patterns of trust and distrust that change depending on one's position within the group.

3.5.3 Practical Implications

This study's practical implications are particularly relevant for the enhancement of deception detection methodologies. First, they demonstrate that incorporating network metrics as predictive features can significantly enhance the performance of deception detection models. Utilizing Graph Neural Networks (GNNs) leverages the full power of network modeling, which has shown a substantial improvement in performance. Such an approach marks a departure from traditional methods, offering a data-driven, network-based analytical tool that better captures the complex dynamics of deceptive behavior within groups.

Secondly, the study has explored the integration of advanced language models, such as GPT, for tagging transcripts with context-level information. This technology facilitates the rapid analysis of large datasets, offering insights that were previously unachievable due to resource constraints. Automated tools for measurement have proven to be more scalable and cost-effective than manual coding, illustrating the potential for developing an end-to-end product that seamlessly transitions from conversation analysis to prediction. This approach could transform into various application scenarios, such as in security and law enforcement for interrogations, in corporate settings for compliance monitoring, or in online platforms for real-time detection of fraud. For example, in military operations, this technology can be pivotal in analyzing communications to identify deceit or misinformation among local interactions or within enemy communications, thereby improving strategic decision-making and operational security.

These enhancements in deception detection technology could significantly utilize the nuances of conversation, interactions, and the meanings behind words, thereby altering and making the processes of discernment more efficient and reliable.

3.5.4 Limitations and Future Directions

As with some empirical deception studies, the data were collected from a specific environment. Testing the generality of the current findings across different communication contexts is necessary. First, the lack of pre-existing relationships among participants in this experimental setup might have led to the observed tendency toward submissive strategies and isolation among deceivers. Conversely, deceivers with a history of collaboration might exhibit more dominant strategies and a higher degree of collective effort. Future research could investigate how established relationships impact deception strategies. Second, the low-stakes nature of the experimental context presents different motivational factors from high-stakes scenarios, potentially affecting the types and intensities of behavioral signals related to deception (Thompson and Hartwig, 2023). Research in high-stakes environments would offer insights into how strong motivation influences relational messages interact in the group. Third, our study was concentrated on small groups (6-8 members), characterized by dense conversation and network connections. Larger groups may demonstrate looser connections, thereby altering group dynamics and deceptive behaviors. Future studies should explore varying group sizes to determine how network density affects collusive deception. For example, criminal networks, where individuals often know each other, stakes are high, and groups are larger, tend to be partitioned into subgroups with close internal interactions (Hutchins and Benham-Hutchins, 2010). Subsequent studies should consider variations in familiarity, motivation, and group size to understand the impact of these factors on deceptive behavior and detection within more complex social structures.

Additionally, there are several methodological constraints within this study that future research should address to refine the analysis of deceptive behaviors. One limitation is that our system only considered the presence and valence of an utterance label, without gauging its

intensity or strength. Consequently, the social structural behaviors in our research model were operationalized on unweighted networks, neglecting the potential significance of the intensity of relational messages between players. Furthermore, the network metrics in this study were confined to degree and community connection scores. Future research should incorporate a broader range of network metrics such as betweenness, closeness, and density, which could offer a more comprehensive view of the network's structure and the relational messages' roles within it. Additionally, the incorporation of more dynamic network analysis techniques could provide a deeper understanding of how deceptive strategies evolve over time. The current study's static approach might not fully capture the temporal variations in communication patterns. Moreover, networks' sparseness at different time points could affect the performance of detection models; thus, incorporating temporal edge features into the GNN model is likely to enhance predictive accuracy. Lastly, this study employed default parameters for the classification algorithms, potentially limiting the performance of the detection models. Fine-tuning of these parameters may improve the performance of detection.

3.6 Conclusion

This study attempts to understand the impact of collusive deception on the verbal behaviors of deceivers within groups, exploring how such behaviors shape interactions with other group members. We conducted an experiment in which colluders were embedded among truth-tellers, acting covertly against the group's interests. Participants' conversations were recorded, and trust-distrust relational messages were annotated with a reliable utterance coding system. Through regression and t-test analysis, we found that deceivers frequently adopt "flight" strategies to evade detection, expressing trust less often and typically isolating themselves from other deceivers. Notably, deceivers also showed collective solidarity in their distrust, especially by expressing less

distrust towards fellow colluders. This interplay between "Secrecy" and "Solidarity" indicates that deceivers will employ whichever submissive strategy best allows them to remain undetected. Furthermore, successful deceivers tend to express more trust towards truth-tellers, possibly as a bonding tactic to blend in, while an increase in expressed distrust within the group may help them further their deceptive goals by obscuring their actions and intentions. This study significantly advances deception research by exploring collusive deception in group settings and underscoring the strategic use of verbal relational messages by deceivers. The incorporation of social network analysis and Graph Neural Networks has enhanced deception detection capabilities, producing more accurate models that fully utilize the complexities of group communication dynamics. Additionally, the integration of GPT for automated context-level tagging points to a future where conversation analysis is scalable, cost-effective, and potentially transformative for various sectors that depend on accurately discerning truthfulness.

Chapter 4: Nonverbal Behaviors, Collusive Deception, and Interpersonal Relationships in Group Communications

4.1 Introduction

This chapter delves into the impact of collusive deception on the nonverbal behaviors of deceivers within interactions with both fellow deceivers and truth-tellers, specifically focusing on the nonstrategic shifts in relational messages of dominance and nervousness. Conducted via Zoom, the study involved participants formed into 5-person teams tasked with evaluating and hiring candidates from provided resumes. Within each team, two members were incentivized to collusively advocate for low-quality candidates. Automated tools were used to extract and analyze facial and head features, revealing that deceivers exhibit reduced dominance and appear less emotionally nervous, yet display heightened cognitive stress cues during interactions with both fellow deceivers.

4.2 Literature Review and Theoretical Backgrounds

4.2.1 Nonverbal Relational Messages

"Nonverbal signals color the meanings of interpersonal relationships" (Burgoon et al, 2021). Relational communication is a fundamental aspect of interpersonal communication, with nonverbal signals playing a central role in enhancing our understanding of these dynamics. Nonverbal relational messages express meanings and sentiments that people may refrain from saying out loud, such as unspoken approval, concealed displeasure, or underlying anxiety in various social situations. For example, during a tense business negotiation, a negotiator's minimal eye contact or tightly crossed arms could subtly signal disagreement or discomfort with the proposed terms. Similarly, a comforting hand on the shoulder during a difficult conversation can

express support and empathy without words. These actions demonstrate the power of nonverbal cues to communicate complex emotions and relational messages subtly and effectively.

Extensive research has been conducted on how nonverbal behavior interplay with relational messages. For instance, the lip corner puller, which facilitates smiles, is commonly associated with expressions of liking, composure, involvement, and trust (Burgoon et al., 2022). Additionally, Hall, Coats, and LeBeau (2005) conducted a meta-analysis in which they examined 27 different nonverbal cues of dominance across 120 studies. In a related study, Dr. Burgoon and her team (2021) explored cross-cultural group deception by measuring a broad array of nonverbal signals from the head, face, torso, and voice. These signals are crucial in predicting participants' perceptions of the relational messages about dominance, nervousness, and trust that they receive from their peers. Their findings revealed that perceptions of dominance were associated with louder voices, more expressive facial behaviors, increased head movement, and more frequent and longer turns at talk. Conversely, signals of nervousness were linked to softer vocal amplitudes, reduced head movement, and fewer and shorter turns at talk. These cues are instrumental in predicting the relational messages conveyed in interactions, highlighting the profound impact of nonverbal cues in understanding and navigating social dynamics.

4.2.2 Deception and Nonverbal Relational Messages

Through the Brunswikian lens model, relational messages can also be used to understand deception and aid in deception detection. This model provides a systematic approach to connect objective behaviors exhibited by social actors to the emotions and cognitions perceived by other interactants, and further links these perceptions to social outcomes. Specifically, Burgoon et al. (2021) provide evidence that objective distal indicators—such as linguistic, vocal, and facial cues—that can be measured and analyzed with automated computer tools, can be correlated with psychological perceptual judgments that represent relational message themes. These themes

include dominance, affection, composure, involvement, similarity, and trustworthiness. These subjective percepts, in turn, can predict communicative outcomes, such as the successful identification of another's deception.

