JOURNAL OF DIGITAL SOCIAL RESEARCH

WWW.JDSR.IO ISSN: 2003-1998

VOL. 4, NO. 4, 2022, 89–129

HIVE MIND ONLINE: COLLECTIVE SENSING IN TIMES OF DISINFORMATION

Shuyuan Mary Ho[a], Jeffrey Nickerson[b] and Qian Zhang[a]

ABSTRACT

This study investigates the efficacy of collective sensing as a mechanism for unveiling disinformation in group interaction. Small group interactions were simulated to experiment on the effects of a group reaction to incentivized deceptive behavior when initiated by social influencers. We use multilevel modeling to examine the individual communication data nested within group interactions. The study advances the use of computational efficacy to support the supposition of collective sensing—by analyzing individual social actors' communicative language and interaction within group contexts. Language-action cues as stigmergic signals were systemically extracted, compared and analyzed within groups as well as between groups. The results demonstrate that patterns of group communication become more concentrated and expressive after a social influencer becomes deceptive, even when the act of deception itself is not obvious to any individual. That is, individuals in the group characterize deceptive situations differently, but communication patterns depict the group's ability to collectively sense deception from circulating disinformation. The study confirms our postulation of using collective sensing to detect deceptive influences in a group.

Keywords: computer-mediated communication, computer-mediated deception, collective sensing, collective intelligence, human sensor networks, language-action cues, multilevel models, information manipulation, disinformation.

[a] Florida State University, U.S.A.

[b] Stevens Institute of Technology, U.S.A.

1 INTRODUCTION

Given this era of prolific social media, false information spreads more efficiently than ever before across our society. The tendency to exaggerate makes false information travel faster (Vosoughi, Roy, & Aral, 2018). That is, more sensational information is more likely to be spread (Berger, 2016; Liu & Li, 2019). False information is generated and manipulated for many reasons, including antisocial disinformation campaigns that can have deleterious effects on democratic processes, as with false claims of electoral fraud (Dozier & Bergengruen, 2021; Holland, Mason, & Landay, 2021), and on public health, as with false claims about the vaccine side-effects (McCarthy, 2020; Milman, 2020). The consequence of false information—whether intentional deception leading to disinformation or unintentional mistakes characterized as misinformation—can influence public opinion, obscures truth, and can be the result of malicious intent, political bias, ignorance, and even competition among social influencers. Mis-informed citizens share their beliefs and opinions online, furthering misinformation. Personally-skewed opinions by social influencers—facilitated by enabling technologies—can manipulate collective perceptions so easily that proven facts and truths can be undermined.

Disinformation is often distributed without ethical consideration, and the ensuing public opinion can have a significant impact on legislative policymaking. Disinformation has a tendency to radicalize American politics (Benkler, Faris, & Roberts, 2018) and polarize Russian politics (Bodrunova, Blekanov, Smoliarova, & Litvinenko, 2019). There is growing evidence of unreliable sources quickly and knowingly spreading half-truths and/or false information through social media and websites for unethical reasons (Kim & Dennis, 2019; Kim, Moravec, & Dennis, 2019). Governments can manipulate facts and blur reality to disrupt and control public narratives. Disinformation spread by political campaigns is an example of a greater threat—information warfare (Boxwell, 2020a, 2020b). The impact of a deceptive social influencer—whether an individual or a state-sponsored agency—can be significant and detrimental to public discourse.

Responding to the threat of disinformation, this study explores the theoretical underpinnings and computational efficacy of collective sensing as a possible avenue to recognize disinformation. Our theoretical stance posits that some of the same kinds of behavior that spread disinformation may also provide clues to detecting it. For example, human sociality leads to herding behavior (Raafat, Chater, & Frith, 2009). This tendency is exploited by those who engage in intentional deception. At the same time, humans are attuned to complex social relationships and anticipate the behavior of others. If, in a goal-driven team activity, expectations are not fulfilled, this may not be enough to infer the presence of deception. But it

may have effects on the nature of interactions that follow deceptive acts, even if individuals are not conscious a deception has taken place. That is, signals may emerge from the collective sensing of the group (Dipple, Raymond, & Docherty, 2014). This study examines the basic mechanisms for analyzing a group's collective reactions—subjectively and objectively—by observing overt online communication behavior over time, during the covert spread of disinformation. Group interactions are examined to find patterns in how groups perceive and react to disinformation from a deceptive influencer. We ask: *Can collective sensing help unveil disinformation*?

This work points toward the possibility of using collective sensing to detect deceptive influence in a group. It also describes a technique for measuring the changes in state of a collective caused by deception through analysis of the collective's stigmergic traces. Stigmergy, a term coined in studies of social insects (Grassé, 1959), is a mechanism in which an action performed by an agent leaves traces on the environment that in turn stimulate new action by another agent. The term has also been applied to traces left in digital environments by humans (Rezgui & Crowston, 2018). Human teams not only have the intuitive ability to coordinate, but also the innate capability to sense and respond to anomalies. In this article, we first conceptualize false information, and identify the complexity and challenges of false information. Then, we conceptualize collective sensing; specifically, the human sensor's ability to detect disinformation based on stigmergy in group interactive contexts. The core components of stigmergy include actors (agents), language-action cues (signs) and group interactive contexts mediated by technologies (environment). Three research hypotheses are raised. Based on these core components, we discuss the research design that stimulates stigmergy in groups' pairwise communication, including data collection and process. Multilevel modeling approaches are adopted to analyze data in responding to hypotheses. Both theoretical and practical implications and study limitations are iterated in the seventh section. The paper concludes with contributions, along with potential directions for future research.

2 INFORMATION: *AUTHENTIC*, *FALSE*, *DIS-*, OR *MIS-*?

Disinformation is not a new phenomenon, but an age-old form of warfare; a strategy utilized by an ill-intentioned opponent to destabilize a situation or a society. This adversary strategy was utilized just after World War II by Mao Tse-Tung to divide China. Disinformation has been especially prominent in the 21st century due to the proliferation of the Internet and social media adoption. Distorted information can spread instantly across the world, making it nearly impossible to propagate the truth or a retraction (Allcott & Gentzkow, 2017; Tendoc Jr., Lim, & Ling, 2017). In this section,

we attempt to define disinformation and conceptualize differences between disinformation and misinformation. Then, we explore the complexity and challenges of uncovering disinformation and differentiating between types of social influences—active vs. passive—based on the degree of disinformation.

2.1 The concept of disinformation

Research has deliberated about what constitutes disinformation. Although there is no unanimous agreement on the definition, many studies have converged on similar ideas. Hernon (1995) differentiated *disinformation* from *misinformation* by the measure of intent. Disinformation refers to inaccurate information as a result of "a deliberate attempt to deceive or mislead"—whereas misinformation is defined as the result of an honest mistake (p. 134). Fetzer (2004a) described disinformation as "misinformation with an attitude" (p. 231), and highlighted how fallacious and incomplete information can be disseminated in an intentional, deliberate, purposeful effort to mislead, deceive or confuse (Fetzer, 2004b, p. 228). It is worth noting that premeditated lies are different than false, mistaken or misleading information as a result of unintentional consequence. When the above criteria exist without the motive of duping people, it is considered *misinformation* (Fetzer, 2004a; Hernon, 1995). When information is knowingly false with the intention to mislead or deceive, these assertions would qualify to be called "lies" (Fetzer, 2004b, p. 232). However, not all false information is asserted deliberately (i.e., misinformation); and not all false claims—even asserted deliberately—can be categorized as lies, if lacking an intent to mislead. A fundamental component of disinformation remains that it is intended to deceive and confuse for some sort of gain (Fetzer, 2004a). Events like political campaigns, advertisements, and editorials tend to attract disinformation because they are driven by the intention of private gain, thus giving these topics the element of ulterior motive (Fetzer, 2004b). Different degrees of disinformation and misinformation are both considered computer-mediated deception (Ho, Hancock, Booth, & Liu, 2016) in this era of social media.

2.2 The challenges in studying disinformation

Disinformation presents many challenges. First, it is a *psychological-behavioral* problem. Manipulation of information is a complex problem of behavioral intent that is fundamentally difficult to observe, detect, or predict. Human behavior often changes for neutral reasons, or simply reflects a change of habit. Although changes in a person's behavior can be

observed, the reasons for those observed behaviors are typically unknowable. Behavioral change can be captured and analyzed to identify patterns. However, changes in a person's behavior may occur for a variety of reasons, and a change in intention may not always be noticeable. Of course, not all behavioral changes reflect malicious behavior. Moreover, behavioral changes of benign actors can result in the propagation of disinformation originated by malicious actors.

While it is difficult to classify a single deceptive individual's behavioral intent with statistical significance (Ho et al., 2015; Ho, Hancock, & Booth, 2017), we conjecture that it may be easier to perceive the changes in a group's collective behavior in response to disinformation. Studying the collective level presents its own challenges. The types of deceptive influence will vary in degrees—from "passive" to "active" (Caddell, 2004; Levine, 2014). Kimmel (1998) classified deceptive influences as either being passive; "withholding relevant information" or active; "bluntly misleading" others (p. 804). Differences in deceptive influence will lead to differences in groups' reactions and responses (Ezeakunne, Ho, & Liu, 2020). Thus, we conjecture that a group's collective sensing and the resulting changes in communication patterns may differ depending on the type of deception.

3 COLLECTIVE SENSING AND THE DETECTION OF DISINFORMATION

To understand the application of collective sensing in the detection of disinformation, we explore language-action cues in collective group interaction. This section culminates with three hypotheses, developed based on the manifestation of collective language-action cues as being representative of collective attributes of a team in response to disinformation instilled by a social influencer. This section starts by conceptualizing how collective sensing works, and then moves to focus on the dependent variables of the hypotheses: *expressiveness*, *cognitive processing*, and *affective processing*.

Collective language-action cues are stigmergic signals yielded by interpersonal communication (Ho & Hancock, 2018), and aggregated they constitute collective attributes. These attributes include levels of *expressiveness, cognitive processing*, and *affective processing*. These attributes can be measured through the number of words used, the number of words associated with insight, causation, discrepancy, certainty, inclusivity and exclusivity, and the number of words associated with positive and negative emotion (Pennebaker & King, 1999; Pennebaker, Mehl, & Niederhoffer, 2003). We posit that stigmergic signals, constituted by collective language-action cues, can be indicative of the subtle intent of an interacting social influencer in group interaction context. Ho (2019) proposed the leader

member exchange as an interactive framework to uncover a deceptive influencer. That is, differences in overall *expressiveness*, *cognitive* and *affective processing* are recognizable in groups with a deceptive social influencer when compared to groups without a deceptive social influencer. Moreover, noticeable differences in group reaction can also be identified when comparing the impacts of a deceptive influencer before and after spreading disinformation.

3.1 Collective sensing

In information systems, a recent look at collectivity highlighted the collective use of technology in team-based processes (Negoita, Lapointe, & Rivard, 2018). Virtual teams are characterized by complex and ongoing relations (Majchrzak, Rice, Malhotra, King, & Ba, 2000) and these relations may produce outputs and state changes that are more than sum of their component parts. Moreover, technology-mediated team processes produce digital traces, which can facilitate analysis that recognizes changes in collective state.

Collective sensing is a component of collective intelligence. Collective intelligence refers to a complex behavioral phenomenon created by simple interaction between individuals within groups that follow basic rules, and is generally defined as "the ability of a group to solve more problems than its individual members" (Heylighen, 1999). Collective intelligence can be observed in group-based or interpersonal pairwise interaction as pairwise interaction can facilitate our cognitive understanding of each other. As language provides the means for effective communication, language-action cues are the vehicles that transmit environment-mediated stigmergic signals to communicators in groups or interpersonal communication. But the collective intelligence literature generally focuses on groups whose members are aligned toward the accomplishment of a shared goal; many of the experimental studies asked groups to build a structure or solve a problem under the assumption all members shared the goal (for example, Woolley, Aggarwal, and Malone (2015); (2010)). Such an assumption cannot be made when a member of a group is intentionally deceptive: such an individual is bent on sabotaging a group, and, if the individual succeeds, there can be no measure of collective intelligence based on group success.

