NORTHWESTERN UNIVERSITY

The Role of Partner Knowledge Inferences in Text Based Collaborative Workspaces

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Psychology

By

By Arielle Elliott

EVANSTON, ILLINOIS

September 2023

Abstract

In shared workspaces users have real-time access to others' actions, allowing greater access to potentially informative cues. However, there has been little exploration about what impact the ability to view each other's work in real time has on group members' perception of both each other and their understanding of the task itself. In a series of four experiments, I examine how visual access to a collaborative partner's real-time typing behaviors shapes the behaviors of others in the collaborative team—similarly to how paralinguistic cues shape face-to-face interactions. First, I test how task difficulty impacts typing in a written production task, and how viewers interpret typing patterns produced under conditions of varying difficulty. In two subsequent experiments, I then test how disfluent typing patterns may more directly influence interaction between group members with one another's output during a collaborative writing task. The results suggest that viewers are sensitive to differences in typing patterns and form judgements about a person's level of task understanding and work quality based on typed disfluencies.

Acknowledgment

First, I would like to express my gratitude to my advisor Sid Horton, for the continuous support, for his patience, motivation, and insights. His guidance and encouragement kept me inspired throughout even the most challenging moments of this project.

I also wish to thank my dissertation committee for their invaluable advice and direction.

I thank my labmates and other Northwestern friends, in particular Joanna Addy, for a cherished time spent together in the lab, for all the stimulating discussions and for all the fun we had.

Finally, I would like to thank my parents, my sister, my husband, and my son for supporting me always and giving life joy.

Table of Contents

List of Tables, Illustrations, Figures, and Graphs

The ability to work well with a group is necessary across many academic and workplace contexts. From groups of software developers to screenwriters for a television show, individuals collaborating as part of a group have to know how to strategize and communicate clearly with one another to establish goals and to ensure that each member of the group is updated on their current progress toward these goals. One of the potential benefits of collaboration is the access it provides to diverse skillsets and knowledge. When collaborating in groups, members can draw upon one another's specialized skills to complete joint projects in ways that might have been impossible alone. For example, consider a situation in which a team of technical writers is tasked with creating a user manual for a new medical device. Each contributing writer will have his or her own subset of expertise to draw from. A group member with extensive experience with federal agencies like the FDA would be able to contribute greater knowledge about relevant government regulations. Likewise, someone who was part of the development of the product itself can provide more technical details about the device. And finally, a group member who has worked directly with physicians or clinicians might provide insight into how to make the writing more accessible to end users. Despite these background differences, the group members would attempt to synthesize their contributions to create a single cohesive product. Importantly, succeeding at this task requires that members have shared understandings of both the demands of the task and one another's current ability to fulfill these demands.

Given the benefits of having a diversity of knowledge and skills across individuals, group members must often work to avoid misunderstandings by having a common mental model of each other and of the task at hand. Both pooling together diverse types of information and developing a collective task model require that group members develop an accurate sense of what each person knows and how others conceptualize the task as it changes over time.

Evaluating other members' knowledge is important for "information pooling," as an incorrect assessment of a partner's knowledge could lead to less effective use of the skills or knowledge they might otherwise bring to the task. Groups who are more accurate at identifying which members have specialized kinds of expertise often complete tasks more successfully than less perceptive groups (Littlepage, Hollingshead, Drake, & Littlepage 2008). Unfortunately, people are often incorrect with such assessments (Littlepage, Robison, & Reddington, 1997). Expertise is not linearly distributed with age or amount of experience, and the relationship between someone's training and their actual level of expertise on a topic may not be perfectly related. Due to these possible complications, knowledge provided through direct communication may not be enough to provide an accurate sense of what a person knows. To fill these gaps, group members can additionally rely on observations of one another's behaviors during the task to infer whether or not other members are experiencing difficulty. This combination of knowledge about one another provided through direct communication and through observation-based inferences may put group members in a better position to work with an understanding of each other's needs.