While the relationship between nonverbal cues, relational messages, and deception has been thoroughly examined, the impact of collusive deception on interpersonal relationships, particularly between deceivers (DD) and between deceivers and truth-tellers (DT), remains less explored. Collusive deception involves multiple parties agreeing to deceive, which adds a layer of complexity to their interpersonal dynamics. Understanding how these dynamics shift during the act of collusion can provide deeper insights into the psychological and emotional effects of collusion on each participant.

Deceivers engage in both strategic and nonstrategic activities: strategic efforts are aimed at managing information, behaviors, and overall image, while nonstrategic activities often inadvertently reveal signs of emotional stress, cognitive load, memory issues, speech disruptions, or failed attempts to suppress information leakage (Burgoon et al, 2022). Nonverbal and verbal messages reflect the socioemotional expressions that constitute relational messages between people (Burgoon et al, 2022). Nonverbal cues reflecting relational messages are typically nonstrategic, as they naturally emanate from an individual's genuine emotional state and cognitive processes, offering authentic insights into their feelings and thoughts during interactions. The authenticity of the interpersonal relationships among colluders is characterized by cooperative solidarity, as they work closely together to pursue a shared deceptive objective, ensuring consistency and reinforcing the group’s false narrative. Consequently, it is anticipated that colluders would exhibit nonverbal cues that unintentionally signal relational messages of collaboration and affirmation, further solidifying their covert alliance.

In this study, we concentrate on the relational message constructs of dominance and nervousness, which are crucial for understanding how deception operates during group deliberations. The significance of dominance, trust, and composure in these interactions is emphasized by Burgoon et al. (2021), who highlight their impact on the dynamics of relational communication. According to their findings, dominance is linked to the broadest array and the highest number of nonverbal cues, while nervousness is also associated with a considerable array of related cues. In contrast, the theme of trust exhibits the fewest nonverbal signals. Given the robust association of dominance and nervousness with observable nonverbal cues, this study will exclusively focus on these two relational messages to deepen our understanding of their roles in interpersonal relationships within the context of collusive deception.

4.2.3 Dominance and Collusive Deception

Dominance is the dynamic process of exerting power through relational and behavioral acts (Dunbar, 2004). Dunbar et al. (2015) noted that deception can be a source of power because deceivers have access to concealed information that truth-tellers lack. This informational asymmetry naturally empowers deceivers, often leading them to be more dominant when interacting with truth-tellers. Dominance, in this context, is often manifested through confident body language, assertive speech patterns, and control over the flow of interaction (Hall et al., 2005).

Conversely, when deceivers interact with one another, they typically possess equivalent levels of information and power. This parity means that neither party has a definitive advantage, leading to interactions that generally feature reduced levels of dominance. In these situations, the equal balance of power creates an environment where each deceiver acknowledges the other's abilities, reducing the need for assertive behavior. As a result, their interactions are more balanced and cooperative, with fewer displays of dominance.

Additionally, during collusive deception, when deceivers collaborate towards a common deceptive goal, the dynamics of their relationship shift towards cooperation. This reduces the need for dominance since assertive or controlling behavior is not necessary. The collaborative nature of collusion requires a coordinated effort where harmony and mutual support are more beneficial than hierarchical displays of power. The shared objective of deceiving others encourages behaviors that promote solidarity and reduce internal competition among the deceivers.

Therefore, we hypothesize that:

H1: Deceivers exhibit less dominance in their nonverbal cues when interacting with another deceiver compared to interactions with a truth-teller.

4.2.4 Nervousness and Collusive Deception

The fear of being caught is a significant contributor to nervousness during deceptive interactions. Zuckerman et al. (1981) highlight that the mental processes involved in deception, such as arousal, affect, cognitive load, and attempted control, are crucial drivers of these stress indicators. When individuals deceive, they experience heightened arousal and emotional strain, manifesting in nonverbal cues like increased nervousness and anxiety. This heightened state of alertness and the cognitive load required to maintain the deception can lead to noticeable stress signals. However, when deceivers interact with their colluders, they face a significantly reduced threat of judgment or detection compared to their interactions with truth-tellers. In a deceptive collaboration, the risk of being caught is lower because all parties involved share the same deceptive intent and are unlikely to expose each other. This reduction in fear within a collaborative deceptive environment naturally leads to a decrease in the expression of stress-related nonverbal cues. Deceivers can communicate more freely and confidently, knowing that their colluders are complicit in the deception.

Furthermore, the Communication Accommodation Theory (Giles, 2016) suggests that people adjust their communication styles based on social dynamics such as power and familiarity. This theory posits that individuals alter their verbal and nonverbal behaviors to accommodate or converge with those of their communication partners. In the context of deception, this means that deceivers are likely to modify their behavior depending on whether they are interacting with truth-tellers or fellow deceivers. Recognizing their colluders as ingroup members who share similar deceptive intents fosters a more relaxed communication style. Ingroup members share common goals and mutual understanding, reducing the pressure on deceivers to constantly monitor their behavior for signs of detection. This shared understanding and alignment in deceptive goals minimize stress markers like nervousness, as the need for vigilant self-monitoring and control diminishes among those with aligned objectives.

Therefore, we hypothesize that:

H2: Deceivers exhibit less nervousness in their nonverbal cues when interacting with another deceiver compared to interactions with a truth-teller.

4.2.5 Head and Face Features

In this study, we specifically focus on nonverbal cues from the face and head because these areas are recognized as primary focal points of attention in human interactions. Paul Ekman (2003) highlighted that facial expressions are universal and key to understanding human emotions, which naturally draw significant attention during interpersonal interactions. By concentrating on these areas, our aim is to capture subtle changes in facial expressions and head movements that may indicate differences in dominance and nervousness between interactions involving deceiver-to-deceiver (DD) and deceiver-to-truth-teller (DT) dynamics.

Many facial and head nonverbal behaviors are associated with perceived dominance. Previous studies have consistently shown that certain postures and facial expressions are strong

indicators of dominance. For instance, more expansive and upright postures, along with head positions that convey authority, are typically linked to higher perceptions of dominance (Hall et al., 2005; Burgoon et al., 2021). Additionally, while less smiling is observed, more expressive facial expressions that demonstrate control and confidence are also associated with dominance (Keating et al., 1981). The opposite behaviors are indicative of less dominance and more submission.

Regarding nervousness, while rigidity is not necessarily equated directly with nervousness, it is often correlated with it through various facial and head cues. Research indicates that rigid and tense behaviors, such as reduced head and face movements and a general decrease in pleasantness, are commonly exhibited during nervous states. This observation is supported by findings from Burgoon and Le Poire (1999), Gregersen (2005), and Pentland et al. (2017), who have all noted that such physical manifestations often accompany feelings of nervousness or tension.

A fuller delineation of signals indicating less dominance and less nervousness is presented in Table 8., which outlines the hypothesized relationships between deceiver-to-deceiver (DD) interactions and nonverbal behaviors. This table systematically categorizes and illustrates how specific face and head cues are expected to manifest differently when deceivers interact with each other compared to when they engage with truth-tellers.

Relational Message Constructs	**Hypothesized Nonverbal Signals**	
Less dominance (H1)	Head	Drooping head
		Reduced head movement
	Facial Emotion	More smiling face
		Fewer facial expressions
Less nervousness (H2)	Head	More nodding
		Increased head movement
	Facial Emotion	More facial expression
		More pleasant facial expressions/ fewer unpleasant facial expressions

Table 8. Summary of the Hypothesized Nonverbal Signals

4.3 Methods

4.3.1 The Mock Hiring Committee Experiment

To investigate the nonverbal relational messages in collusive deception, the study replicated the methodology of Dunbar et al. (2014) using an online platform (Zoom) for synchronous video communication. This setup simulated a hiring committee scenario where participants, grouped into teams of four to five, were tasked with evaluating and hiring the best candidate from a set of provided resumes. The experiment involved multiple phases, including review of resumes and interview notes, interview notes presentations, and group discussions.

Participants were each instructed to review interview notes documenting one candidate's performance and prepare a summary for the other committee members. Two participants in each group were randomly assigned the role of deceivers, while the remaining participants served as truth-tellers. The truth-tellers received interview notes for candidates with either high or medium-quality resumes. In contrast, the deceivers received notes corresponding to low-quality resumes. Deceivers were informed that their objective was to act as corporate spies from a competitor company, aiming to manipulate the group into hiring an unqualified candidate. Success for the deceivers meant getting one of the two least suitable candidates hired. They were permitted to fabricate details during their presentations and group discussions. Truth-tellers, unaware of the deceivers' roles, were instructed to aim for the most qualified candidate's hiring to win the game. They were not informed that some participants would advocate for unqualified candidates.