Collective sensing is a more specific process that can apply in situations in which deception may be present. It refers to an emergent phenomenon that comes from the ability of a group to sense more than the individual can (Bennati, 2018). Small changes in individual responses to individual contributions when aggregated provide indications of collective state. The concept has been studied in biology where, for example, collective or quorum sensing has been shown to operate in cells, in ant

colonies (Gordon, 2014), and in animal flocks (Berdahl, Torney, Ioannou, Faria, & Couzin, 2013). This work has been extended to model humans as sensors in urban environments (Blaschke, Hay, Weng, & Resch, 2011; Resch, 2013), and to recognize the ways such sensing might help identify hazards (Yang, Ahn, Vuran, & Kim, 2017).

As with many group level phenomena, it may be difficult to recognize what the group has sensed without somehow aggregating the traces of their interaction. In situations that are deceptive, we postulate that collective sensing can recognize deception without any individual being able to call it out. The recognition of deception in a group will manifest in subtle communication patterns. These in turn will leave stigmergic traces that can be analyzed computationally (Crowston, O'sterlund, Howison, & Bolici, 2017; Dipple et al., 2014). This conceptual framing is shown in *Figure 1*.

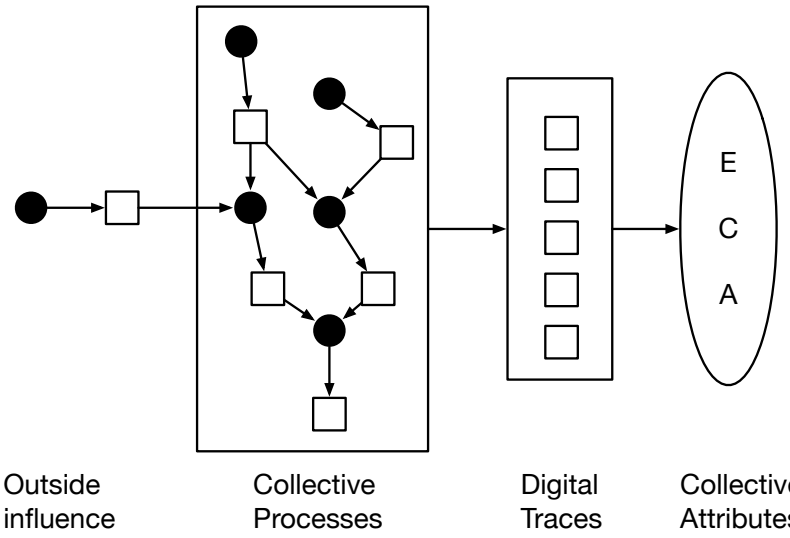

Figure 1. From outside influence to collective attributes. Squares indicate digital artifacts such as text messages interchanged through team collaboration software. Circles indicate people.

In Figure 1, a virtual team (Majchrzak et al., 2000), is engaged in a process of collaboration when an outside influence affects a team member. For example, a corrupting outside influence might present a team member with an incentive to deceive other team members. The collective processes of the team continue. Even though team members are not aware of deceptive acts taking place, those acts may act like a pebble thrown onto a foggy pond, leaving barely detectable ripples. The sensemaking processes of the team may change in subtle ways. The digital traces of these collective processes, when extracted and analyzed, may be characterized by collective attributes, such as the degree of the expressiveness of the team conversation, as well as the level of cognitive and affective processing. Moreover, deception has different styles, and these different styles affect the collective differently.

The core components of stigmergy include actors (*agents*), cues (*signs*) and contexts as mediated by cooperative technologies such as social media (*environment*). Differences in context, differences among the parties involved, differences in time, place, and even communication medium—each influence not only what is communicated, but also how communication is perceived. In the context of computer-mediated communication, language-action cues can become important indicators in identifying computer-mediated deception (Ho et al., 2016). Cues found in language help enable collective understanding, and thus provide communication context for collective sensing. In this regard, communicative signals shared by communicators are also critically important in identifying deceptive intent in computer-mediated communication (CMC).

A simple example of collective sensing can be seen in ants' or bees' ability to map out their environment (Gordon, 2019). Individually, these insects experience limited capacity in processing information; however, collectively they can decide on the fields to exploit any potential dangers about to occur. Collective cognition is demonstrated and communicated through a stigmergic signal (Heylighen, 1999), or a genome (Malone, Laubacher, & Dellarocas, 2010). Based on these biological analogs, Dipple et al. (2014) proposed a macro-level view of the communication mechanism that triggers responses in human society. It includes three components: the agents, the environment, and the signs. The agent's ability to coordinate, to sense, or to detect anomalies depends on their interpretation of meaning as mediated by the manifestation of stigmergic signals and signs when interacting. These core components also correspond to the fundamentals of human sensors' ability to interpret and sense deceptive communication.

Agents. To combat intentional false information, or disinformation, our ability to understand the efficacy of collective sensing with regards to inferring subtle intent—as manifested in language-action cues by social actors or influencers—is vital in anticipating disinformation. Ho et al. (2017) examined and compared social actors' reactions and responses to manipulation by a deceptive social influencer. Based on an interactive framework, collective language-action cues from human sensors interactions were framed as stigmergic signals that collectively sensed during computer-mediated deception within a group context (Ho & Hancock, 2018, 2019).

Signs. Computer-mediated communication enables information to be transferred to a message receiver through words and patterns of communication cues. These cues and patterns can reveal both overt and covert intent of a message sender. However, in this cue-lean environment, the availability of such cues is effectively limited to the text itself—without the physical cues in face-to-face communication. Nonetheless, even in 'cue

lean' text-based communication, there can be linguistic and syntactical clues of deception. For example, Newman, Pennebaker, Berry, and Richards (2003) suggested that the overuse of sensory or spatiotemporal words, and changes in the diversity and complexity of language can be indicative of deception. Zhou, Burgoon, Nunamaker Jr., and Twitchell (2004) suggested that deceivers tend to be more casual and expressive in their linguistic style. The level of detail (too much or too little) are clues to deception in both face-to-face and CMC settings. Deceivers using CMC particularly tend to be wordier than truth-tellers, but the additional detail provided is not necessarily relevant or meaningful in context (Zhou & Zhang, 2004). Moreover, Hancock, Curry, Goorha, and Woodworth (2008) discovered that deceivers tend to use more sense-based words (e.g., seeing, touching), fewer self-oriented and more other-oriented pronouns in text-based CMC. Enabled with multiple cues and immediate feedback, richer media can increase the ability of message receivers to perceive and thus facilitate the detection of deception (Kahai & Cooper, 2003). However, simply knowing these linguistic cues does not help conversational partners to improve deception detection. Hancock et al. (2009) suggested that certain language-action cues (e.g., first-person references, words of emotion or inhibition, etc.) have been shown to be effective indicators to distinguish deceivers from truth tellers. While, in general, humans are not good at detecting deception, Ho et al. (2016) computationally identified deceivers' strategies through the use of salient language-action cues, which included the use of words associated with *affective* processes, *cognitive* processes, self- and other- references, as well as the use of peripheral expressions and overall wordiness.

The importance of immediacy cues and the representation of these cues illustrate psychological elements of communication (J. K. Burgoon, Blair, Qin, & Nunamaker, 2003; Judee K. Burgoon & Buller, 1994; Judee K. Burgoon, Buller, Dillman, & Walther, 1995). A message sender often employs cues to associate (or distance) him/herself *physically* or *psychologically* from the content of a message (Mehrabian, 1968, p. 203). Buller and Burgoon (1994) noted that deceivers often use both verbal and *non*verbal means to "distance [themselves] from others, to disaffiliate, and to close off scrutiny or probing communication" (p. 204). Similarly, social distance theorists (c.f. DePaulo, Kashy, Kirkendol, Wyer, & Epstein, 1996) suggested that a deceptive actor will try to minimize potential cues to reduce the cognitive load associated with deception, by adopting a cue-lean communication mode or style and thereby limiting opportunities for others to question or engage in conversation. In face-to-face communication, a deceiver can create psychological distance by exhibiting literal (physical) distance (e.g., standing/ sitting remotely from the conversational party), or by choosing to interact via the telephone rather than meeting physically

(Mehrabian, 1968). Likewise, in CMC, psychological distance can be created through word choice and phraseology—that is, by minimizing immediacy (Buller & Burgoon, 1994). Word choice and overall tone that suggest negative feelings (such as disappointment, frustration, or even anger) can be a sign of distancing, while word choice and tone suggesting a positive relationship—perhaps conveying humor or praise—can foster a positive, trusting relationship between communicating actors.

Environments. Social media has significantly changed people's information behavior in producing, sharing, using, and disseminating information contents. With the pervasiveness of social media, disinformation and fake news circulates more readily and reaches more people (Garrett, 2017). Chen, Sin, Theng, and Lee (2015), for example, identified that students are prone to exchanging misinformation in social media. Mocanu, Rossi, Zhang, Karsai, and Quattrociocchi (2015) identified that people who tend to interact with conspiracies' information sources are more likely to be exposed to intentionally false claims. Although people are increasingly aware of the presence of unsubstantiated or untruthful rumors, once fake news has already reached its targets, it is highly unlikely to be corrected. Furthermore, this form of disinformation tends to foster a collective credulity because of its pervasiveness (Mocanu et al., 2015).

Social media provides an environment for people to communicate, share interests, opinions and information (Chen et al., 2015). It has gained popularity and priority in setting policies (Hernon, 1995), and political campaigns (Garrett, 2017). In social media, individuals tend to subscribe to and engage in activities or information reflecting their viewpoints. As a result, regardless of the authenticity of the information, people tend to endorse and affirm information they agree with. Garrett (2017) describes this social media phenomenon as an echo chamber or a filter bubble. People tend to believe information shared within their trusted circle of friends and colleagues without verification or validation of the reality and truth. This creates a real threat to our understanding of reality and truth from the overwhelming amount of unverified information and the unprecedented proliferation of conspiracy-related disinformation (Mocanu et al., 2015). The social resources available for fighting this phenomena is finite, while its proliferation is moving at an accelerated rate (Garrett, 2017).

3.2 Hypothesis 1: Expressiveness

We consider the *expressiveness* of communication within groups that include a deceptive influencer, and we expect to observe the differences through groups' *expressiveness* in response to the influencer's deceptive intent. Trust influences and impacts not only interpersonal relationships, but also the relationships within, between and among groups (Hosmer, 1995). Group

trust depends on the interaction and relationships between group members and their influencers. Members develop exchange relationships with their social influencers (Wayne, Shore, & Liden, 1997), and group trust can be undermined or enhanced by a social influencer's behavior and leadership style. Dansereau, Graen, and Haga (1975), Liden and Graen (1980) proposed a dyadic exchange between the social influencer and subordinates. Liden, Wayne, and Stilwell (1993) suggested that social influencers tend to develop different leadership styles, relationships or exchange with different members. While different leadership styles engender different group communication patterns and performance outcomes (Liden, Erdogan, Wayne, & Sparrowe, 2006), members can also develop different types of social exchange relationships with peer group members as well as with their immediate social influencers. One's trust toward the social influencer can be impacted (i.e., reduced or lost) if they feel betrayed. This breach of trust occurs as a result of incongruence (Morrison & Robinson, 1997) arising from a violation against the reciprocal exchanged agreements (often referred to as a "psychological contract") by and among group members (Robinson, 1996; Simons, 2002). The loss of group trust can also result in the loss of an influencer's credibility within the group, which can further impact his/her leadership (c.f., Simons, 2002). The group's perception of the social influencer's credibility is a primary source of influence leveraged by the influencer to manage the group. However, the credibility assessments/perceptions can change over time (George, Giordano, & Tilley, 2016). Loss of credibility within the group can trigger suspicion and motivate other members to exchange conversations in the awareness of uncertainty and deception—depending on the group sensitivity (Ho, 2009). The sensitivity of the group members does not depend on average individual intelligence (Woolley et al., 2015; 2010), but more on the average social sensitivity of group members, and the way group members interact (i.e., expressiveness, cognitive and affective processing).