The process of negotiating differences between individuals with the purpose of working cohesively toward a common goal is referred to as *coordination*. As highlighted above, whether or not individuals in a group successfully coordinate with one another depends in part on the accuracy of each member's understanding of both their fellow group member's knowledge and also the degree to which each member's task understanding overlaps. In lieu of attaining a complete picture of what others may know, people interacting with others frequently rely on particular kinds of readily accessible information to make probabilistic inferences about what others may know (Clark & Marshall, 1981). When actively engaged in a collaborative interaction, such information may arise not only from the *content* of direct communication between group members, but also from *how* that information is delivered. In face-to-face contexts prosody or paralinguistic cues such as facial expression and gesture can provide useful information, providing support for inferences about what another person does or does not understand. However, even in settings with full access to an array of potentially informative signals, the cognitive processes behind our ability to use these types of information to generate these is not yet completely understood. Moreover, in settings with limited access to such rich multimodal signals, such as online or written contexts, not only is the understanding of *how* we use cues not fully understood, but the question of *what* cues people naturally attend to—and what factors determine the use of such cues—takes on a new dimension.

While some types of work lend themselves naturally to collaboration, such as engineering or design-based work, other activities have been traditionally more solitary. In particular, writing and editing are common tasks in which even collaborative efforts often include a high degree of independent work. Traditionally, individuals contributing to a writing-intensive task will most often work asynchronously. That is, they typically write independently and make changes to one another's work separately (Couture & Rymer, 1991). As technologies have changed and developed, though, the range of possibilities for collaborative writing has expanded. Google Docs and different forms of "wikis" have increased the potential for collaborative writing to become highly synchronous, allowing multiple individuals to easily contribute to a single document at the same time. These developments add a new set of considerations for how users in these situations make decisions about their own work in relation to written contributions from others.

As is the case with other forms of synchronous collaborations, synchronous collaborative writing requires that individuals remain attentive to changes as they are being made in real time and to respond accordingly. This requirement naturally raises questions about when and how individuals in these situations attend to the contributions of others and what kinds of actions allow people to coordinate their writing more effectively. For potential answers to these questions, I will consider the literature on face-to-face collaborative interaction, and whether the types of interactive considerations that exist in physically co-present collaborative settings bear any relation to the interactive needs of collaborative writing, in which individuals interact remotely and can "view" their partners only through the written text that appears on the screen.

More particularly, I will look to findings on how teams communicate with each other during face-to-face tasks, identifying which factors are unique to particular media and which crosscut various forms of collaborative work more broadly. Based on these findings, the experiments I propose will address the role of onscreen typing patterns as a potential cue that is unique to collaborative writing. These experiments will test the hypothesis that the collaborative activities of group members can be influenced by visual access to a partner's real-time typing behavior--similarly to how paralinguistic cues such as speech disfluencies shape face-to-face interactions.

Experiments 1 and 2 examine how typing patterns can be used to inform representations of another individual in an online communicative context. Experiments 3a and 3b investigate how onscreen typing patterns impact a real-time collaborative interaction. Experiment 3a specifically explores how typing patterns may potentially influence one's own level of attention or effort towards the task, while 3b expands on Experiment 2, testing how self-reported perceptions are shaped by disfluency during collaboration.

1.0 The Role of Partner Knowledge on Coordination

To identify how para-linguistic cues potentially shape collaborative behavior, we must consider which behaviors are prone to being shaped by the kinds of information that such cues might convey. The experiments outlined here will primarily seek to identify *how* individuals form expectations about other members through observation of typing behaviors, and *what* individuals may do with such information once they have it. As part of motivating this work, it is necessary to consider types of collaborative coordination, and how such coordination may be impacted by differences in knowledge or expectations across members of the group. The types of collaborative decisions I will discuss relate to how group members navigate through the interdependence inherent to being part of a team. Collaborative coordination can be viewed as how group members contribute individually to the collaborative task in ways that are mutually dependent on the work and decisions of others within the group (Malone & Crowston 1994). This often includes any steps that group members take to understand their role within the group—whether they do so directly by discussing how tasks are distributed between members or indirectly by attending to what others are doing in order to gain a sense of which needs are not being met and in turn what they can contribute. In other words, group members can use both *explicit* and *implicit* strategies to coordinate their behaviors, depending on the current needs of the group.