Each session was facilitated by a trained research assistant who used videotaped instructions to introduce the experiment, ensuring consistency across sessions. The entire experiment was video recorded, and transcripts of the sessions were saved for detailed analysis.

The experiment's flow is shown in Figure 9., with further detailed experiment procedures available in Appendix G.

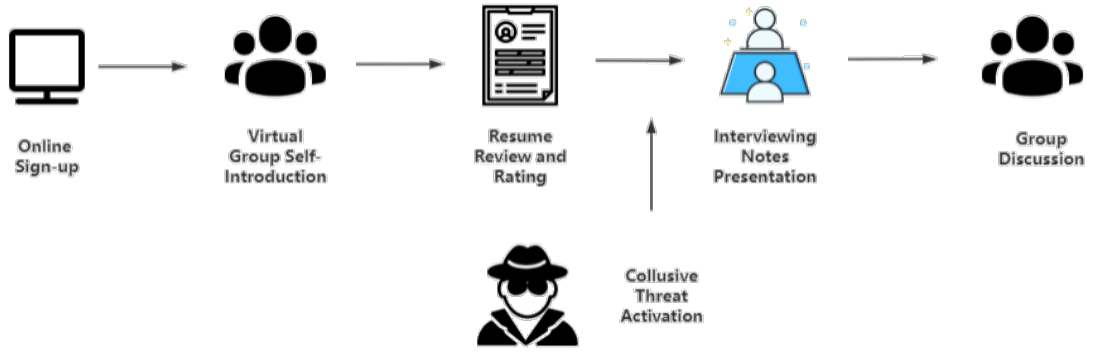


Figure 9.The Mock Hiring Committee Experiment Procedures

This experimental design is particularly effective in studying the subtle nuances of nonverbal relational messages within the context of collusive deception. First, the setup creates an optimal environment for exploring collusion, with multiple deceivers strategically embedded within the group. These participants engage in covert deception to achieve shared gains, operating secretly to subvert the goals of the larger group. This configuration aligns with our research objectives to examine the dynamics of collusion. Second, the experiment is structured to promote extensive interaction among participants, which is essential for analyzing relational messages. The process includes both structured interview note presentations and unstructured discussions, providing diverse interaction formats that elicit a range of verbal and nonverbal behaviors. Third, the use of video communication platforms like Zoom simplifies the setup compared to in-person laboratory settings, while still maintaining control over the experimental environment. This remote setup not only makes the experiment more accessible but also provides a rich array of facial and head cues for detailed analysis. These aspects together enhance our ability to observe and interpret the subtle nuances of nonverbal communication within the context of group collusion.

4.3.2 Data Preprocessing and Feature Extraction

Before the experiment commenced, an audiovisual check was conducted to ensure that each participant's video camera and microphone were fully functional. Videos were trimmed according to timestamps generated from Zoom's auto-created transcripts. These timestamps were manually adjusted to ensure perfect alignment with the conversation, focusing particularly on segments during the interview note presentation and group discussion. These segments are critical as they occurred after the introduction of collusive intent, where deceivers were instructed to manipulate the group into hiring an unqualified candidate.

The trimmed video clips were processed using the Face++ API software (Megvii, 2022), which estimates head pose, smile, and facial emotions. The software measures the pitch, roll, and yaw angles, representing the rotation angles of a detected head in coordinates with the camera serving as the origin, with a range from -180 to 180 degrees. An increase in the pitch angle indicates a head lift. An increase in the roll angle is observed with a clockwise head rotation. An increase in the yaw angle occurs when the head turns from left to right. The Face++ API was also utilized to measure the intensity of six basic emotions (anger, disgust, fear, happiness, sadness, and surprise) and smiles detected on a subject’s face. These variables are scaled from 0 to 100, where a higher value indicates greater confidence in the presence of the specific emotion.

For analytical purposes, data were aggregated by calculating the pooled mean and standard deviation. These metrics refer to the average intensity and the variability of the extracted features from the corresponding video clips, respectively. In summary, a total of 22 features describing head pose and facial emotions were extracted from the video clips, as detailed in Table 9.

Modality	**Feature Name**	**Aggregation**	**Number of Features**	**Features**

Head pose	Head pose intensity	Mean	3	Mean of pitch, roll, and yaw angles
	Head pose variability	Sd	3	Sd of the same head poses as above
Emotions	Emotional intensity	Mean	7	Mean of happiness, sadness, surprise, fear, anger, disgust, and neutral
	Emotional variability	Sd	7	Sd of the same emotions as above
Smile	Smile intensity	Mean	1	Mean of smile
	Smile variability	Sd	1	Sd of smile
Note: Sd stands for standard deviation				

Table 9. 22 Features Extracted from Video Clips

4.3.3 T-tests

To study the different nonverbal relational messages that deceivers convey towards fellow deceivers and truth-tellers, we operationalize the interactions as follows: Deceiver-to-Deceiver (DD) interactions are defined as periods when a deceiver listens to another deceiver speaking, and Deceiver-to-Truth-Teller (DT) interactions as times when a deceiver listens to a truth-teller speaking. This focus on listening phases is chosen for three key reasons. First, in group settings, individuals often "talk to the room," directing comments generally rather than to a specific person (Dunbar et al., 2021). This behavior makes it challenging to determine the exact target of spoken comments, thereby complicating the analysis of directed nonverbal cues. Second, while speaking, facial features change significantly due to movements involved in speech, such as mouth movements. These changes can engage various facial muscles, affecting the clarity and interpretation of facial cues. By focusing on listening moments, we minimize the influence of speech-related facial movements, allowing for a clearer analysis of other expressive features. Third, when individuals speak, they often engage in strategic behaviors, consciously managing how they

present information, their behavior, and their personal image to the audience. Interpersonal Deception Theory (IDT) emphasizes that deceivers intensify these strategic activities to effectively manipulate others' perceptions (Buller and Burgoon, 1996). In contrast, listeners' nonverbal cues are generally less controlled and more spontaneous, offering a more genuine reflection of their relational stance and emotional state. Since the aim of this study is to analyze nonstrategic activities that convey relational messages, focusing on the nonverbal cues of deceivers while they are listening, rather than speaking, is particularly suitable.

To analyze the differences in nonverbal behaviors between Deceiver-to-Deceiver (DD) interactions and Deceiver-to-Truth-Teller (DT) interactions, we calculate the difference in nonverbal features observed in these two scenarios for each deceiver. A t-test is then used to determine whether these differences are statistically significant. This method is chosen over a plain ANOVA because it accounts for individual variations in nonverbal behaviors, which is important when comparing the same individuals' reactions in different situational contexts.

4.4 Results

4.4.1 Summary Statistics

Participants for the study were recruited from two large public universities in the Western US. A total of 22 group experiment sessions were initially formed, but one session with four participants was subsequently removed. The remaining 21 sessions involved a total of 105 participants, comprising 69 females, 34 males, and two who did not report their gender. Deceiver won 10 games. The demographic breakdown of the participants was as follows: 54.3% were white; 19% were Asian; 9.5% were Hispanic/Latino; and 3.8% were black. The average age of the

participants was 21.4 years (SD = 2.1), with ages ranging from 18 to 31 years. Additionally, 80% of the participants were native English speakers.

4.4.2 Dominance

The data provide partial support for Hypothesis 1, which suggests that deceivers exhibit less dominance in their nonverbal cues when interacting with fellow deceivers as opposed to interactions with truth-tellers. Although there is no significant difference in the frequency of smiling between these interactions, other indicators suggest a reduction in dominance. Specifically, deceivers show fewer upright head positions, fewer head movements, and fewer facial expressions when interacting with other deceivers. These findings align with the expectation that the relational dynamics among deceivers involve less dominance than those between deceivers and truth-tellers.