Ho and Hancock (2018); (2017) suggested that groups will often sense acts that are reflective of deception, and this can result in more conversations within the group. Thus, we anticipate that groups with a deceptive influencer will stimulate more conversations, and thus hypothesize that groups with a deceptive influencer actively spreading disinformation will exhibit a higher overall expressiveness than groups without a deceptive influencer. That is, we would expect the overall expressiveness of groups with a deceptive influencer will be higher than groups without, because a deceptive influencer actively spreading disinformation would stimulate more expression and words. However, the group will display lower levels of *expressiveness* if the deceptive influencer conceals his/her intent. Accordingly, we hypothesize:

H1. Communication within groups that include a deceptive influencer actively spreading disinformation will display higher levels of expressiveness, but when the act of deception is concealed and passive, the group will display lower levels of expressiveness.

3.3 Hypothesis 2: Cognitive processing

Ekman and Friesen (1969) characterized deception as the purposeful concealment of the truth, either by omission or commission. Deception typically involves a persuasive, strategic process by which a deceiver transmits messages that have been deliberately distorted and/or manipulated, with the intention of misleading or misdirecting a receiver into reaching a wrong conclusion, or otherwise fostering a false belief, often for the deceiver's own benefit (Buller & Burgoon, 1996). Schultz (2002) speculated on a set of behavioral cues—e.g., deliberate markers, meaningful errors and verbal cues—to uncover a deceptive influencer's behavior. Greitzer, Kangas, Noonan, Brown, and Ferryman (2013); (2012) further created a behavioral/ psychosocial model to identify deceptive influencers. These psychosocial indicators include disgruntlement, not accepting feedback, anger management issues, confrontational behavior, and self-centeredness. Regarding verbal cues in asynchronous interpersonal communication, Zhou et al. (2004) suggested that deceptive actors tend to be more casual and expressive in their linguistic style. Level of detail (too much or too little) may also be indicative of deception, and that deceptive actors tend to be wordier than non-deceptive actors, but the additional detail is not necessarily relevant or meaningful in context (Zhou & Zhang, 2004). Brown, Greitzer, and Watkins (2013); (2013) and Taylor et al. (2013) also found language differences linked to psychological instability such as self-centeredness, negativity affect, and more cognitive processes when compared with their co-workers in asynchronous communication (i.e., email). Brown, Watkins, et al. (2013) translated observed linguistic cues into behavioral categories identified as corresponding to behaviors significantly associated with deceptive actors.

As an deceptive influencer attempts to conceal his/her hidden agenda, this unconscious/subconscious behaviors can trigger nonverbal *leakage* (Ekman & Friesen, 1969), and groups interacting with this deceptive influencer can be expected to use more words associated with *cognitive* processes. In group interactive context, the differences in *cognitive* language-action cues can be identified, not only between deceptive influencers and non-deceptive influencers, but also between groups with a deceptive influencer and groups without one (Ho et al., 2017). While the clues regarding a single deceptive influencer's behavior tend to be subtle or even unnoticeable, the language-action cues of groups interacting with a

deceptive social influencer can be found when compared to groups without one. That is, the group's collective *cognitive processes* (i.e., words connoting inclusion, exclusion, certainty and insight), and these types of language-action cues will result in changing patterns of group communication. The group will sense and reflect a change with more *cognitive* response to the spread of disinformation. As an influencer attempts to disguise deceptive intent, groups may collectively react to the information behavioral change by the influencer, and thus display less *cognitive* processing (i.e., certainty, inclusion, suggestions, or insight). Accordingly, we hypothesize that:

> *H2. Communication within groups that include a deceptive influencer actively spreading disinformation will display higher levels of cognitive processing, but when the act of deception is concealed and passive, the group will display lower levels of cognitive processing*.

3.4 Hypothesis 3: Affective processing

Pennebaker and King (1999); (2003) suggested that a deceptive actor would display more negative emotion. When a deceptive actor breaches a psychological contract by deceiving (or attempting to deceive) group members, the dynamics of group communication become more complex than deception in interpersonal communication. The deceptive influencer's persuasive strategy must account for multiple, interactive perspectives, which often requires a deceptive influencer to leverage and combine *cognitive* and *affective* processes. That is, regardless of whether trust within a group is built on cognitive factors, affective factors or a combination, when trust is violated by a deceptive influencer, subordinates' perception and commitment may be negatively influenced (Griffith, Connelly, & Thiel, 2011). In a group interactive context, Griffith et al. (2011) also suggested that a deceptive influencer's changed behavior can infect an associated group with negative emotions. While the behavioral change of a deceptive influencer may initially prompt analytical discussion amongst social actors (i.e., the use of cognitive-process words reflecting uncertainty, discrepancy, insight, causation, question, etc.), the interaction within groups having a deceptive influencer will suppress over emotional response (e.g., confusion, concern, emotion, frustration, or even apathy) (Ho et al., 2017). We thus posit that the group will display lower collective *affective processes* after deception has been initiated. When a deceptive influencer conceals his/her hidden agenda, groups including a deceptive influencer are likely to react to the concealment and stimulate more *affective* processing whereas if a deceptive influencer actively spreads disinformation, the interacting group may display less *affective* processing. Accordingly, we hypothesize that:

> *H3. Communication within groups that include a deceptive influencer actively spreading disinformation will display lower levels of affective*

processing, but when the act of deception is concealed and passive, the group will display higher levels of affective processing.

These hypotheses are represented diagrammatically in *Figure 2*.

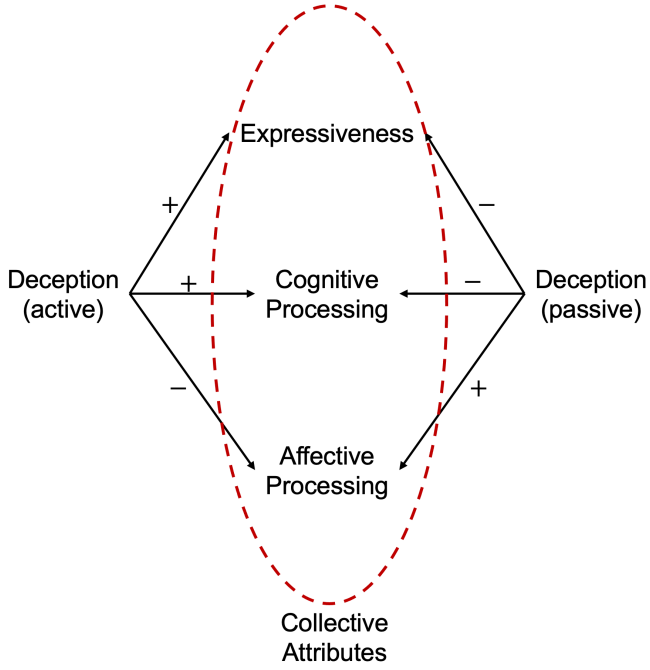

Figure 2. Diagrammatic representation of the hypotheses.

4 METHOD

Communication logs from experimental groups—depicting objective behavioral traces of the group interaction during deceptive situations—were analyzed and compared. Social influencers were randomly assigned to receive either a treatment or a placebo. Treatments allow deception to naturally occur within social influencers in treatment groups. Both intragroup (within groups that contain a deceptive influencer) and intergroup (between groups with and without a deceptive influencer) comparisons were analyzed. Small group interactions were designed and simulated to facilitate interactions (i.e., communication) when actors are given opportunities to interact with one another closely in a synchronous cue-lean, text-based CMC environment (Ho & Warkentin, 2017). As group size and task characteristics will reflect differences of people interacting in groups, Hackman and Vidmar (1970) empirically proved that optimal performance is found in groups having four to five members (pp. 48-49). Wheelan (2009) also confirmed that groups containing three to six members were significantly more productive than larger groups.

4.1 Ethical considerations

There are always ethical issues to confront when performing research related to deception. The investigators worked with IRBs to design interventions that would be unlikely to cause discomfort or stress. All experiments were debriefed in person, with the experimental design revealed to participants. All participants had opportunities to opt out of the research and/or to discuss the experiments or responses with an IRB representative. The goal of the design was to minimize potential harm while potentially providing benefit. The issues are complex. Deception itself—for example, financial fraud—can cause harm. And the surveillance to detect deception can also cause harm (as can experiments), unless designed carefully.

4.2 Data collection

Two identical experiments were conducted at two different research institutions. The first experiment was conducted at a large northeastern university in 2008[1] using Blackboard as the data collection platform. This participant group consisted of 26 participants (62% males, 38% females), ranging in age from 20 to 65 years. The second experiment was conducted at a large southeastern university in 2015[2], with all data collected using Google+ Hangout. This participant group consisted of 27 participants (63% males, 37% females), ranging in age from 18 to 65 years. Participants were recruited based on convenience sampling strategies in both experiments, largely from the student population of the respective academic institutions. Recruited participants were then randomly assigned to virtual groups. The participants' names were replaced with pseudo-names to protect privacy. We removed two (2) groups' data from the study because the influencers in these groups did not follow instructions: they didn't perform deceptive acts to spread disinformation. There are a total of 10 experimental groups of interaction data selected as the final dataset.

4.3 Research design

Between these experiments, a total of five (5) control groups and five (5) treatment groups were involved. Each group consisted of four to six members, and each group included one social influencer. All virtual teams were put into a scenario where they were to compete with one another by

[1] Human subject research was approved by the University's Institutional Review Board (IRB) under protocol #07-276.

[2] Human subject research was approved by the University's Institutional Review Board (IRB) with the protocols 2013.10910, 2014.12923, and 2015.15316.

solving brain teasers and math problems within a predetermined timeframe. Each team's collective goal is to be number one in the competition, and they are provided shared public information on where they stand in the competition. Each member shares in a collective financial incentive based on the standing of the member's team at the end of the competition. The social influencers in the treatment groups were financially incentivized to introduce false information, whereas the social influencers in the control groups were not incentivized (Ho & Warkentin, 2017). This additional financial incentive, which is private information given only to the social influencers in the treatment groups, affords social influencers the ability to make autonomous decisions in communicating information—whether authentic or false—within their own group. This financial incentive was provided in a communication to the influencer by a confederate to the experimenter. This incentive was communicated as an intervention part way through the experiment, so that baseline data could be collected before the intervention, and it corresponds to the outside influence shown in Figure 1. Ground truth was collected to differentiate between false and authentic information for the experiments. Participants were tasked to collaborate on problem-solving assignments, and communicated with each other using message board, instant messages, and/or chat. Communication data in both experiments were collected over five (5) consecutive days.

4.4 Manipulation checks

We conducted two types of manipulation checks to ensure that participants perceive, comprehend, and react as expected to the portion of the manipulations. The first manipulation check was conducted to make certain the characteristics of social influencers. The second manipulation check was conducted to confirm the effectiveness of the incentive, and to minimize the confounding effects of the experiment design.