Explicit coordination typically includes overt, directed planning of task contributions, as well as spoken or written feedback to others evaluating the need for particular actions (March & Simon, 1958). *Implicit coordination*, in contrast, stems from the anticipation of other members' actions (Cannon-Bowers, Salas, & Converse, 1993; Weick & Roberts, 1993). For instance, a team of students working together to write a group essay using Google Docs might engage in explicit coordination by sending messages to one another to determine what sections each person should be responsible for. Then, they might implicitly coordinate by attending to what others are writing and individually deciding that since the other member is covering topic A, they should contribute by working on a topic that is conceptually distant enough to avoid any premature overlap. By doing so, the group members are not only basing their writing decisions on their partners' current actions, but also on the partners' anticipated actions as they attempt to write with each other's anticipated contributions in in mind.

The literature on collaborative work has primarily assumed that individuals switch between explicit and implicit coordination strategies depending on the group's momentary needs (Serfaty, Entin, & Deckert, 1993; Salas, Fowlkes, Stout, Milanovich, & Prince, 1999; Stout, Cannon- Bowers, & Salas, 1996). Other research areas, however, such as the literature on *conversational coordination*, have suggested that both implicit and explicit strategies can be used simultaneously, with implicit coordination being an on-going default process (Clark, 1996). Explicit coordination may then come to the fore when implicit means do not provide the needed information (Holler, Kendrick, Casillas, & Levinson, 2015). However, as this implicit/explicit distinction originates from the literature on collaborative work, I will primarily discuss the two strategies as being separate and frame the discussion of the two in terms of what actions teams take.

Within the collaborative work literature, implicit communication is often seen as more efficient for higher-speed, rapidly changing tasks, such as military aviation or even a shared document with a large number of contributors. Coordinating implicitly may be beneficial in these situations as frequent, extended communication may hinder progress (Cannon-Bowers & Salas, 2001; Mathieu et al., 2000). Thus, much of the discussion on implicit coordination in these contexts centers around its proposed benefit for reducing the time and effort spent on coordination as a whole (Entin & Serfaty, 1999). As a result, high stakes circumstances such as military work (Oransu, 1990) are often the settings of choice for research on implicit coordination. While some of these settings differ significantly from that of collaborative writing, I will argue that particular findings on coordination strategy can apply broadly, although later I will also examine how particular tasks might shape coordination.

As I shall discuss, both explicit and implicit coordination require people to make inferences about other group members' knowledge. However, the *role* of partner knowledge may differ between the two. With explicit coordination, an accurate understanding of partner knowledge is required for efficient planning and task allocation. As a result, activities focused on identifying each member's expertise and weaknesses are a primary aspect of explicit communication. Implicit coordination, though, may require frequent knowledge inferences to an even greater degree, given that implicit forms of coordination are often based around unspoken expectations and a presumption of having a shared understanding of the task situation. Given that a shared task model is critical for groups to be able to work in concert without constant communication, in the next sections I will consider different ideas for how teams might develop such shared understandings. Understanding the function of partner knowledge inferences on coordination generally will provide foundation for interpreting the impact of different settings and tasks.

1.1 Explicit Coordination

Direct communication about the collaborative task itself, whether in the form of spoken feedback, requests for information, or even written schedules, is categorized as part of explicit coordination. Explicit coordination also includes the process of deciding how to distribute sub-tasks, and periodic "check-ins" between team members. Explicit coordination has been traditionally viewed as one of the most important factors in determining a group's success (Foushee, 1984). Communicating with one another directly is also the coordination strategy teams most naturally employ. In an observational study of early software design team meetings, Olson, Olson, Carter, and Storrosten (1992) found that 20% of time on average was spent in explicit coordination activities. While my research questions here focus on implicit cues but will be briefly considering explicit coordination to highlight the kinds of coordination that groups and teams prioritize. Since explicit coordination plays such a large role in early collaboration, it is important to identify what enables a team to explicitly coordinate successfully.

To explicitly coordinate, as with any form of communication (whether task-based or more open-ended), speakers must tailor what they do or say based on beliefs about their partners' needs (Brennan & Clark, 1996; Grice, 1975). In task-oriented dialogue, the types of knowledge speakers must account for can be considered at many levels, from concrete immediate knowledge, such as knowing that other members of the group are currently discussing the same topic, to much more abstract knowledge about the larger scope of the task and task goals. For example, *mutual organizational awareness* is a type of high-level knowledge that includes an awareness of other team members' workload and needs (Bolstad & Endsley, 1999). For example, this could involve knowledge about whether or not another team member is struggling with their understanding of what they are supposed to do, and then being able to identify what information should be provided to help them get back on track (MacMillan, Entin, & Serfaty, 2004).