	Nonverbal signals	**Measurement**	**T-statistics**	**P-value**	**Nonverbal Signals Interpretation**
Head	Upright head	Mean head pitch	**-2.15**	**0.04***	Drooping head
	Nodding	SD head pitch	**-2.71**	**0.01***	Less nodding
	Head movement				Reduced head movement
		SD head roll	**-2.09**	**0.04***	
		SD head pitch	**-2.17**	**0.04***	
Facial Emotion	Smiling	Mean smile	-1.44	0.16	Less smiling face (not significant)
	Emotional variability	SD happiness	**-1.88**	**0.07^{+}**	less facial expression/ more rigid facial expression
		SD sadness	**-3.34**	**0.00****	
		SD surprise	**-1.70**	**0.10^{+}**	
		SD fear	**-2.11**	**0.04***	
		SD anger	-0.30	0.77	
		SD disgust	**-1.71**	**0.10^{+}**	
		SD neutral	**-2.62**	**0.01***	
	Emotional intensity	Mean happiness	-1.04	0.31	More neutral, less sadness, surprise, fear facial emotions
		Mean sadness	**-1.74**	**0.09^{+}**	
		Mean surprise	**-1.84**	**0.07^{+}**	
		Mean fear	-2.11	**0.04***	
		Mean anger	0.70	0.49	

		Mean disgust	-1.05	0.30	
		Mean neutral	**1.75**	**0.09**$^{+}$	

*Note: +p < 0.1, *p < 0.05, **p < 0.01*

Table 10. T-test Results for Nonverbal Behaviors Related to Dominance and Nervousness

4.4.3 Nervousness

H2, which hypothesized that deceivers would exhibit less nervousness in their nonverbal cues when interacting with another deceiver compared to interactions with a truth-teller, is not supported by the full scope of the results. The findings present a complex picture.

Emotionally, the results indicate a decrease in negative facial expressions among deceivers when interacting with each other, evidenced by fewer displays of sadness, surprise, and fear, and more neutral expressions. These observations are in line with H2, suggesting that deceivers feel less threatened and therefore less nervous when among fellow deceivers, leading to fewer negative emotional expressions. Contrary to the hypothesis, from a cognitive perspective, the data demonstrate a rigid effect characterized by reduced head movements and less variability in facial expressions. This rigidity suggests a higher level of control or suppression, which is typically associated with an increased cognitive load rather than a relaxed state. Furthermore, the analysis revealed no significant difference in nodding behaviors between interactions with deceivers and truth-tellers.

In summary, while emotional cues suggest less nervousness as hypothesized in H2, the cognitive indicators contradict this, showing signs of increased cognitive load and potential stress.

Relational Message Constructs	**Hypothesized Nonverbal Signals**		**Results**
Less dominance (H1)	Head	Drooping head	Support
		Reduced head movement	Support
	Facial Emotion	More smiling face	Not Support
		Less facial expression	Support
	Head	More nodding	Not Support

Less nervousness (H2)		More head movement	Not Support
	Facial Emotion	More facial expression	Not Support
		More pleasant facial expressions/ less unpleasant facial expression	Partially Support

Table 11. Summary of the Nonverbal Cue Testing Results

4.5 Discussion

4.5.1. Major findings and alternative explanations

The study investigates the nonverbal behaviors of deceivers in collusive deception scenarios, revealing that deceivers display less dominance when interacting with other deceivers, demonstrated by less upright head positioning, fewer head movements, and reduced facial expressions. This adheres to the assumption that deceivers share equal information and power, and they cultivate cooperative solidarity in their nonstrategic interpersonal connections. Additionally, while emotional cues show decreased nervousness among deceivers, marked by fewer negative facial expressions, cognitive indicators suggest heightened nervousness, as evidenced by increased rigidity in head movements and facial expressions.

The mixed result of nervousness could be explained by different levels of appraisals on the situation. Cognitive appraisal theory, initially proposed by Lazarus (1991), offers a valuable framework for understanding how individuals evaluate and react to specific events or interactions. In the context of deceptive interactions, deceivers first assess whether the interaction poses a threat, challenge, or benefit. When interacting with fellow deceivers who share similar goals and are aware of the deception, the situation is likely perceived as less threatening compared to interactions with truth-tellers who may catch them lying. Therefore, this reduces the likelihood of experiencing strong negative emotions such as fear, sadness, or surprise that are typically triggered by perceived threats. Following the initial assessment, deceivers evaluate their ability to manage or control the situation. In group settings, each member of the group loses some control over what information

is provided and how each member behaves (Vernham et al., 2016). Deceivers must not only regulate their own information but also collaborate to manage the impressions of other deceivers. They need to listen attentively to other deceivers, considering ways to enhance the manipulation effect on the false narratives or mitigate exposure risks, thereby experiencing a significant cognitive load. Despite the non-threatening nature of the situation, as deceivers perceive that they have less control over the outcome of the interaction, this introduces a different type of nervousness. This can lead to what is observed as the "rigid effect," such as reduced head movement and a controlled facial expression.

Moreover, this study primarily focuses on head and face movements to assess nonverbal cues. Incorporating a broader range of nonverbal features could provide a more comprehensive understanding of nervousness. According to Burgoon et al. (2021), the vocal channel often emerges as a predominant signal of nervousness. This predominance is likely due to the higher controllability of kinesic elements and the relative difficulty in concealing emotional indicators in the voice. By expanding our analysis to include vocal elements such as tone, pitch, and speech rate, we might capture additional nuances of nervousness that are not as easily controlled or masked, thus providing a richer and more detailed profile of the nonverbal dynamics in deceptive interactions.

4.5.2. Implications

This study investigates the nonverbal dynamics within collusive deception scenarios, shedding light on the non-strategic activities between deceivers regarding their expressions of dominance and nervousness. This aspect of deception has not been widely studied before, and our findings contribute to understanding the complex interplay of emotions and cognitions that characterize deceptive behaviors. By identifying the subtleties in these interactions, our findings

broaden the theoretical frameworks on deception, revealing how relational messages are conveyed in a collaborative deceptive environment.

The use of a mock hiring experiment represents a novel approach to the study of group deception. Unlike many traditional deception studies that utilize competitive games such as Mafia or Resistance, which are structured around clear adversarial roles, the mock hiring experiment provides a more naturalistic and professional setting. This experiment more closely mimics real-world professional interactions, offering a context that is likely to provoke more genuine behaviors and responses. The setup not only provides a structured task but also allows for naturalistic dialogue and interaction, enabling the collection of richer data on the nonverbal cues used in professional settings. Additionally, the use of Zoom for this experiment facilitates the capture of high-quality audiovisual data that reflects contemporary communication practices.

4.5.3. Limitations and Future Directions

The current study is limited by its small sample size, with the group serving as the unit of analysis. This has potentially underpowered our statistical analysis, which may account for some of the non-significant findings. To address this issue, future studies should aim to collect more data to enhance the statistical power and robustness of the results, allowing for more definitive conclusions.

While this study focused primarily on facial and head movements, other nonverbal cues such as vocal tones and linguistic patterns were not examined. These elements play a critical role in conveying relational messages within interactions. Future research should incorporate a broader array of nonverbal cues to provide a more comprehensive understanding of how different modalities contribute to the communication of deception and relational messages.

The current research focused on the constructs of dominance and nervousness. However, other relational dimensions such as affection-hostility and involvement-detachment are also worth

exploring. Affection-hostility examines the warmth or hostility within interactions, shedding light on how attitude may shift from deceiver-to-deceiver (DD) to deceiver-to-truthteller (DT) interactions. Involvement-detachment assesses how engaged or detached deceivers are when interacting with fellow deceivers, offering insights into their engagement levels. Future studies could expand the range of relational messages examined to include these and potentially other dimensions, enriching the theoretical and practical understanding of relational communication in deceptive contexts.

4.6. Conclusion

This research attempts to investigate the impact of collusive deception on the nonverbal behaviors of deceivers, examining the shifts in relational messages of dominance and nervousness. We conducted a mock hiring experiment, where participants underwent a multi-phase screening process that included presenting the qualifications of a single candidate and then engaging in a group deliberation about the five candidates under consideration. In this setup, two members were incentivized to advocate for low-quality candidates, with rewards tied to candidate selection. Utilizing automated tools for face and head feature analysis, the results demonstrated that deceivers exhibit reduced dominance and appear less nervous emotionally yet display heightened cognitive stress cues during interactions with fellow deceivers.