Different types of influence—*active* vs. *passive*—in social influencers' behavior (Caddell, 2004; Kimmel, 1998; Levine, 2014) could influence participants' behavior, perception and observation. The study divides social influences based on different influences style. That is, *active* and *passive* influences, characterized as different manipulation styles, were performed by social influencers yielding discrete deceptive acts/behaviors. Specifically, an *active* deceptive act refers to active steps taken to sabotage their group performance, intentionally spreading disinformation by submitting incorrect answers in violation of group's collective consensus and response. A *passive* deceptive act refers to silence, and/or a failure to act—concealment—by not submitting the group's collective response. That is, one is an act of commission, and the other an act of omission.

Each experiment lasts for five (5) consecutive days. In the first manipulation check, the consistency of manipulation styles across different experiments is ensured. Baseline data (i.e., data from Day 1 to Day 2) were collected to compare and determine consistency between the treatment and controlled groups. The study confirms that manipulation styles—groups of *active* influences compared with groups of *passive* influences—before the manipulation of social influences were identical and consistent.

In the second manipulation check, the data quality for intergroup comparison is ensured to confirm the effectiveness of the incentive, and to minimize the confounding effects of the experiment design. The study confirms that data—collected across treatment and controlled over *active* and *passive* influences—before the manipulation of social influences were identical and consistent. We found no significant differences in ways people communicated before the introduction of the incentive between two online platforms and across two institutions. The treatment data compared with controlled data includes data only from Day 3 through Day 5. When communication patterns observed after the incentive were compared to those captured before the incentive (intergroup comparison), differences in the treatment groups' reaction to the incentive were captured in terms of word count and affect process across active vs. passive influences. The fact that the communication difference persists despite different types of influences and across different online platforms validates the design of the experimentation.

4.5 Data cleaning and processing

Raw datasets of participants' conversations and message exchanges were collected, archived, and cleaned prior to analysis. In the first experiment, one participant's data was excluded from analysis because the participant did not complete the entire study, leaving the final dataset to consist of data from 25 participants. In the second experiment, all participant data was included in the analysis, so the final dataset consisted of data from 27 participants.

Collected data was processed using the Linguistic Inquiry and Word Count (LIWC) (Pennebaker, Chung, Ireland, Gonzales, & Booth, 2007). LIWC is a computerized text analyzer for computational linguistics analysis. Newman et al. (2003) adopted the LIWC to investigate linguistic style and distinguish between true and false stories, correctly classifying 67% of deceptive text, which was far better than the 52% accuracy from untrained human judges (p. 671). The use of LIWC to identify the psychological aspects of words used was also validated in Tausczik and Pennebaker's (2010) study. LIWC parses out words from text based on psychological constructs (Newman et al., 2003; Pennebaker & King, 1999;

Pennebaker et al., 2003). LIWC incorporates a dictionary that categorizes words across multiple dimensions of language use. The operating principle of LIWC is that, by analyzing a body of text on a word-by-word basis, with each word corresponding to one or more of the LIWC dimensions and categories, the text itself can reveal psychological processes. The output of LIWC analysis reflects the percentage of words from within the LIWC dictionary that appear in a given body of text against an overall word count. In other words, by taking the LIWC word count as a percentage of overall word count, the total counts of categorized words were normalized by the total length of the text messages exchanged per each experimental group communication.

The total word count (WC) between the treatment groups and control groups across all group communications became the baseline dataset for both experiments. Overall, the cleaned active influence dataset consists of a total count of 20,452 words in 9,682 total lines of chat. The cleaned passive influence dataset consists of a total count of 13,086 words in 9,477 total lines of chat.

Analysis focused on participants' use of words as cues depicting *cognitive processes* (CP) and *affective processes* (AF), in both the treatment and control groups (**Table 1**). The unit of analysis in both experiments was the words used (i.e., words that correspond to LIWC categories of interest), at the individual level, per group, per time period. From the raw data, we normalized the dataset for empirical investigation, and derived group-level data from communication between influencers and their respective group actors while collaborating to solve the puzzles during game play interaction. At the end of day two, the influencers were subtly incentivized to misbehave. Thereafter, the datasets for both experiments were divided into two distinct timeframes: data collected during days one and two (pre-incentive) and data collected during days three though five (post-incentive).

Table 1. Coding schema extracted from LIWC categories.

LIWC CATEGORIES	CODING SCHEMA	EXAMPLES	# of WORDS in CATEGORIES
Affective Process	affect	happy, cried, abandon	915
Positive Emotion	posemo	love, nice, sweet	406
Negative Emotion	negemo	hurt, ugly, nasty	499
Anxiety	anx	worried, fearful, nervous	91
Anger	anger	hate, kill, annoyed	184
Sadness	sad	crying, grief, sad	101
Cognitive Process	cogmech	cause, know, ought	730
Insight	insight	think, know, consider	195
Causation	cause	because, effect, hence	108
Discrepancy	discrep	should, would, could	76
Certainty	certain	always, never	83
Inhibition	inhib	block, constrain, stop	111
Inclusive	incl	and, with, include	18
Exclusive	excl	but, without, exclude	17

Descriptive statistics for outcomes of interest are summarized in **Table 2**. Specifically, mean, standard deviation, and range of total word count for both control and treatment groups in the active influence dataset were greater than those in the passive influence dataset. In terms of cognitive process-related words used, the passive influence dataset showed larger means than those in the active influence dataset for both groups, whereas standard deviations and ranges appeared to be similar in both experiments. Regarding affect-related word use, the mean for both groups was very close in both the active and passive datasets, and there was only a slight difference in standard deviation between these two experiments.

Table 2. Descriptive statistics of total word count, word use related to cognitive process, and word use related to affect.

Outcome	Group membership	Experiment 1 (active)			Experiment 2 (passive)		
		Mean	Standard deviation	Range	Mean	Standard deviation	Range
WC	Control	367.23	246.46	928.00	186.41	141.15	710.00
	Treatment	315.47	355.12	1661.00	186.94	119.74	447.00
CP	Control	12.34	4.18	25.65	17.07	4.76	29.55
	Treatment	13.68	5.20	23.00	14.64	3.86	20.56
AF	Control	6.10	2.34	12.50	6.78	3.18	17.86
	Treatment	5.69	2.99	14.08	6.70	3.16	16.39

Note. WC: the overall total word count; CP: word use related to cognitive process; AF: word use related to affective process.

We found several zero values within the extracted language-action cues as variables. These zero values may be a result of the sluggishness of the Internet speed, which may also explain why some participants did not make much conversation during the interaction. Even with small sample sizes, our experiments still indicated consistency and demonstrated significant mean differences in terms of total word count and affect-related word use.

5 MULTILEVEL MODELING OF DATA ANALYSIS

We speculate that patterns of group interaction may change after the introduction of a social influencer (intragroup comparison), and further assume that groups with a deceptive influencer may react differently when compared to groups without a deceptive influencer (intergroup comparison). In testing null hypotheses, we compare how groups react to a possible deceptive social influencer in both intergroup and intragroup comparison. Intergroup comparison refers to "between" group comparison where the results of the treatment groups (with a deceptive influencer) are compared with the results of the control groups (without a deceptive influencer). By contrast, intragroup comparison refers to "within" group

comparison where the results before a social influencer becomes deceptive are compared with the treatment results after a social influencer becomes deceptive.

Multilevel modeling is an essential approach to analyzing the hypotheses, because the data structure across the two experiments (active vs. passive influence) is nested. Specifically, repeated measures over five consecutive days are nested within individuals, which are then nested within each group (Raudenbush & Bryk, 2002). The data structure has three levels. However, for intergroup comparison, we used two-level linear mixed models to account for dependency of group members at the individual level due to relatively small numbers of groups (McNeish & Wentzel, 2017). Moreover, for intragroup comparison, we compare the means of the three outcome variables within the treatment groups during pre-incentive and post-incentive states using averages of repeated measures outcome scores at each of the two states using regressions and two-level linear mixed models. The outcome variable—word count—is measured on the time-level (i.e., each day of the experiments). The group membership *G* (0: control; 1: treatment) is measured at the participant level.

Regarding the between-groups (intergroup) comparison, we formulate the following model with μ_i representing the mean of an outcome variable for participant *i*:

$$\mu_i = d_0 + d_{btn}G_i + \alpha_1 D_1 + \cdots \alpha_{K-1}D_{K-1} + \varepsilon_i, \quad (1)$$

where K is the number of groups. D_k $(k = 1, \ldots, K-1)$ are a set of dummy variables ($D_k = 1$ if the participant is in group k and 0 otherwise) that represent the K groups. Therefore, d_{btn} is the average difference of an outcome variable between control and treatment groups across individuals controlling for groups. Equation 1 is called a level-2 model in the multilevel modeling framework in contrast to the model about repeated measures at level 1, and is used for each outcome variable in our study.

Regarding within-group (intragroup) comparison, we examined and compared whether the outcome was higher after *active* incentive is implemented, or lower after *passive* influence was implemented "within" treatment groups only. Our dataset contains two treatment groups for *active* influence, and three treatment groups for *passive* influence. With two treatment groups for active influence, we used a regression model controlling for group membership. By contrast, with three treatment groups, we used a multilevel model, and set the change scores of individuals at level 1 and groups at level 2. We began the analysis by obtaining the changes of means for *total word count*, *cognitive processes*, and *affective processes* from pre-incentive to post-incentive phrases. These change scores were used as the dependent variables. Then, we compared the

outcome for active (using regression analysis) versus passive influence (using multilevel modeling).

SAS Proc Mixed was used for multilevel analyses. Below, we discuss the analysis results regarding the three sets of hypotheses.

5.1 Hypothesis 1: Expressiveness

To address the question of whether the treatment groups use more or fewer words in total than the control groups after incentive—i.e., active or passive, respectively, we use the following model.

$$WC_{ti} = \mu_i + e_{ti}. \quad (2)$$

Here, $t = 3, 4, 5$ represent days for post-incentive measures; $i = 1, 2, \ldots, N$ individuals, and WC_{ti} is the overall word count measured at t for individual i. Equation 2 is called the level-1 model. For between-groups (intergroup) comparison, we found word count was not statistically significantly higher in treatment groups after incentive was introduced in the active influence dataset ($d_{btn} = 23.00$, p > .10), or it was not statistically lower in treatment groups with passive influence dataset ($d_{btn} = 4.88$, p > .10) (**Table 3**).

Table 3. Results from multilevel data analysis.

Outcome	Active Influence Experiment		Passive Influence Experiment	
	Between-Groups Difference	Within-Group Difference (for 2 groups)	Between-Groups Difference	Within-Group Difference (for 3 groups)
WC	23.00	135.46**	4.88	-41.94
CP	2.66	.19	-6.00**	-.75
AF	.99	*.62/-1.84***	-.93	1.84**

Note. **: p< 0.05; *: p< 0.10.

WC: the overall total word count; CP: word use related to cognitive process; AF: word use related to affective process. Between-Groups Difference: mean difference between control and treatment groups after incentive was introduced; Within-Group Difference: mean difference before and after an incentive has been accepted by a deceptive influencer within the treatment group.

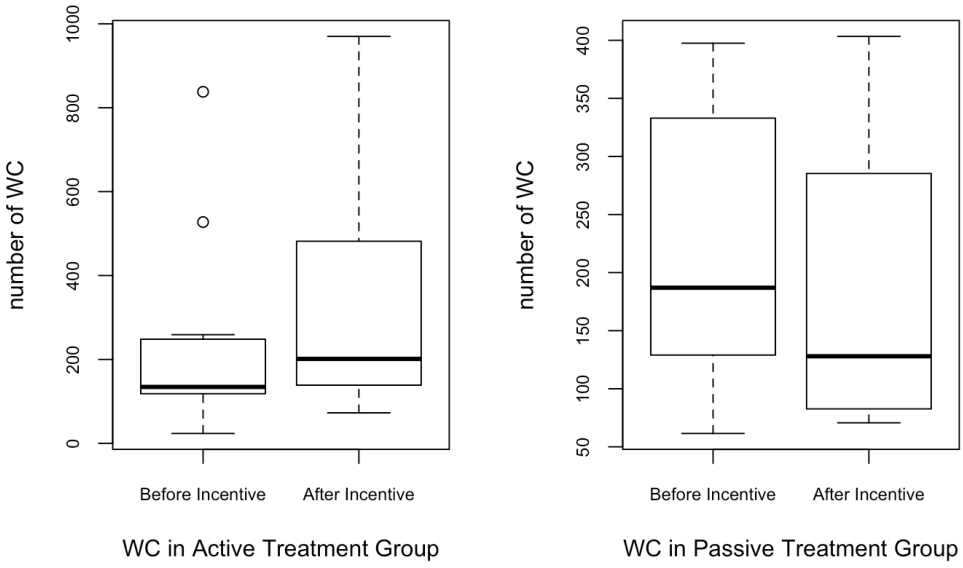

Figure 3. Differences of word counts within treatment groups (active vs. passive).