Certain explicit coordinative behaviors have been shown to be beneficial for aiding group members' ability to have an accurate sense of one another's knowledge and in turn to communicate with each other more efficiently. Among these is prioritizing the discussion of individual intentions about future actions. Groups with members that provide direct information about their expected upcoming decisions perform better than groups that communicate primarily about the current state of the task (Harbers, Jonker, & Van Reimsdijk, 2012). As shown by Olson et al. (1992) discussion of intentions plays a significant role in what groups communicate about, though it accounts for less overall coordination time than discussion of what has been completed so far. Of the groups observed by Olson et al., the majority of time spent on explicit coordination activities was spent on summaries and on walk-throughs of designs as they currently were, rather than on intentions or upcoming decisions. This kind of forward-focused explicit coordination can be viewed as distinct from other forms of explicit task-relevant negotiation, such as error correction or semantic coordination. Communication centered on coordinating a group's upcoming actions is often viewed as more time efficient (Blickensderfer, Cannon-Bowers, & Salas, 1997; Harbers, Jonker, & Van Reimsdijk, 2015), as it is fundamental to a group's ability to progress through a task.

Group structure is another factor proposed to facilitate knowledge estimation between group members and make explicit coordination more efficient. Macmillan et al. (2004) suggests that the way responsibilities and resources, such as access to information or tools, are distributed across members impacts the productivity of a team's explicit coordination. In a simulated naval mission, Macmillan et al. asked teams of six military officers to 'regain control of an allied country' through completing a series of sequential tasks. The necessary 'weapons' and information required to complete each mission's sub-task were either distributed so that individual members could work on their own sub-tasks with greater or lesser independence. When resources were distributed such that members could act more independently, the time spent on communication centered more on coordinating upcoming decisions, rather than on explaining previous actions or ensuring members have a mutual understanding of the task state. Additionally, Macmillan et al. found that individuals needed to communicate directly less often, as the clear resource allocation allowed for more accurate awareness of other member's needs. Indeed, when the resources were allocated independently, each group member's assessment of their partner's task competence more closely matched the other person's own self-assessment than when resources were distributed equally across members, regardless of their respective roles.

Determining how to allocate resources may not be as clean cut in other collaborative settings. In circumstances with greater ambiguity around which resources are required for which tasks, the role of deciding how tasks are assigned to group members becomes more open-ended. For instance, a task in which resource allocation plays a large defining role, as described in Macmillan et al. (2004), decisions made about resource allocation must coincide with task distribution decisions. Collaborative writing, however, allows for fluid resource or information distribution, as the completion of writing tasks are not generally tied to which group member is in possession of which resource and thus, without a clear task-resource connection, task distribution depends more heavily on how group members perceive one another's expertise in areas relevant to each sub-task. Group member roles may be less sharply defined, allowing for potential errors that stem from assigning tasks to non-experts (DeSanctis & Poole, 1994; McGrath, 1991). This fluidity may also suggest that while more rigid tasks often lead to sub-task allocation via external factors (such as members' work history or educational background), less well-defined tasks may involve greater reliance on interaction-based knowledge inferences.

Finally, factors within the structure of the group can also aid each individual's ability to identify other members' expertise. For instance, a group that has members from diverse domains may be more accurate at identifying individual expertise in each task area than groups with members who come from similar backgrounds, given that differences among heterogeneous group members are more immediately noticeable (Libby et al., 1987). Groups may also make inferences about one another's expertise based on *indirect* behaviors that may suggest particular kinds of topic expertise. For instance, a person who is more active in a conversation or collaborative interaction will often be perceived as having a higher level of expertise (Treem, 2013). This use of indirect behavioral cues from the interaction itself to develop a view of other group member's knowledge base will be discussed in section 2.