Our findings deepen comprehension of the complex interplay between emotions and cognition in collusive deception, broadening theoretical frameworks on deception and relational communication. Through a novel approach using a mock hiring experiment, conducted via Zoom, we provide a naturalistic setting that mirrors professional interactions in a group setting, yielding richer data on nonverbal cues in real-world contexts and advancing research in the field. Moving

forward, research should aim to increase expanding sample sizes, explore diverse nonverbal cues, and investigate additional relational message constructs.

Appendix

Appendix A: Definitions of Terminology

Term	Identification
Relational Message	Nonverbal and verbal messages that signal how people regard one another and define their interpersonal relationships.
Interpersonal Group Communication	Interpersonal group communication refers to the process by which small groups of individuals exchange information, feelings, and meanings through verbal and non-verbal messages among themselves.
Collusive Deception	The collective action by a group of deceivers with hidden objectives aimed at undermining the interests of a larger group.
Trust	Positive beliefs in another's competence, benevolence, integrity, and the predictability of their actions, leading to a willingness to rely on them, which is reflected in cooperative behaviors.
Distrust	Skepticism about these qualities and results in reluctance to depend on others, often leading to defensive behaviors like information withholding and strict formal agreements.

Appendix B. The Original Utterance Coding Taxonomy Proposed in Chen, 2023

Label	**Description/Definition**	**Illustrative Examples**
Agreeing Statement (Agreeing)	Utterances that express agreement, acceptance, compliment, or willingness to cooperate	Yeah, I agreed to that That's also what I, I was gonna say that Yeah, true
Disagreeing Statement (Disagreeing)	Utterances that express disagreement or unwillingness to cooperate	No, I'm against No, you can't. You can't do that But that's not a guarantee as well
Justification (Justification)	Utterances that explain (especially the speaker's own) decision, choice, action, or behavior	Because I'm a very broke person (I) Just had a feeling
Reasoning (Reasoning)	Utterances that analyze the game (such as the role of a participant, the composition of a mission team, etc.) in a logical and evidence-based way	So, if it's a fail, then we know the deceivers inside the four, right? But if not fail, we know they are not deceiver.
Statement of Certainty (Certainty)	Utterances that express certainty about something (especially about other participants' game role)	I think I have a very good idea of who the deceivers are Well, now I have an idea who could be a deceiver on this team
Statement of Uncertainty (Uncertainty)	Utterances that express uncertainty, lack of understanding, lack of clarity, and confusion	I can't tell for sure I don't know. I'm so confused
Statement of Positive Sentiment (Positive)	Utterances that express optimism and confidence in general (not specific about another participant or mission team)	That's actually good! Well done! It's an easy game
Statement of Negative Sentiment (Negative)	Utterances that express a pessimistic, hopeless, or despairing mood in general (not specific about another participant or mission team)	It's a sellout silly game It's not adding up
Trusting Statement (Trusting)	Utterances that express trust towards another participant or a mission team	He seems trustworthy I don't think he's a deceiver I think you'd be a good team leader

Distrusting Statement (Distrusting)	Utterances that express distrust or suspicion towards another participant or a mission team	You just like to watch the world burn Julie, you're the odd girl out pretending
Accusation (Accusation)	Utterances that accuse another participant of being a deceiver or being guilty of a failed mission explicitly	Your both deceivers That makes it more sense that you are a deceiver
Game Role Self disclosure (Self-Disclosure)	Utterances in which the speaker discloses his/her truth-teller identity directly	I'm a truth-teller We've already cleared my name
Indirect Role Disclosure (Indirect Disclosure)	Utterances in which the speaker indicates or implies his/her truth-teller role implicitly, indirectly, or unconsciously	Okay guys keep winning Deceivers already won twice
Game Rule Statement (Game Rule)	Utterances that restate the game rule	We need two deceiver votes We can't have the same team
Game Rule Question (Rule Question)	Questioning utterances that inquire about the game rule	By the way, how many deceivers (do) we have? Does this take from our time?
Game Replay (Game Replay)	Utterances that bring up or recall a fact, detail, or situation in previous game rounds	They did it with (Participant Number) Five last time So, we have fail(ed), failed the previous mission
Suggestion (Suggestion)	Utterances that suggest a mission team or strategy (suggestion of team leader not included)	You have to be, you have to pick (Participant Number)1, 2, 5, 4 Well, next time nominate me, and we'll see
Question (Question)	Questioning utterance that inquire about other participants' choice, opinion, or rationale (not about game rules)	Okay, why do you think that? Why are we believing these two then?
Nomination (Nominate)	Utterances that intent to nominate a team leader (including self-nomination)	I nominate Mikaela I want you to be the leader
Promise (Promise)	Utterances in which the speaker assures that he/she will do a particular thing	I swear I will, I will pass the mission
Note: notations in parentheses under each label are the abbreviated labels that are used in results presentation. The exemplary utterances may not be grammatical.		

Appendix C: Node Ranking Regression Results

	Received Trust	Prestige	HITS	PageRank
	-1	-2	-3	-4
Spy	-0.310***	-0.278***	-0.205***	-0.125**
	(0.051)	(0.051)	(0.051)	(0.051)
SpyWin	-0.064	-0.098**	-0.031	-0.045
	(0.048)	(0.049)	(0.049)	(0.049)
Spy:SpyWin	0.257***	0.196**	0.131*	0.130*
	(0.077)	(0.078)	(0.077)	(0.078)
Male	0.014	-0.130***	-0.111***	-0.086**
	(0.039)	(0.039)	(0.039)	(0.039)
NativeEngSpeaker	0.071*	-0.048	-0.021	-0.023
	(0.038)	(0.039)	(0.038)	(0.039)
GameExperience	0.080**	0.046	0.062	0.045
	(0.038)	(0.038)	(0.038)	(0.038)
Intercept	0.476***	0.677***	0.606***	0.583***
	(0.047)	(0.047)	(0.047)	(0.047)
Observations	324	324	324	324

Note: *p<0.1; **p<0.05; ***p<0.01

Appendix D: Node Ranking Algorithms

1. Received Trust

$$Received\ Trust\ = \left|IN_i^{(+)}\right|$$

2. Prestige (Shahriari and Jalili, 2014):

Denote the number of positive incoming links to node *i* as $\left|IN_i^{(+)}\right|$ and those with negative sign as $\left|IN_i^{(-)}\right|$. The prestige of node *i* (Pr_i) is calculated as

$$Pr_i = \frac{\left|IN_i^{(+)}\right| - \left|IN_i^{(-)}\right|}{\left|IN_i^{(+)}\right| + \left|IN_i^{(-)}\right|}$$

3. HITS (Shahriari and Jalili, 2014):

Denote the subgraph with only positive links by $G^+(V,E^+)$ and the one with negative connections by $G^-(V,E^-)$ HITS algorithm runs separately on these subgraphs and the update equations are

$$h_i^{(+)} = \sum_{j \in IN_i^{(+)}} a_j^{(+)};\ a_j^{(+)} = \sum_{j \in OUT_i^{(+)}} h_j^{(+)}$$

$$h_i^{(-)} = \sum_{j \in IN_i^{(-)}} a_j^{(-)};\ a_j^{(-)} = \sum_{j \in OUT_i^{(-)}} h_j^{(-)}$$

where h_i indicates the hub vector and a_i the authority vector for the nodes. '+' sign above hub and authority vectors indicates the values for G^+ and "-" in G^-. The sum in the above update equations (for node *i*) is carried out to the authority and hub sets of node *i*. $OUT_i^{(+)}$ shows the set of nodes who receive positive links from node *i* and $OUT_i^{(-)}$ indicates those receiving negative links from node *i*. Similarly, $IN_i^{(+)}$ and $IN_i^{(-)}$indicate the set of nodes

pointing to node *i* with positive and negative links respectively. One can rewrite the above equations as

$$a^{(+)}(t+1) = A^{+T}A^{+}a^{(+)}(t); (t+1) = A^{-T}A^{-}a^{(-)}(t)$$

$$h_{\square}(t+1) = A^{+T}A^{+}h^{(+)}(t); h^{(-)}(t+1) = A^{-T}A^{-T}h^{(-)}(t)$$

where $a^{(+)}(t)$ and $h^{(+)}(t)$ show authority and hub vectors in G^+ at time t. A^+ and A^- are the adjacent matrices of G^+ and G^-.