However, regarding within-group (intragroup) comparison, according to *Figure 3* boxplots, groups that include a deceptive influencer displays higher total word count after an influencer has initiated an active process of deception yet lower total word count after an influencer has initiated a passive process of deception. Our statistical test results show that word count significantly increased after *active* acts of spreading disinformation by 135.46 ($p < .05$) but did not significantly decrease after *passive* acts of spreading disinformation (**Table 3**). Therefore, hypothesis H1 is supported. Communication within groups that include a deceptive influencer—actively spreading disinformation—displays higher expressiveness, but when the act of deception is concealed and passive, the group displays lower expressiveness.

5.2 Hypothesis 2: Cognitive processing

To address the question of whether communication within the treatment groups will reflect more or fewer words relating to cognitive process than in the control groups after incentive—i.e., *active* or *passive*, respectively, we can use the following model.

$$CP_{ti} = \mu_i + e_{ti}. \tag{3}$$

Here, $t = 3, 4, 5$ represent days for post-incentive measures; $i = 1, 2, \ldots, N$ individuals, and CP_{ti} is cognitive process measured at t for

individual i. For between-groups (intergroup) comparison, we compared the average words relating to cognitive process between control and treatment groups during phases before and after incentive was implemented. Results showed that the use of cognitive process-related words was not statistically significantly higher for the treatment groups after incentive of active influence ($d_{btn} = 2.66,\ p > .10$) was introduced (**Table 3**). For passive influence data, by contrast, true to our expectation, the use of *cognitive* process-related word significantly decreased after incentive of passive influence ($d_{btn} = -6.00,\ p < .05$) was introduced (**Table 3**).

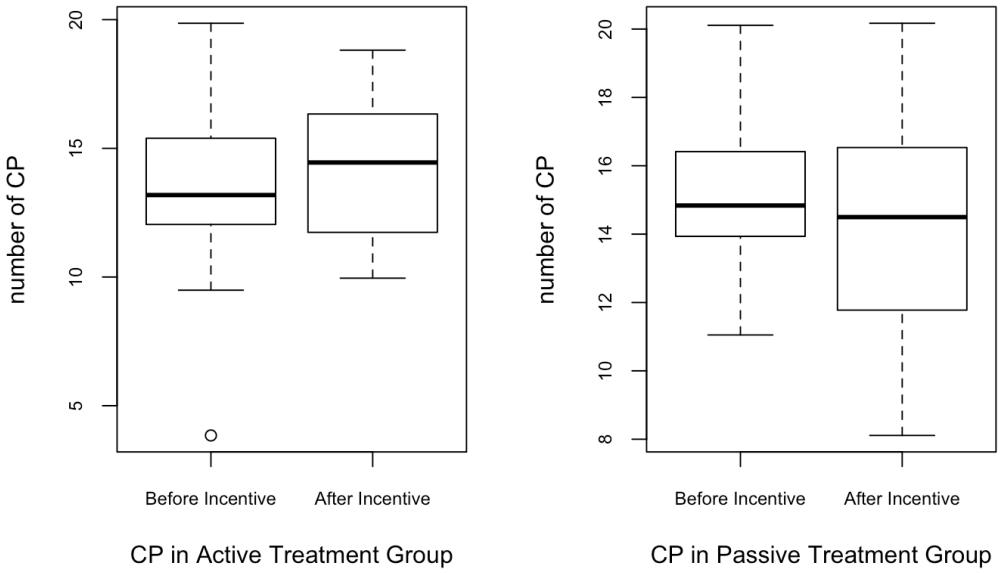

Figure 4. Differences of cognitive process within treatment groups (active vs. passive).

Regarding the within-group (intragroup) comparison, according to *Figure 4* boxplots, groups with a deceptive influencer display a higher mean value in cognitive process after an influencer has initiated an active process of spreading disinformation, but a lower mean value in cognitive process after an influencer has initiated a passive process of spreading disinformation. We compared the use of words related to cognitive processes for the *active* and *passive* influence datasets. For the *active* influence dataset, the use of cognitive process-related words did not show a statistically significant increase; likewise, for the *passive* influence dataset, the use of cognitive process-related words also did not show a statistically significant decrease (**Table 3**). Although groups that include a deceptive influencer do not display a noticeable difference—either higher or lower—in cognitive process after this influencer has initiated a process of deception as a result of the within-group (intragroup) comparison, the hypothesis H2 is

nonetheless supported by the between-groups (intergroup) comparison. Communication within groups that include a deceptive influencer—actively spreading disinformation—displays higher cognitive process, but when the act of deception is concealed and passive, the group displays lower cognitive process.

5.3 Hypothesis 3: Affective processing

To address the question of whether the treatment groups use less or more affect-related words than the control groups after incentive—i.e., active or passive, respectively, we can use the following model.

$$AF_{ti} = \mu_i + e_{ti}, \tag{4}$$

Here, $t = 3, 4, 5$ represent days for post-incentive measures; $i = 1, 2, \ldots, N$ individuals, and AF_{ti} is affect measured at t for individual i. For the active influence data, we found that affect scores on average were not statistically lower for the treatment groups after incentive ($d_{btn} = .99$, $p > .10$). For the passive influence data, the means of affect-related word usage was not statistically higher for the treatment groups after an incentive ($d_{btn} = -.93$, $p > .10$) was introduced (**Table 3**).

Regarding the within-group comparison, according to *Figure 5* boxplots, groups that include a deceptive influencer display a lower mean value in collective affective processing after an influencer has initiated an active process of spreading disinformation—yet a higher mean value in affective processing after an influencer has initiated a passive process of spreading disinformation. We compared the use of words related to affect processes for both the *active* and *passive* influence datasets for within-group (intragroup) comparison. For the active influence data, affective-process related words showed different patterns in the two treatment groups; one group did not show significantly lower affect-related words whereas the other group showed statistically significantly lower affect-related words ($d_{btn} = -1.84$, $p < .05$). For the passive influence data, there was a statistically significant increase in affective process-related words after incentive ($d_{btn} = 1.84$, $p < .05$). While there was no sufficient evidence to support the between-groups (intergroup) comparison, we assert that groups including a deceptive influencer using a passive influence strategy do, in fact, display higher affective process after this influencer has initiated a process of deception. For the active influence dataset, one of the two groups showed statistically significantly lower affective-process-related word usage. Thus, hypothesis H3 is support. Communication within groups that include a deceptive influencer—actively spreading disinformation—displays lower affective process, but when the act of

deception is concealed and passive, the group displays higher affective process.

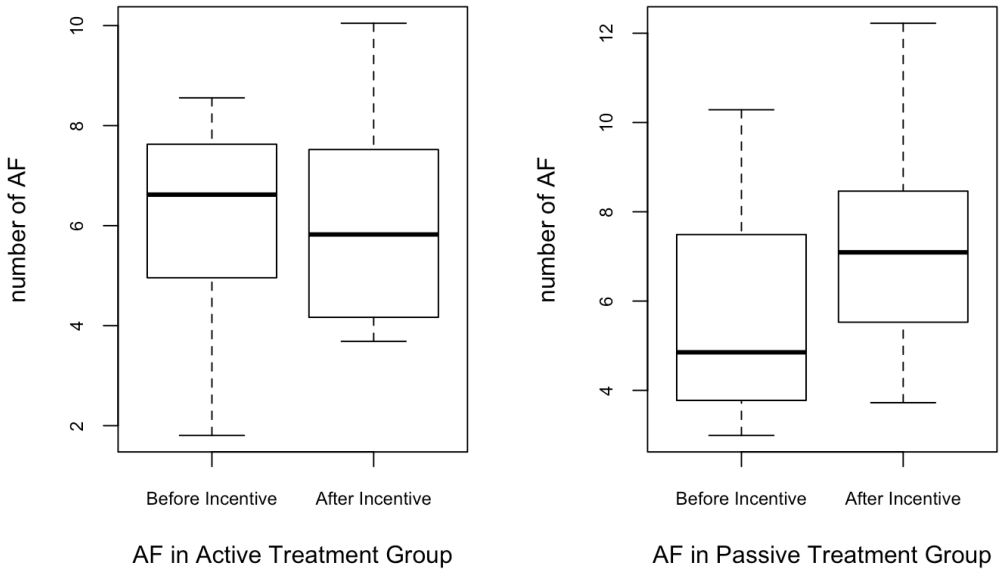

Figure 5. Differences of affective process within treatment groups (active vs. passive).

In summary, all three research hypotheses are supported. Communication within groups will exhibit higher *expressiveness* (H1) and lower *affective process* (H3) after a deceptive influencer has initiated a process of spreading disinformation. These hypotheses are supported by the intragroup comparison approach. Communication within groups that include a deceptive influencer actively spreading disinformation will exhibit higher *cognitive process* (H2) when compared with groups that do not have a deceptive influencer (i.e., no influence of deception). This hypothesis is supported by the intergroup comparison approach.

6 DISCUSSION AND IMPLICATIONS

This study addresses group communication affected by a social influencer with deceptive intent. Rather than collecting perception data using survey instruments, this study conducts experiments with two types of deceptive influence; *active* and *passive* to collect objective measures of language-action cues in group dynamics. Repeated measures were used to distinguish the difference between collective processes before and after an intervention that triggered the deceptive behavior. Specifically, communication and interactions between groups with a deceptive influencer were compared to groups without a deceptive influencer. The groups' collective reaction to situations before and after an influencer became deceptive was further

compared. In this section, we discuss the theoretical and practical implications of our findings. Moreover, the implications of both intergroup comparison (i.e., collective language-action cues change between control and treatment groups) and intragroup comparison (i.e., collective language-action cues change before and after treatment) are further elaborated to understand the dynamics of collective sensing.

6.1 Theoretical implications

This study supports our supposition of using collective sensing to detect deceptive influences in a group. The results suggest that an outside influence that affects a single individual—a social influencer—can cause a perturbation in the collective attributes of the team. Furthermore, this study also demonstrates the computational feasibility of detecting a shift of deceptive intent in a collective context, even though the deceptive act may not be detectable by any participating individual on the team.

Notably, the closure of the study is reached through analysis of the stigmergic traces—extracted language-action cues—of a collective engaged in teamwork. As the deceptive influencer's communicative intent was hidden in the threads of communication (intergroup comparison), no individuals picked up on this deceptive influencer's behavioral changes after they were incentivized to deceive. Even though no humans consciously recognized the deception, the collectives' behavioral changes in reaction to deception were computationally noticeable. Groups with a deceptive influencer used significantly fewer words associated with cognitive processes than groups without a deceptive influencer (from the intergroup perspective). Groups with a deceptive influencer showed a significant difference in total word count—as well as words reflecting affective processes—once an influencer had been active in the act of spreading disinformation (from the intragroup perspective).