1.2 Implicit coordination

The definition of implicit coordination has varied within the collaborative literature, with some defining it broadly as any coordination that is based on unspoken expectations about other group members' actions (Wittenbaum & Stasser, 1996). On this definition, the *method* used to coordinate is the most relevant factor for distinguishing between implicit and explicit coordination. Alternatively, the distinction between explicit and implicit coordination has been framed by some in terms of intentionality. In this view, explicit coordination includes any form of premeditated planning, such as making a schedule, while implicit coordination typically involves more emergent forms of behavior. Espinosa, Lerch, and Kraut (2002) describe this as 'a form of shared cognition' developing through ongoing contact and unconsciously used by members to make coordinated actions. Furthermore, when a team is under a higher degree of pressure or uncertainty, they may often switch from using primarily explicit strategies of communication to more implicit modes of coordination (Serfaty, Entin, & Deckert, 1993). This shift of strategy, from explicit to implicit, enables group members to avoid the overhead in time required by explicit communication, and to focus their attention towards the task itself (Salas, Fowlkes, Stout, Milanovich, & Prince, 1999). Based on an analysis of group interactions within in a simulated naval/military environment, Serfaty et al. (1993) found evidence to suggest that this kind of switch can allow groups to maintain performance despite the potential for added ambiguity or time pressure. Additionally, certain circumstances involving unexpected changes that do not allow team members to coordinate explicitly, such as sudden technical difficulties, can also lead to a switch to implicit coordination (Entin & Entin, 2000).

It is generally agreed upon that a group's ability to successfully coordinate implicitly centers on the degree of knowledge overlap between group members (Cannon-Bowers & Salas, 2001; Crowston et al., 1998; Weick et al., 1993). This kind of knowledge overlap is often discussed in terms of a "shared mental model." A shared mental model between team members has been proposed to include common mental representations of the following elements: the task itself, the characteristics of each team member, and the anticipated interactions between team members (Hinds & Wiseband, 2003). Implicit coordination is largely based on the ability to form expectations about one's collaborative partners actions, which requires having an accurate sense of their knowledge. To form such expectations, it is especially important to have a sense of not only how much a partner knows (e.g., in terms of expertise), but *what* they know. More specifically, to make decisions as a whole, groups need to know where their overall knowledge converges and where it differs. How a group arrives at a shared mental model and what types of knowledge is most beneficial to share remain under debate. I will discuss the primary possibilities for how group shared mental models are proposed to develop, as well as how these models may be structured.

One idea is that shared knowledge between group members can overlap on hierarchically organized levels, with knowledge overlap on deeper, less accessible levels being most beneficial to a group's implicit coordination. Cannon-Bowers and Salas (1999) propose that each member of a team constructs his or her own hierarchically structured mental model of the task. Individuals are thought to develop several distinct models for each aspect of both their understanding of the task and of their fellow team members. These models are nested at different levels, such that information about the external environment would then be the broadest and most immediately accessible level. This would also be the level in which the highest degree of overlap occurs most naturally. Nested within that would be the team environment, which would include the relational structure and norms of the team. Within that level would be a mental model of the individual's own role and sub-tasks. This nested structure can continue as far as necessary, depending on the information requirements of the task. The overall degree to which each of these mental models overlaps across individuals may facilitate their ability to anticipate each other's decisions and to adapt implicitly (Kleinman & Serfaty, 1989). The more similar the task understanding among individuals, the more effective they will be at both implicit coordination and overall task performance (Mathieu, Heffner, Goodwin, Salas, & Cannon- Bowers, 2000). Overlap in more specific aspects of these shared models, such as knowledge of individual responsibilities, can allow for better implicit coordination as group members who closely understand these nuances would then be able to identify actions they can take to assist the other members.

It has also been proposed that rather than the organization of mental models being a determining factor for what information is more beneficial to be highly shared, the *content* of the shared model is crucial to this determination (Druskat & Pescosolido, 2002). Druskat and Pescosolido emphasize the role of expectations in shaping group member behavior patterns. In contrast with the hierarchical view, in which the impact of any particular element of information on coordination depends on that information's level of specificity, this view suggests that an accurate understanding of how group members are likely to behave and interact with each other is the most beneficial type of information to share. In this view, a group that has clearly defined expectations about, for instance, who in the group is best at managing conflicts, would be better able to act in concert without explicit communication than a group with highly shared task knowledge, but disparate expectations of group norms. This again differs from the hierarchical view which places task information and relational information as equals, both of which can have individually specific sub-information nested beneath.