4. Modified PageRank (Shahriari and Jalili, 2014):

Apply PageRank separately on G^+ (subgraph with positive links) and G^- (subgraph with negative links). The update rules are as

$$PR_i^+(t+1) = \alpha \sum_{j \in IN_i} \frac{PR_j^+(t)}{|OUT_j^{(+)}|} + (1-\alpha)\frac{1}{N}$$

$$PR_i^-(t+1) = \alpha \sum_{j \in IN_i} \frac{PR_j^-(t)}{|OUT_j^{(-)}|} + (1-\alpha)\frac{1}{N}$$

where $|OUT_i^{(+)}|$ and $|OUT_i^{(-)}|$ are the number of positive and negative outgoing links from node *j*, respectively. The algorithms start with some initial condition for PR^+ and PR^- vectors and after enough iterations it converges to the final rank vectors for G^+ and G^-. The final rank vector (*PR*) is calculated by

$$PR = PR^+ - PR^-.$$

Appendix E: GNN Model Structure

```
SignedGCNModel(
 (layer1): SignedGCNLayer(
    (conv_pos): GCNConv(12, 24)
    (conv_neg): GCNConv(12, 24)
  )
  (layer2): SignedGCNLayer(
    (conv_pos): GCNConv(24, 2)
    (conv_neg): GCNConv(24, 2)
  )
)
```

Appendix F: GPT Prompt

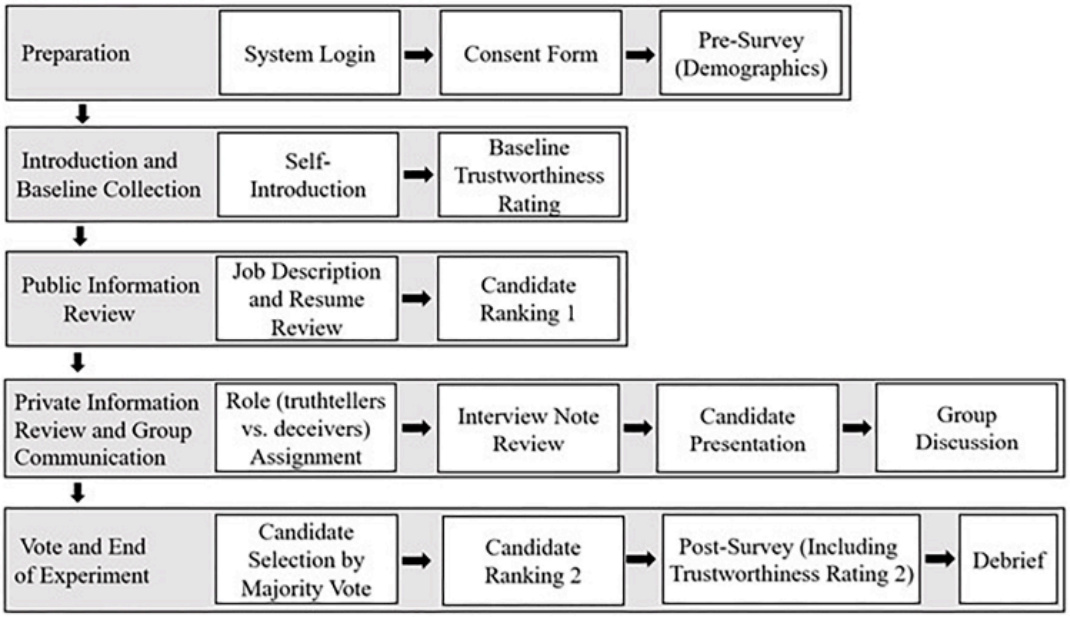
Goal:
You are an advanced AI tasked with analyzing player interactions during a round of the "Resistance" game. Your primary function is to evaluate comments made during discussions to ascertain the sentiments players harbor towards each other, assigning corresponding sentiment scores. These scores should range from -1 to 1, where -1 indicates the player is perceived as a deceptive spy, and 1 suggests the player is viewed as a trustworthy villager or leader candidate.
Rules:
- Players are identified by labels p1 through p8.
- The speaker and the mentioned player must not be the same individual. speaker != mentionedPlayer. no self-mention.
- Focus solely on tagging conversations that exhibit clear positive or negative sentiment. Accompany each identified sentiment with a brief reasoning and the index of the conversation from the input where the sentiment was observed.
- A facilitator is present within the game context but does not participate in the sentiment analysis. no facilitator in speaker or mentionedPlayer
- The output should be in JSON format, detailing the sentiment scores as specified, along with the index of the conversation from which each sentiment was derived.
Example Input:
{ "1. P2: 'I trust p4 completely, but I'm unsure about p5. p6 seems to be hiding something.'
2. P5: 'What player would you pick for the mission?'
...
10. P3: 'I will pick p2, p4 for the mission.'
11. P7: 'I agree with your choices.'
...
66. P8: 'I have a feeling p4 is a villager and I would like nominate her for this round's leader.'
"}

Expected Output:

```
{  "sentimentScores": [
   {     "speaker": "p2",
     "mentionedPlayer": "p4",
     "score": 1,
     "reason": "Expressed complete trust.",
     "conversationIndex": 1    },
   {     "speaker": "p2",
     "mentionedPlayer": "p6",
     "score": -1,
     "reason": "Suggested deception.",
     "conversationIndex": 1    },
   {     "speaker": "p3",
     "mentionedPlayer": "p2",
     "score": 1,
     "reason": "Selected for team.",
```

```
    "conversationIndex": 10    },
  {     "speaker": "p3",
    "mentionedPlayer": "p4",
    "score": 1,
    "reason": "Selected for team.",
    "conversationIndex": 10    },
  {     "speaker": "p7",
    "mentionedPlayer": "p3",
    "score": 1,
    "reason": "Supported p3's choice.",
    "conversationIndex": 11    }
  {     "speaker": "p8",
    "mentionedPlayer": "p4",
    "score": 1,
    "reason": "seems to be a villager and nominated for leader.",
    "conversationIndex": 66    }
 ]
}
```

Appendix G: Mock Hiring Experiment Procedures



(from Dunbar et al., 2023)

References

Bai, C., Bolonkin, M., Kumar, S., Leskovec, J., Burgoon, J. K., Dunbar, N. E., & Subrahmanian, V. S. (2019, August). Predicting dominance in multi-person videos. In *IJCAI* (pp. 4643-4650).

Bales, R. F., Strodtbeck, F. L., Mills, T. M., & Roseborough, M. E. (1951). Channels of communication in small groups. *American Sociological Review*, *16*(4), 461-468.

Bateson, G. (1951). "Information and codification," in *Communication: The social Matrix of Psychiatry*, eds J. Ruesch and G. Bateson (New York, NY: W.W. Norton), 168–211.

Bateson, G. (1958). *Naven: A survey of the problems suggested by a composite picture of the culture of a New Guinea tribe drawn from three points of view* (Vol. 21). Stanford University Press.

Boltz, M. G., Dyer, R. L., & Miller, A. R. (2010). Jo are you lying to me? Temporal cues for deception. *Journal of Language and Social Psychology,* 29(4), 458-466.

Bond Jr, C. F., & DePaulo, B. M. (2008). Individual differences in judging deception: accuracy and bias. *Psychological bulletin*, *134*(4), 477.

Bonjour, T., Aggarwal, V., & Bhargava, B. (2022, August). Information theoretic approach to detect collusion in multi-agent games. In *Uncertainty in Artificial Intelligence* (pp. 223-232). PMLR.

Boshoff, W. H., & Paha, J. (2021). List price collusion. *Journal of Industry, Competition and Trade*, *21*(3), 393-409.

Brown, G., & Yule, G. (1983). Discourse Analysis Cambridge University Press. *Cambridge, England.*

Buller, D. B., & Burgoon, J. K. (1996). Interpersonal deception theory. *Communication theory*, *6*(3), 203-242.

Buller, D. B., Burgoon, J. K., Buslig, A. L., & Roiger, J. F. (1994). Interpersonal deception VIII: Further analysis of nonverbal and verbal correlates of equivocation from the Bavelas et al.(1990) research. *Journal of Language and Social Psychology*, *13*(4), 396-417.

Buchan, N. R., Croson, R. T., & Solnick, S. (2008). Trust and gender: An examination of behavior and beliefs in the Investment Game. *Journal of Economic Behavior & Organization*, 68(3-4), 466-476.