It is noteworthy that collective sensing and associated processing of the results could detect changes in an influencer's communication patterns without participants being given any cues as to the possibility of deception by researchers, thus eliminating the possibilities of participants' social desirability bias (Krumpal, 2013), truth bias (Street & Masip, 2015) or halo effect susceptibility (Cooper, 1981). That is, participants were not given the knowledge on whether the influencer was deceptive, thus the individual and collective reactions to the changes in the influencer's behavior were considered native and intuitive. We conjecture that the changes in group behavior that caused changes in collective attributes are the results of an accumulation of subconscious reactions to subtle changes in the social influencer's behavior. Members of the group form a human sensor network that engages in collective sensing that detects the subtle changes in

communication triggered by a deceptive influencer. This study empirically demonstrates the efficacy of observing changes in group interactions/ dynamics as a valuable means of detecting a social influencer's deceptive act of spreading disinformation, and is in contrast to the statistically insignificant results from direct observation of a deceptive influencer's behavior.

6.1.1 Intergroup comparison

In intergroup comparison, our findings suggest that no difference in *expressiveness*, *cognitive* or *affective processes* is found between control and treatment groups before the introduction of any incentive. These findings support hypothesis H2 with statistical significance and highlights the validity of our research design (**Table 3**). Interestingly, when a deceptive influencer concealed the act of deception in a *passive* influence (i.e., the deceptive influencer did nothing to facilitate the group's collective goal of winning), the group tends to reflect less cognitive load. This *passive* deception possibly explains how control and treatment groups differ in the use of cognitive process words. That is, the influencer was not actively concealing his/her deceptive intent, but simply omitting information. Thus, a statistically significant difference was observed between the treatment groups and the control groups. We note a modest increase in the number of cognitive process words used by the control groups versus the treatment groups. This finding implies that a deceptive influencer may withdraw and hide his/her intent to purposely avoid the possibility of triggering the group's cognitive thinking. However, when the influencers are *actively* deceptive (actively facilitating disruption of the group' collective goal), they are successful at concealing deceptive intent, and no statistically significant difference was observed between treatment groups and the control groups (even after removing an outlier who seemed to remain silent or talked little before and after incentive. See left boxplot in *Figure 4*). *Figure 4* illustrates that the passive deception style discouraged group dynamics, which showed a statistically significant less cognitive word count for treatment groups than for control groups. The study objectively examines group language-action cues to prove that language-action cues can be discernibly different in a group that includes an influencer with deceptive intent

6.1.2 Intragroup comparison

In intragroup comparison, our findings that support the hypotheses H1 and H3, are even more intriguing. These hypotheses provide insight into how a group's collective language-action cues change when a social influencer has been compromised and becomes deceptive. Our results support the hypotheses that treatment groups display modified patterns in

communication after the influencer had been compromised. For example, the results show that the treatment groups used a different number of words relating to affective processes after the group influencer was compromised. Thus, our study affirms the overarching research question, showing with statistical significance that group communication can be expected to change once a compromised influencer demonstrates an intent to deceive.

We speculate that the distinction between *active* and *passive* deception may provide additional insight into patterns of group dynamics that include a deceptive influencer. That is, a deceptive influencer will seek to distance him- or herself from the group, either by using communication modes that provide the group with fewer cues as to their influencer's behavioral intent, or by using a linguistic style that leaves little room for questioning or additional details. We can anticipate that a change in communication will likely result in an alteration to the established interaction dynamic—particularly because the content and quality of the message will have been changed in an effort to disguise the influencer's deceptive intent. Further, we surmise that groups can notice and sense changes in an influencer's communication style from changes in overall *expressiveness* (i.e., word counts), as well as *cognitive* and *affective*-based language-action cues.

The study further illustrates how *active* versus *passive* deception may be expected to manifest in group communication. Considering overall *expressiveness* as illustrated in the intragroup comparison, we note that the groups with a deceptive influencer actively spreading disinformation (active treatment dataset) showed an increase in total word count. Not only did the change in group dynamic raise the group members' collective suspicion, but these suspicions were further fueled as group members became aware of various telltale signs from the action(s) the influencer took to sabotage the group (i.e., a clear instance of *active* deceptive intent being acted out). By comparison, groups with a social influencer *passively* spreading disinformation (passive treatment dataset) showed a decrease in the amount of communication. We conjecture that, while respective groups may have likewise observed a change in group interaction, there was no overt action for them to notice and react to, and as a result less group interaction overall. With respect to *cognitive* process as addressed by the intragroup comparison, there was a small change in use of words relating to *cognitive* processes before vs. after the incentive treatment (incentive) was introduced. The change in the influencer's deceptive intent was unquestionably sensed by collective, but not by the individual members. The increase of groups' cognitive load as a result of the influencer's adopted reticence—was sensed (even if not understood) by the group members, leading to a shift in the group dynamic reflecting *affect*-based reactions. This

leads to the intragroup comparison of hypothesis H3, that while group members may have communicated with one another in an attempt to gain insight or a sense of certainty, ultimately the emotional (affective) component of the dialog presided over the rational/ cognitive component, to infect the group with confusion, concern and frustration. With respect to the use of affective-process related words, our datasets show a statistically significant increase in the use of words related to affective processes after an influencer was passively spreading disinformation.

This study advances disinformation research by contemplating the efficacy of pattern recognition in collective sensing in group communication, and measuring collective sensing using a multilayer modeling approach that analyzes the groups' collective language-action cues as forms of stigmergic signals. The novelty of this research lies in the advancement of measuring and understanding groups' collective reactions (in terms of language-action cues) to a deceptive social influencer. The research design is novel in that it provides for objective patterns of group interaction in response to the deceptive social influencers. While it is always challenging to detect social influencer's deceptive intent, this research demonstrates a possibility for an early warning system that can identify potential disinformation based on collective sensing by objectively analyzing and comparing groups' interactive information behavior.

6.2 Practical implications

Disinformation is a complex social problem that can manifest in any group, organization, or community context, especially given the increasing prevalence of online communications in our daily lives. This problem of disinformation is further complicated when all observed human interaction is entirely online. Unlike social bots, social influencers can masquerade their deceptive intent easier in online communication. Although fact-checking[3,4] is useful, it cannot thwart the spread of disinformation, and does not minimize the impact of disinformation on society as illustrated by the Capitol riots (Dozier & Bergengruen, 2021; Holland et al., 2021) and the false claims of coronavirus (McCarthy, 2020; Milman, 2020). Past research has suggested that analyzing a deceptive social influencer's language-action cues in group context does not offer statistical significance because the deceptive influencer's behavior involves only one datapoint (Ho et al., 2017). Moreover, surveillance or monitoring an individual's online communication triggers privacy concerns, which does not warrant a reliable measure. In contrast, the present study empirically supports our

[3] IFCN Fact-Checking: https://ifcncodeofprinciples.poynter.org/know-more/the-code-and-the-platforms
[4] Media Bias/Face Check (MBFC): https://mediabiasfactcheck.com/methodology

supposition for using collective sensing to inform the early identification of disinformation. That is, just as other software tools have been built that autonomously operate (Seidel et al., 2020), tools could be specifically constructed to automatically pick up group-level signals of disinformation

6.3 Generalizability

One issue to be considered is generalizability. Deception occurs in a variety of contexts, and it is not at all justifiable to say that deceit in a card game is the same as deceit in stock market trading. The situation described here can arguably be generalized to corporate settings in which much communication is through enterprise social media. That is, when communication in small teams is mediated through textual communication. Many of the treatment conditions—incentivized deceptive behavior—in the experiment are similar to the organizational communication mediated by technology (i.e., enterprise social media). Other situations—for example, broadcasts to thousands of followers on twitter—are quite different from the experimental conditions, in that there is not a small nucleus of people paying attention to accomplishing a common goal, but instead potentially thousands of people with fractured attention. In such cases, the subtlety of perception seen in the experimental situation is most likely swamped by other effects related to valence of language and bandwagon effects.

Another kind of generalizability can come from considering different sorts of situations. For example, the present experiment design only posited at most one deceptive person in the group. There was no real way for any but the leader to effectively sway the group by revealing, suppressing, or distorting information. While it is possible that other people in the virtual team could also engage in deceptive behavior, the incentives were against that behavior, and the investigators did not see it. In real world situations, deception can involve multiple actors in a coordinated plan and effort e.g., a denial and deception (D&D) campaign. Russian Deception Warfare in Crimea and Ukraine is a coordinated effort of Soviet *dezinformatsiya* campaign to influence the public opinion (Bouwmeester, 2017). At the other end of the spectrum might be experiments that effectively incentivized cooperative or selfless behavior on behalf of the deceptive social influencer. It would be interesting to contemplate if collective sensing might pick up on hidden charity or helpfulness, and if that changed group outcomes.

6.4 Limitations

This study has a small sample size, which may increase the chance of measurement errors. However, based on research performed by Maas and Hox (2005), it is sufficient to have 30 individuals, each with 5 repeated

measures, to obtain accurate fixed effect estimates and statistical inferences. In fact, the actual datasets being processed and analyzed are language-action cues with the total word count of 20,452 words in 9,682 total lines of chat (active influence data), and the total word count of 13,086 words in 9,477 total lines of chat (passive influence data). Comparatively speaking, the total word count in the passive influence dataset is much smaller than the active influence dataset. However, even with a smaller dataset, the cognitive and affective processing in the passive influence from intergroup and intragroup comparisons were still found to be statistically significant. The highly controlled and structured research design allowed us to collect, process and compare collective language-action cues as repeated measures representative of the objective perspectives in group interaction.

Although the study is based on three variables: *expressiveness*, *cognitive processing*, and *affective processing*, the approach of deriving the variables was rigorous. Both cognitive and affective processes were derived from hundreds of indictors in LIWC (as illustrated in **Table 1**). However, the total word count is considered a single variable. With intragroup comparison, even though cognitive processing was not statistically significant in both datasets, we still observed a statistically significant difference after the financial incentive was accepted (*cognitive* processing increased in active influence, but decreased in passive influence). As hundreds of indicators are included in this affective processing, indicators may include variables that contradict one another (such as positive vs. negative emotions), and thus introduce measurement errors that could compromise the weights of the parsed words until we have a bigger sample size.

By design, there was no difference in the manipulation instructions given to the social influencers in the treatment groups across the two institutions. The differences between *active* and *passive* deception as manifested by the social influencers could be a confounding effect as a result of the intrinsic differences in the confederates' disposition/personalities, and could also reflect the external differences of the online learning environment or the cultural differences between two institutions.

The study does not examine scenarios in which a designated "ethical" social influencer of a treatment group might refuse to influence the interacting group unethically. Future iterations of the research could provide insight by comparing individuals who do not deceive the group, to see if the ethical choice made in reaction to an attempt to compromise also introduces ripple effects into collective processing. Our analysis remains focused on collective sensing in terms of how group members collectively react and respond in situations where an influencer becomes deceptive (i.e., including situations of both "*active*" or "*passive*" deception), and not the types of activities a social influencer would engage in.

To the best of our knowledge there are no theoretical standards or statistical references for collective sensing in this context. That is; there is no base or standard for judgment when choosing variables or judging effectiveness. Future research may also benefit from a revised design that manipulates incentives for *active* versus *passive* deception, and designs that vary within one experiment the proportion of commissive and omissive deceptive acts. Such studies might be considered steps toward building machine learning systems that help recognize deception by paying attention to traces of collective behavior. Experiments like the ones described here, might, if scaled, provide a way to generate enough data for the training of robust classifiers. Once trained, such classifiers could be tested in non-game environments to if the learning transfers to the less structured and more fluid environments of persistent conversation in social media.

In the experiments presented, the platform that shared textual information was different in the different experimental locations. Moreover, the experiment took place over a long timeframe. Both of these differences mean there could be confounds: for example, world events might change the sentiment toward deception, or small differences in response time might encourage larger amounts of communication. While the platforms and settings shared many feature similarities, and we didn't see obvious behavioral differences in the way the platforms were used or the nature of participants behavior across experiments, it would be useful to better understand the extent to which differences in platforms can lead to differences in behavior, including deceitful behavior.

In sum, there are several ways that future researchers might build on both the findings and the experimental design discussed in this paper. Simultaneous studies with more conditions and standardized technologies might seek to better understand how collective sensing might pick up on particular forms of behavior. Here, the emphasis of the present study has been on collective sensing over deceptive social influences, but other kinds of antisocial or eusocial behavior might also be studied.

7 CONCLUSION

Disinformation and fake news as a type of computer-mediated deception generated with a subtle deceptive intent can harm the stability of society. The impact of disinformation creates a lack of trust across the society, and this spreads quickly on social media. In this age where communication is in every fiber, and influences the stability, of the society, disinformation prevails and disrupts the core values of ethics and trust in the society (McCarthy, 2020; Milman, 2020). When a social influencer becomes deceptive, the deceptive behavior can influence a group's communication

and performance. For example, the supposition of collective sensing presented here offers a potential explanation of the 2021 U.S. Presidential Election in the prevalence of disinformation: citizens engaged in collective sensing recognized disinformation, even if many individuals could not.

The study models collective sensing based on the core components of stigmergy including human sensors (agents), language-action cues as stigmergic signals (signs), and social media (environments) and shows collective sensing can detect disinformation. The language-action cues as repeated measures observed from interacting individuals during group interaction can indicate when deception is present, and especially when the interaction within a group provides context for groups to unknowingly become a network of sensors. Collective sensing can discern the trustfulness of information, and subtle changes in a social influencer's intent. The research delineates a method for computationally identifying an influencer's deceptive intent through analyzing collective language-action cues nested within and between group interaction, as an efficacious means of outing a deceptive influencer. More generally, this study suggests that collective sensing may integrate information and communication behavioral patterns at the organizational level to recognize a subtle spread of disinformation.

FUNDING STATEMENT AND ACKNOWLEDGMENTS

The authors wish to thank the National Science Foundation EAGER grant #1347113 and #1347120, 09/01/13—08/31/15. The first author wishes to also thank the Florida State University Council for Research and Creativity Planning Grant #034138, 12/01/13—11/30/14. The authors acknowledge the game development, data collection and analysis efforts of Shashanka S. Timmarajus, Aravind Hariharan, Wenyi Li, and many research participants. The first author wishes to thank Conrad Metcalfe for his editing assistance.

REFERENCES

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. ***Journal of Economic Perspectives***, *31*(2), 211-236. doi:10.1257/jep.31.2.211

Benkler, Y., Faris, R., & Roberts, H. (2018). *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*: Oxford University Press.

Bennati, S. (2018). On the role of collective sensing and evolution in group formation. ***Swarm Intelligence***, *12*, 267-282. doi:10.1007/s11721-018-0156-y

Berdahl, A., Torney, C. J., Ioannou, C. C., Faria, J. J., & Couzin, I. D. (2013). Emergent sensing of complex environments by mobile animal groups. ***Science***, *339*(6119), 574-576.

Berger, J. (2016). *Contagious: Why things catch on* (1st ed.): Simon & Schuster.

Blaschke, T., Hay, G. J., Weng, Q., & Resch, B. (2011). Collective sensing: Integrating geospatial technologies to understand urban systems—An overview. ***Remote Sensing***, *3*(8), 1743-1776. doi:10.3390/rs3081743

Bodrunova, S. S., Blekanov, I., Smoliarova, A., & Litvinenko, A. (2019). Beyond left and right: Real-world political polarization in Twitter discussions on inter-ethnic conflicts. ***Media and Communication***, *7*(3), 119-132. doi:10.17645/mac.v7i3.1934

Bouwmeester, H. (2017). Lo and behold: Let the truth be told—Russian deception warfare in Crimea and Ukraine and the return of 'Maskirovka' and 'Reflexive Control Theory'. ***Netherlands Annual Review of Military Studies 2017***, 125-153. doi:10.1007/978-94-6265-189-0_8

Boxwell, R. (2020a, April 4, 2020). The blame game: The origins of Covid-19 and the anatomy of a fake news story. *South China Morning Post Magazine*. Retrieved from https://www.scmp.com/magazines/post-magazine/long-reads/article/3078417/how-chinas-fake-news-machine-rewriting-history

Boxwell, R. (2020b, April 4, 2020). How China's fake news machine is rewriting the history of Covid-19, even as the pandemic unfolds, Opinion. *Politico*. Retrieved from https://www.politico.com/news/magazine/2020/04/04/china-fake-news-coronavirus-164652

Brown, C. R., Greitzer, F. L., & Watkins, A. (2013). *Toward the development of a psycholinguistic-based measure of insider threat risk focusing on core word categories used in social media.* In Proceedings of the 2013 Americas Conference on Information Systems, Chicago, Illinois, 1-8.

Brown, C. R., Watkins, A., & Greitzer, F. L. (2013, January 7-10). *Predicting insider threat risks through linguistic analysis of electronic communication.* In Proceedings of the 2013 46th Hawaii International Conference on System Sciences, Wailea, Hawaii, 1849-1858. doi:10.1109/HICSS.2013.453

Buller, D. B., & Burgoon, J. K. (1994). Deception: Strategic and nonstrategic communication. ***Strategic interpersonal communication***, 191-223.

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

Burgoon, J. K., Blair, J. P., Qin, T., & Nunamaker, J. F. (2003). Detecting deception through linguistic analysis. ***Intelligence and Security Informatics***, *2665*, 91-101.

Burgoon, J. K., & Buller, D. B. (1994). Interpersonal deception: III. Effects of deceit on perceived communication and nonverbal behavior dynamics. ***Journal of Nonverbal Behavior***, *18*(2), 155-184. doi:10.1007/BF02170076

Burgoon, J. K., Buller, D. B., Dillman, L., & Walther, J. B. (1995). Interpersonal deception. IV. Effects of suspicion on perceived communication and nonverbal behavior dynamics. ***Human Communication Research***, *22*(2), 163-196.

Caddell, J. W. (2004). *Deception 101--Primer on deception*. (1-58487-180-6). Strategic Studies Institute: U.S. Army War College Retrieved from https://ssi.armywarcollege.edu/pubs/display.cfm?pubID=589

Chen, X., Sin, S.-C. J., Theng, Y.-L., & Lee, C. S. (2015). Why students share misinformation on social media: Motivation, gender and study-level differences. ***The Journal of Academic Librarianship***, *41*(5), 583-592. doi:10.1016/j.acalib.2015.07.003

Cooper, W. H. (1981). Ubiquitous halo. ***Psychological Bulletin***, *90*(2), 218-244. doi:10.1037/0033-2909.90.2.218

Crowston, K., O'sterlund, C. S., Howison, J., & Bolici, F. (2017). *Work features to support stigmergic coordination in distributed teams.* In Proceedings of the Academy of Management Annual Meeting Proceedings, Briarcliff Manor, NY 10510, 14409. doi:10.5465/AMBPP.2017.14409abstract

Dansereau, F., Graen, G., & Haga, W. J. (1975). A vertical dyad linkage approach to leadership within formal organizations: A longitudinal investigation of the role making process. ***Organizational Behavior and Human Performance***, *13*(1), 46-78. doi:10.1016/0030-5073(75)90005-7

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. ***Journal of Personality and Social Pyschology***, *70*(5), 979-995. doi:0022-3514/96

Dipple, A., Raymond, K., & Docherty, M. (2014). General theory of stigmergy: Modeling stigma semantics. ***Cognitive Systems Research***, *31-32*, 61-92. doi:10.1016/j.cogsys.2014.02.002

Dozier, K., & Bergengruen, V. (2021, January 6, 2021). Incited by the President, Pro-Trump rioters violently storm the Capitol. *TIME*. Retrieved from https://time.com/5926883/trump-supporters-storm-capitol/

Ekman, P., & Friesen, W. B. (1969). Nonverbal leakage and clues to deception. ***Psychiatry***, *32*, 88-106.

Ezeakunne, U., Ho, S. M., & Liu, X. (2020, October 18-21, 2020). *Sentiment and retweet analysis of user response for fake news detection.* In Proceedings of the Proceedings of the 2020 International Conference on Social Computing, Behavioral-Cultural Modeling & Prediction and Behavior Representation in Modeling and Simulation (SBP-

BRiMS'20), Washington D.C., 1-10 (Paper No. 37). Retrieved from http://sbp-brims.org/2020/proceedings/papers/working-papers/SBP-BRiMS_2020_paper_37.pdf

Fetzer, J. H. (2004a). Disinformation: The use of false information. ***Minds and Machines***, *14*(2), 231-240. doi:10.1023/B:MIND.0000021683.28604.5b

Fetzer, J. H. (2004b). Information: Does it have to be true? ***Minds and Machines***, *14*(2), 223-229. doi:10.1023/B:MIND.0000021682.61365.56

Garrett, R. K. (2017). The "echo chamber" distraction: Disinformation campaigns are the problem, not audience fragmentation. ***Journal of Applied Research in Memory and Cognition***, *6*(4), 370-376. doi:10.1016/j.jarmac.2017.09.011

George, J. F., Giordano, G., & Tilley, P. A. (2016). Website credibility and deceiver credibility: Expanding prominence-Interpretation Theory. ***Computers in Human Behavior***, *54*, 83-93. doi:10.1016/j.chb.2015.07.065

Gordon, D. M. (2014). The ecology of collective behavior. ***PLOS Biology***, *12*(3), e1001805. doi:10.1371/journal.pbio.1001805

Gordon, D. M. (2019). The ecology of collective behavior in ants. ***Annual review of entomology***, *64*, 35-50. doi:10.1146/annurev-ento-011118-111923

Grassé, P.-P. (1959). La reconstruction du nid et les coordinations interindividuelles chezBellicositermes natalensis etCubitermes sp. la théorie de la stigmergie: Essai d'interprétation du comportement des termites constructeurs. ***Insectes Sociaux***, *6*(1), 41-80. doi:10.1007/BF02223791

Greitzer, F. L., Kangas, L. J., Noonan, C. F., Brown, C. R., & Ferryman, T. (2013). Psychosocial modeling of insider threat risk based on behavioral and word use analysis. ***e-Service Journal***, *9*(1), 106-138. doi:10.2979/eservicej.9.1.106

Greitzer, F. L., Kangas, L. J., Noonan, C. F., Dalton, A. C., & Hohimer, R. E. (2012). *Identifying at-risk employees: Modeling psychosocial precursors of potential insider threats.* In Proceedings of the 2012 45th Hawaii International Conference on System Sciences, Maui, Hawaii, 2392-2401. doi:10.1109/HICSS.2012.309

Griffith, J. A., Connelly, S., & Thiel, c. E. (2011). Leader deception influences on leader-member exchange and suborginate organizational commitment. ***Journal of Leadership & Organizational Studies***, *18*(4), 508-521. doi:10.1177/1548051811403765

Hackman, J. R., & Vidmar, N. (1970). Effects of size and task type on group performance and member reactions. ***Sociometry***, *33*(1), 37-54. doi:10.2307/2786271

Hancock, J., Birnholtz, J., Bazarova, N., Guillory, J., Perlin, J., & Amos, B. (2009). *Butler lies: Awareness, deception and design.* In Proceedings of the CHI'09, Boston, MA.