Burgoon, J. K. (2021). Prelude: Relational communication and the link to deception. *Detecting trust and deception in group interaction*, 3-9.

Burgoon, J. K., & Bacue, A. E. (2003). Nonverbal communication skills. *Handbook of communication and social interaction skills*, 179-219.

Burgoon, J. K., Buller, D. B., & Floyd, K. (2001). Does participation affect deception success? A test of the interactivity principle. *Human Communication Research*, *27*(4), 503-534.

Burgoon, J. K., Buller, D. B., Guerrero, L. K., Afifi, W. A., & Feldman, C. M. (1996). Interpersonal deception: XII. Information management dimensions underlying deceptive and truthful messages. *Communications Monographs*, *63*(1), 50-69.

Burgoon, J. K., & Hale, J. L. (1984). The fundamental topoi of relational communication. *Communication Monographs*, *51*(3), 193-214.

Burgoon, J. K., & Hale, J. L. (1987). Validation and measurement of the fundamental themes of relational communication. *Communications Monographs*, *54*(1), 19-41.

Burgoon, J. K., Manusov, V., & Guerrero, L. K. (2021). *Nonverbal communication.* Routledge.

Burgoon, J. K., & Le Poire, B. A. (1999). Nonverbal cues and interpersonal judgments: Participant and observer perceptions of intimacy, dominance, composure, and formality. *Communications Monographs*, 66(2), 105-124.

Burgoon, J. K., Wang, R. X., Chen, X., Ge, T. S., & Dorn, B. (2022). How the Brunswikian Lens Model Illustrates the Relationship Between Physiological and Behavioral Signals and Psychological Emotional and Cognitive States. *Frontiers in Psychology*, *12*, 781487.

Burgoon, J. K., Wang, X., Chen, X., Pentland, S. J., & Dunbar, N. E. (2021). Nonverbal behaviors "speak" relational messages of dominance, trust, and composure. *Frontiers in Psychology*, *12*, 624177.

Burgoon, J. K., & Qin, T. (2006). The dynamic nature of deceptive verbal communication. *Journal of Language and Social Psychology*, *25*(1), 76-96. https://doi.org/10.1177/0261927x05284482.

Cartwright, D., & Zander, A. F. (1978). *Group Dynamics: Research and Theory*. Harper and Row.

Chen, X. (2023). *Intelligent Systems for Understanding Trust and Deception in Human Behavior* (Doctoral dissertation, The University of Arizona).

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological bulletin*, *129*(1), 74.

DeSanctis, G., & Gallupe, R. B. (1987). A foundation for the study of group decision support systems. *Management science*, *33*(5), 589-609. https://doi.org/10.1287/mnsc.33.5.589.

Dorris, B. (2018). Report to the nations, 2018 Global study on occupational fraud and abuse. *New York, Association of Certified Fraud Examiners.*

Dorn, B., Dunbar, N. E., Burgoon, J. K., Nunamaker, J. F., Giles, M., Walls, B., ... & Subrahmanian, V. S. (2021). A system for multi-person, multi-modal data collection in behavioral information systems. *Detecting trust and deception in group interaction*, 57-73.

Dunbar, N. E. (2004). Theory in progress: Dyadic power theory: Constructing a communication-based theory of relational power. *Journal of Family Communication*, *4*(3-4), 235-248.

Dunbar, N. E., Burgoon, J. K., Chen, X., Wang, X., Ge, S., Huang, Q., & Nunamaker, J. (2023). Detecting ulterior motives from verbal cues in group deliberations. *Frontiers in Psychology,* 14, 1166225.

Dunbar, N. E., Dorn, B., Hansia, M., Ford, B., Giles, M., Metzger, M., ... & Subrahmanian, V. S. (2021). Dominance in groups: How dyadic power theory can apply to group discussions. *Detecting trust and deception in group interaction*, 75-97.

Ekman, P. (2003). Emotions revealed: recognizing faces and feelings to improve communication and emotional life. New York. *NY: Times books*.

Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *semiotica*, 1(1), 49-98.

Elangovan, A. R., & Shapiro, D. L. (1998). Betrayal of trust in organizations. *Academy of management review*, *23*(3), 547-566.

Evans, J. R., & Michael, S. W. (2014). Detecting deception in non-native English speakers. *Applied Cognitive Psychology*, 28(2), 226-237.

Feely, A. J., & Harzing, A. W. (2003). Language management in multinational companies. *Cross Cultural Management: an international journal*, 10(2), 37-52.

Foddy, M., Platow, M. J., & Yamagishi, T. (2009). Group-based trust in strangers: The role of stereotypes and expectations. *Psychological Science*, *20*(4), 419-422.

Forsyth, D. R. (2010). Group dynamics (5th ed.). Wadsworth, Cengage Learning.

Fuller, C. M., Biros, D. P., & Delen, D. (2011). An investigation of data and text mining methods for real world deception detection. Expert systems with applications, 38(7), 8392-8398.

Ge, S. (2022). Trust and Distrust Network in Group Deception: An Exploratory study.

Giles, H. (Ed.). (2016). *Communication accommodation theory: Negotiating personal relationships and social identities across contexts*. Cambridge University Press.

Giles, H., & Baker, S. C. (2008). Communication accommodation theory. In W. Donsbach (Ed.), *The international encyclopedia of communication*. Chichester: Wiley. cdoi. org/10.1002/9781405186407.wbiecc067.

Gregersen, T. S. (2005). Nonverbal cues: Clues to the detection of foreign language anxiety. *Foreign language annals,* 38(3), 388-400. doi: 10.1111/j.1944-9720.2005.tb 02225.x

Guha, R., Kumar, R., Raghavan, P., & Tomkins, A. (2004). Propagation of trust and distrust. In *Proceedings of the 13th international conference on World Wide Web* (pp. 403-412).

Hall, J. A., Coats, E. J., & LeBeau, L. S. (2005). Nonverbal behavior and the vertical dimension of social relations: a meta-analysis. *Psychological bulletin*, *131*(6), 898.

Harrison McKnight, D., & Chervany, N. L. (2001). Trust and distrust definitions: One bite at a time. In *Trust in cyber-societies: Integrating the human and artificial perspectives* (pp. 27-54). Springer Berlin Heidelberg.

Hauch, V., Blandón-Gitlin, I., Masip, J., & Sporer, S. L. (2015). Are computers effective lie detectors? A meta-analysis of linguistic cues to deception. *Personality and social psychology Review*, *19*(4), 307-342.

Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual review of psychology,* 53(1), 575-604.

Ho, S. M., & Benbasat, I. (2014). Dyadic attribution model: A mechanism to assess trustworthiness in virtual organizations. *Journal of the Association for Information Science and Technology*, *65*(8), 1555-1576.

Ho, S. M., Hancock, J. T., & Booth, C. (2017). Ethical dilemma: Deception dynamics in computer-mediated group communication. *Journal of the Association for Information Science and Technology*, *68*(12), 2729-2742.

Hutchins, C. E., & Benham-Hutchins, M. (2010). Hiding in plain sight: criminal network analysis. *Computational and Mathematical Organization Theory*, *16*, 89-111.

Iacono, C. S., & Weisband, S. (1997, January). Developing trust in virtual teams. *In Proceedings of the Thirtieth Hawaii International Conference on System Sciences* (Vol. 2, pp. 412-420). IEEE.

Islam, A., Corney, M., Mohay, G., Clark, A., Bracher, S., Raub, T., & Flegel, U. (2011). Detecting collusive fraud in enterprise resource planning systems. In Advances in Digital Forensics VII: 7th IFIP WG 11.9 International Conference on Digital Forensics, Orlando, FL, USA, January 31–February 2, 2011, Revised Selected Papers 7 (pp. 143-153). Springer Berlin Heidelberg.

Islam, M. N., Haque, S. R., Alam, K. M., & Tarikuzzaman, M. (2009, December). An approach to improve collusion set detection using MCL algorithm. In *2009 12th International Conference on Computers and Information Technology* (pp. 237-242). IEEE.

Jensen, M. L., Meservy, T. O., Burgoon, J. K., & Nunamaker, J. F. (2008). Video-based deception detection. *Intelligence and security informatics: Techniques and applications,* 425-441.

K, A., T, I., & W, M. (2022). Fraud Detection in Stock Market Using Collusion Clustering Algorithm. *Advances in Intelligent Systems and Technologies*.