Hancock, J., Curry, L. E., Goorha, S., & Woodworth, M. (2008). On lying and being lied to: A linguistic analysis of deception in computer-medicated communication. ***Discourse Process**, 45*(1), 1-23. doi:10.1080/01638530701739181

Hernon, P. (1995). Disinformation and misinformation through the Internet: Findings of an exploratory study. ***Government Information Quarterly**, 12*(2), 133-139. doi:10.1016/0740-624X(95)90052-7

Heylighen, F. (1999). Collective intelligence and its implementation on the Web: Algorithms to develop a collective mental map. ***Journal of Computational & Mathematical Organization Theory**, 5*(3), 253-280. doi:10.1023/A:1009690407292

Ho, S. M. (2009). *Behavioral anomaly detection: A socio-technical study of trustworthiness in virtual organizations.* (Ph.D. Information Systems). Syracuse University, Syracuse. Retrieved from http://libezproxy.syr.edu/login?url=http://proquest.umi.com/pqdweb?did=2112815091&sid=1&Fmt=2&clientId=3739&RQT=309&VName=PQD Available from ProQuest ProQuest database. (47)

Ho, S. M. (2019, January 8, 2019). *Leader member exchange: An interactive framework to uncover a deceptive insider as revealed by human sensors.* In Proceedings of the Proceedings of the 2019 52nd Hawaii International Conference on System Sciences (HICSS-52), Maui, Hawaii, 3212-3221. doi:hdl.handle.net/10125/59757

Ho, S. M., Fu, H., Timmarajus, S. S., Booth, C., Baeg, J. H., & Liu, M. (2015, June 4-6). *Insider threat: Language-action cues in group dynamics.* In Proceedings of the Proceedings of the 2015 ACM SIGMIS Computers and People Research (SIGMIS-CPR'15), Newport Beach, CA, 101-104. doi:10.1145/2751957.2751978

Ho, S. M., & Hancock, J. T. (2018, January 3-6). *Computer-mediated deception: Collective language-action cues as stigmergic signals for computational intelligence.* In Proceedings of the Proceedings of the 2018 51th Hawaii International Conference on System Sciences (HICSS-51), Big Island, Hawaii, 1671-1680. doi:hdl.handle.net/10125/50098

Ho, S. M., & Hancock, J. T. (2019). Context in a bottle: Language-action cues in spontaneous computer-mediated decepion. ***Computers in Human Behavior**, 91*, 33-41. doi:10.1016/j.chb.2018.09.008

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. doi:10.1002/asi.23849

Ho, S. M., Hancock, J. T., Booth, C., & Liu, X. (2016). Computer-mediated deception: Strategies revealed by language-action cues in spontaneous communication. ***Journal of Management Information Systems***, *33*(2), 393-420. doi:10.1080/07421222.2016.1205924

Ho, S. M., & Warkentin, M. (2017). Leader's dilemma game: An experimental design for cyber insider threat research. ***Information Systems Frontiers***, *19*(2), 377-396. doi:10.1007/s10796-015-9599-5

Holland, S., Mason, J., & Landay, J. (2021, January 6, 2021). Trump summoned supporters to "wild" protect, and told them to fight. They did. *Reuters*. Retrieved from https://www.reuters.com/article/us-usa-election-protests/trump-summoned-supporters-to-wild-protest-and-told-them-to-fight-they-did-idUSKBN29B24S

Hosmer, L. T. (1995). Trust: The connecting link between organizational theory and philosophical ethics. ***Academy of Management Review***, *20*(2), 379-403. Retrieved from http://www.jstor.org/stable/258851

Kahai, S. S., & Cooper, R. B. (2003). Exploring the core concepts of media richness theory: The impact of cue multiplicity and feedback immediacy on decision quality. ***Journal of Management Information Systems***, *20*(1), 263-299.

Kim, A., & Dennis, A. R. (2019). Says who? The effects of presentation format and source rating on fake news in social media. ***MIS Quarterly***, *43*(3), 1025-1039. doi:10.25300/MISQ/2019/15188

Kim, A., Moravec, P. L., & Dennis, A. R. (2019). Combating fake news on social media with source ratings: The effects of user and expert reputation ratings. ***Journal of Management Information Systems***, *36*(3), 931-968. doi:10.1080/07421222.2019.1628921

Kimmel, A. J. (1998). In defense of deception. ***American Psychologist***, *53*(7), 803-805. doi:10.1037/0003-066X.53.7.803

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: a literature review. ***Quality & Quantity: International Journal of Methodology***, *47*(4), 2025-2047. doi:10.1007/s11135-011-9640-9

Levine, T. R. (2014). Active deception detection. ***Policy Insights from the Behavioral and Brain Sciences***, *1*(1), 122-128. doi:10.1177/2372732214548863

Liden, R. C., Erdogan, B., Wayne, S. J., & Sparrowe, R. T. (2006). Leader-member exchange, differentiation, and task interdependence: Implications for individual and group performance. ***Journal of Organizational Behavior***, *27*(6), 723-746. doi:10.1002/job.409

Liden, R. C., & Graen, G. (1980). Generalizability of hte vertical dyad linkage model of leadership. ***Academy of Management Review***, *23*(3), 451-465. doi:10.2307/255511

Liden, R. C., Wayne, S. J., & Stilwell, D. (1993). A longitudinal study on the early development of leader-member exchanges. ***Journal of Applied Psychology***, *78*(4), 662-674. doi:10.1037/0021-9010.78.4.662

Liu, F., & Li, M. (2019). A game theory-based network rumor spreading model: based on game experiments. ***International Journal of Machine Learning and Cybernetics***, *10*(2019), 1449-1457. doi:10.1007/s13042-018-0826-5

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. ***Methodology: European Journal of Research Methods for the Behavioral and Social Sciences***, *1*(3), 86-92. doi:10.1027/1614-2241.1.3.86

Majchrzak, A., Rice, R. E., Malhotra, A., King, N., & Ba, S. (2000). Technology adaption: The case of a computer-supported inter-organizational virtual team. ***MIS Quarterly***, *24*(4), 569-600. doi:10.2307/3250948

Malone, T. W., Laubacher, R., & Dellarocas, C. (2010). The collective intelligence genome. ***MIT Sloan Management Review***, *51*(3), 21-31.

McCarthy, T. (2020, April 14, 2020). 'It will disappear': the disinformation Trump spread about the coronavirus-timeline, Assorted. *The Guardian*. Retrieved from https://www.theguardian.com/us-news/2020/apr/14/trump-coronavirus-alerts-disinformation-timeline

McNeish, D., & Wentzel, K. R. (2017). Accommodating small sample sizes in three-level models when the third level is incidental. ***Multivariate Behavioral Research***, *52*(2), 200-215. doi:10.1080/00273171.2016.1262236

Mehrabian, A. (1968). Methods & designs: Some referents and measures of nonverbal behavior. ***Behavior Research Methods & Instrumentation***, *1*(6), 203-207.

Milman, O. (2020, March 31, 2020). Seven of Donald Trump's most misleading coronavirus claims, Assorted. *The Guardian*. Retrieved from https://www.theguardian.com/us-news/2020/mar/28/trump-coronavirus-misleading-claims

Mocanu, D., Rossi, L., Zhang, Q., Karsai, M., & Quattrociocchi, W. (2015). Collective attention in the age of (mis)information. ***Computers in Human Behavior***, *51*(Part B), 1198-1204. doi:10.1016/j.chb.2015.01.024

Morrison, E. W., & Robinson, S. L. (1997). When employees feel betrayed: A model of how psychological contract violation develops. ***Academy of Management Review***, *22*(1), 226-256.

Negoita, B., Lapointe, L., & Rivard, S. (2018). Collective information system use: A typological theory. ***MIS Quarterly***, *42*(4), 1281-1301. doi:10.25300/MISQ/2018/13219

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.

Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2007). *The development and psychometric properties of LIWC2007*. Retrieved from http://www.liwc.net/LIWC2007LanguageManual.pdf

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. ***Journal of Personality and Social Pyschology***, *77*(6), 1296-1312. doi:10.1037/0022-3514.77.6.1296

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. ***Annual Review of Psychology***, *54*, 547-577. doi:10.1146/annurev.psych.54.101601.145041

Raafat, R. M., Chater, N., & Frith, C. (2009). Herding in humans. ***Trends in Cognitive Sciences***, *13*(10), 420-428. doi:10.1016/j.tics.2009.08.002

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd ed.). Thousand Oaks, California: Sage Publication.

Resch, B. (2013). People as sensors and collective sensing-contextual observations complementing geo-sensor network measurements. In J. M. Krisp (Ed.), *Progress in Location-Based Services* (pp. 391-406): Springer-Verlag Berlin Heidelberg.

Rezgui, A., & Crowston, K. (2018). *Stigmergic coordination in Wikipedia.* In Proceedings of the Proceedings of the 14th International Symposium on Open Collaboration, Paris, France, 1-12. doi:10.1145/323391.3233543

Robinson, S. L. (1996). Trust and breach of the psychological contract. ***Administrative Science Quarterly***, *41*(4), 574-599. doi:10.2307/2393868

Schultz, E. E. (2002). A framework for understanding and predicting insider attacks. ***Computers & Security***, *21*(6), 526-531.

Seidel, S., Berente, N., Lindberg, A., Lyytinen, K., Martinez, B., & Nickerson, J. V. (2020). Artificial intelligence and video game creation: A framework for the new logic of autonomous design. ***Journal of Digital Social Research***, *2*(3), 126-157. doi:10.33621/jdsr.v2i3.46

Simons, T. (2002). Behavioral integrity: The perceived alignment between managers' words and deeds as a research focus. ***Organization Science***, *13*(1), 18-35. doi:10.1287/orsc.13.1.18.543

Street, C. N. H., & Masip, J. (2015). The source of the truth bias: Heuristic processing? ***Scandinavian Journal of Psychology***, *56*, 254-263. doi:10.1111/sjop.12204

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. ***Journal of Language and Social Psychology***, *29*(1), 24-54. doi:10.1177/0261927X09351676

Taylor, P. J., Dando, C. J., Ormerod, T. C., Ball, L. J., Jenkins, M. C., Sandham, A., & Menacere, T. (2013). Detecting insider threats through language change. ***Law and Human Behavior***, *37*(4), 267-275. doi:10.1037/lhb0000032

Tendoc Jr., E. C., Lim, Z. W., & Ling, R. (2017). Defining "fake news." A typology of scholarly definitions. ***Digital Journalism***, *6*(2), 137-153. doi:10.1080/21670811.2017.1360143

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. ***Science***, *359*(6380), 1146-1151. doi:10.1126/science.aap9559

Wayne, S. J., Shore, L. M., & Liden, R. C. (1997). Perceived organizational support and leader-member exchange: A social exchange perspective. ***The Academy of Management Journal***, *40*(1), 82-111.

Wheelan, S. A. (2009). Group size, group development and group productivity. ***Small Group Research***, *40*(2), 247-262. doi:10.1177/1046496408328703

Woolley, A. W., Aggarwal, I., & Malone, T. W. (2015). Collective intelligence and group performance. ***Current Directions in Psychological Science***, *24*(6), 420-424. doi:10.1177/0963721415599543

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. ***Science***, *330*, 686-688. doi:10.1126/science.1193147

Yang, K., Ahn, C. R., Vuran, M. C., & Kim, H. (2017). Collective sensing of workers' gait patterns to identify fall hazards in construction. ***Cognitive Systems Research***, *82*, 166-178. doi:10.1016/j.autcon.2017.04.010

Zhou, L., Burgoon, J. K., Nunamaker Jr., J. F., & Twitchell, D. P. (2004). Automating linguistics-based cues for detecting deception in text-based asynchronous computer-mediated communication. ***Group Decision and Negotiation***, *13*(1), 81-106. doi:10.1023/B:GRUP.0000011944.62889.6f

Zhou, L., & Zhang, D. (2004, Jan. 5-8). *Can online behavior unveil a deceiver?* In Proceedings of the Proceedings of the 2004 Hawaii International Conference on System Sciences (HICSS-37), Hilton Waikoloa Village Big Island, Hawaii.