Kadushin, C. (2004). Introduction to social network theory. *Boston, Ma*, *307*.

Kamvar, S. D., Schlosser, M. T., & Garcia-Molina, H. (2003, May). The eigentrust algorithm for reputation management in p2p networks. In *Proceedings of the 12th international conference on World Wide Web* (pp. 640-651).

Keating, C. F., Mazur, A., Segall, M. H., Cysneiros, P. G., Kilbride, J. E., Leahy, P., ... & Wirsing, R. (1981). Culture and the perception of social dominance from facial expression. *Journal of personality and social psychology*, *40*(4), 615.

Kramer, R. M. (1999). Trust and distrust in organizations: Emerging perspectives, enduring questions. *Annual review of psychology*, *50*(1), 569-598.

Kumar, S., Bai, C., Subrahmanian, V. S., & Leskovec, J. (2021, May). Deception detection in group video conversations using dynamic interaction networks. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 15, pp. 339-350).

Lam, C. (2011). Linguistic politeness in student-team emails: its impact on trust between leaders and members. *IEEE Trans. Profess. Commun*. 54, 360–375.

Lazarus, R. S. (1991). Progress on a cognitive-motivational-relational theory of emotion. *American Psychologist*, 46(8), 819–834. https://doi.org/10.1037/0003-066X.46.8.819

Lewicki, R. J., McAllister, D. J., & Bies, R. J. (1998). Trust and distrust: New relationships and realities. *Academy of management Review*, *23*(3), 438-458.

Lewicki, R. J., & Tomlinson, E. C. (2003). Trust and trust building. *Beyond intractability,* 5(2), 114-139.

Luhmann, N. (1979). Trust and Power (John A. Wiley and Sons, Chichester).

Mazrooei, P., Archibald, C., & Bowling, M. (2013, June). Automating collusion detection in sequential games. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 27, No. 1, pp. 675-682).

McCullagh, P. D., & Landers, D. M. (1976). Size of audience and social facilitation. *Perceptual and Motor Skills*, *42*(3_suppl), 1067-1070. https://doi.org/10.2466/pms.1976.42.3c.1067.

Megvii, (2022). Face++ Cognitive Services Face++ API's Reference. https://www.faceplusplus.com/attributes/

Moreland, R. L. (2010). Are dyads really groups?. *Small Group Research*, *41*(2), 251-267.

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and social psychology bulletin*, *29*(5), 665-675

Ortega Ruiz, R., Rey Alamillo, R. D., & Casas Bolaños, J. A. (2012). Knowing, building and living together on internet and social networks: The ConRed cyberbullying prevention program. *International Journal of Conflict and Violence*, 6 (2), 302-312.

Pak, J., & Zhou, L. (2014). Social structural behavior of deception in computer-mediated communication. *Decision Support Systems*, *63*, 95-103.

Palshikar, G. K., & Apte, M. M. (2008). Collusion set detection using graph clustering. *Data mining and knowledge Discovery*, *16*, 135-164.

Pentland, S. J., Spitzley, L., Chen, X., Wang, X., Burgoon, J. K., & Nunamaker, J. F. (2021). Behavioral indicators of dominance in an adversarial group negotiation game. *Detecting Trust and Deception in Group Interaction*, 99-122.

Pentland, S. J., Twyman, N. W., Burgoon, J. K., Nunamaker, J. F. Jr., and Diller, C. B. (2017). A video-based screening system for automated risk assessment using nuanced facial features. J. *Manag. Inform. Syst.* 34, 970–993. doi: 10.1080/ 07421222.2017.1393304

Rahman, D. (2014). The power of communication. *American Economic Review*, *104*(11), 3737-3751. http://www.jstor.org/stable/43495353

Serva, M. A., Fuller, M. A., & Mayer, R. C. (2005). The reciprocal nature of trust: A longitudinal study of interacting teams. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, *26*(6), 625-648.

Shahriari, M., & Jalili, M. (2014). Ranking nodes in signed social networks. *Social network analysis and mining*, 4, 1-12.

Smed, J., Knuutila, T., & Hakonen, H. (2006). Can we prevent collusion in multiplayer online games. In *Proceedings of the Ninth Scandinavian Conference on Artificial Intelligence* (Vol. 9).

Smed, J., Knuutila, T., & Hakonen, H. (2007). Towards Swift and Accurate Collusion Detection. In *GAMEON* (pp. 103-107).

Spitzley, L. A., Wang, X., Chen, X., Pentland, S. J., Nunamaker, J. F., Burgoon, J. K., & Dunbar, N. E. (2022). Non-invasive measurement of trust in group interactions. *IEEE Transactions on Affective Computing*, *14*(3), 2389-2401.

Sporer, S. L., & Schwandt, B. (2006). Paraverbal indicators of deception: A meta-analytic synthesis. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *20*(4), 421-446.

Sutar, R., Chaudhary, P., & Yadav, V. (2022). Prevalence of collusion in cancer communications: A meta-analysis. *Psycho-Oncology*, *31*(3), 372-387.

Tajfel, H., & Turner, J. C. (2004). The social identity theory of intergroup behavior. *In Political psychology* (pp. 276-293). Psychology Press.

Tang, J., Chang, Y., Aggarwal, C., & Liu, H. (2016). A survey of signed network mining in social media. *ACM Computing Surveys (CSUR),* 49(3), 1-37.

Thompson, A. D., & Hartwig, M. (2023). The language of high-stakes truths and lies: Linguistic analysis of true and deceptive statements made during sexual homicide interrogations. *Legal and criminological psychology*, 28(1), 34-44.

Vallve-Guionnet, C. (2005, April). Finding colluders in card games. In *International Conference on Information Technology: Coding and Computing (ITCC'05)-Volume II* (Vol. 2, pp. 774-775). IEEE.

Valacich, J. S., Dennis, A. R., & Nunamaker Jr, J. F. (1992). Group size and anonymity effects on computer-mediated idea generation. *Small Group Research*, *23*(1), 49-73.

Vernham, Z., Granhag, P. A., & Mac Giolla, E. (2016). Detecting deception within small groups: A literature review. *Frontiers in psychology*, *7*, 186099.

Vrij, A. (1993). Credibility judgments of detectives: The impact of nonverbal behavior, social skills, and physical characteristics on impression formation. *The Journal of social psychology*, *133*(5), 601-610.

Watzlawick, P., Bavelas, J. B., & Jackson, D. D. (2011). Pragmatics of human communication: A study of interactional patterns, pathologies and paradoxes. WW Norton & Company.

Wasserman, S., & Faust, K. (1994). Social network analysis: Methods and applications.

Wu, J., & Liu, Y. (2020). Deception detection methods incorporating discourse network metrics in synchronous computer-mediated communication. *Journal of Information Science*, *46*(1), 64-81.

Wu, J., & Zhou, L. (2015). DOBNet: exploiting the discourse of deception behaviour to uncover online deception strategies. *Behaviour & Information Technology*, *34*(9), 936-948

Yan, J. (2010, July). Collusion detection in online bridge. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 24, No. 1, pp. 1510-1515).

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., ... & Sun, M. (2020). Graph neural networks: A review of methods and applications. *AI open*, *1*, 57-81.

Zhou, L., Burgoon, J. K., Twitchell, D. P., Qin, T., & Nunamaker Jr, J. F. (2004). A comparison of classification methods for predicting deception in computer-mediated communication. *Journal of Management Information Systems*, *20*(4), 139-166.

Zhou, L., Wu, J., & Zhang, D. (2014). Discourse cues to deception in the case of multiple receivers. *Information & management*, *51*(6), 726-737.

Zhou, L., & Zhang, D. (2006). A comparison of deception behavior in dyad and triadic group decision making in synchronous computer-mediated communication. *Small Group Research*, *37*(2), 140-164.

Zhou, L., Zhang, D., & Sung, Y. W. (2013). The effects of group factors on deception detection performance. Small Group Research, 44(3), 272-297. https://doi.org/10.1177/1046496413484178.

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In *Advances in experimental social psychology* (Vol. 14, pp. 1-59). Academic Press.

Zuckerman, M., & Driver, R. E. (2014). Telling lies: Verbal and nonverbal correlates of deception. In *Multichannel integrations of nonverbal behavior* (pp. 129-147). Psychology Press