Klimoski and Mohammed (1994) also suggest that a successful shared mental model requires that group members understand both *what* their group member's behavioral norms are and *why* these norms developed. This would mean that knowledge of what information other group members have access to plays a larger role than in the hierarchical shared model concept, as a person's knowledge is a large part of what determines their actions. Based on this information, one might speculate that inferences about a group member's knowledge that originate from direct observation as they complete their task, might be particularly beneficial for mutual understanding. As in this view, the cause of why group members have displayed a behavioral pattern is seen as a foundation for effective shared mental models, and in turn implicit coordination.

Regardless of whether the content or the organizational structure of knowledge determines what information is most important for a group to share, there remains the question of *how* the information becomes shared between group members. One possibility that has been proposed is that a shared mental model develops gradually and without direct intention, as a byproduct of collaboration itself (Cooke et al., 2003). This idea of automatic knowledge convergence being the source of a shared mental model is in contrast with other research suggesting that a shared mental model primarily must stem from a concerted effort between members to ensure their task understandings are consistent with one another (Fiore et al., 2001; Shah & Breazeal, 2010). This would mean that the ability to correctly anticipate a team member's action would be reserved for groups that actively strive to establish shared knowledge.

Stout et al. (1999) suggests that pre-planning communication is what facilitates implicit coordination, with non-verbal cues emerging as an important source of information *after* people have engaged in explicit planning to establish a shared mental model sufficient for current purposes. Stout et al. laid out nine planning strategies that potentially lead to a highly overlapping shared mental model, which include such as things as explicit discussion of goals, the handling of unexpected events, and sequencing of sub-tasks. The importance of such strategies has been further supported by evidence that teams are able to use implicit coordination strategies more effectively when members have provided information about themselves without being requested to do so (Oser, Prince, Morgan, & Simpson, 1991). This mirrors the idea that the providing intention-related information specifically improves explicit coordination, and also suggests that the eventual ability to rely on implicit forms of coordination is facilitated by prior explicit communication.

Finally, Rico et al. (2008) proposes that the way a shared mental model develops is also dependent on the type of knowledge that is being shared. In this case, both task understanding and information about the group's structure and behavioral norms would fall under a category referred to as *static knowledge*. Static knowledge is defined as information that is largely unchanging throughout the course of the entire task (Endsley & Jones, 2001). Rico et al. then suggests that while static knowledge can become shared through gradual convergence, explicit communication and a directed effort to share such information is more typically relevant for static knowledge. For instance, a shared understanding of group member roles could develop over time as a group works together and observes each other's behaviors, but it is traditionally more associated with the use of some explicit coordination and communicated decisions. This contrasts with Rico's proposed *dynamic* knowledge structures, which are defined as a continually changing situation model of the task and the group relationship. This would include beliefs about other group member's current state of mind (such as level of fatigue), the level of trust between group members, and modifications to the workspace, any of which may change over the course of the task. Again, this suggests that real-time observation of fellow group members during task completion would allow a group to coordinate more effectively. Shared dynamic models would then emerge naturally from the processes of working on a task.

Rico et al. (2008) argues that models of dynamic information play a larger role in facilitating implicit coordination, as anticipating another person's action is suggested to be inherently based on changing information. This is consistent with the view of implicit coordination as especially beneficial in high pressure circumstances. For instance, if a restaurant has a surge of high traffic, the service team might implicitly coordinate by arranging sets of silverware in advance so that the waitstaff can set up tables as customers leave at a faster pace.

This response of taking on new tasks without explicit direction would be based on a shared awareness of the current situational demands. While this distinction between models of static and dynamic team mental models is not addressed directly by proponents of the intentional view of mental model convergence, Shah and Breazeal (2010), for instance, imply that dynamic situational information would also need intentional measures for implicit coordination to work. They provide examples of pilots providing one another with unsolicited spoken updates on situational changes, stating that such behaviors lead to improved effectiveness via implicit coordination. In contrast with the business-related collaborative contexts described by Rico et al. (2008), effortful and non-effortful processes might differ in their usefulness depending on the task and its setting.

In summary, the ability to identify other group member's needs and to develop an internally consistent task model across team members determines how successfully a group will be able to coordinate, whether implicitly or explicitly. The coordination strategy a group chooses may depend on external factors such as time pressure and task consistency, as well as on the degree to which they have developed a shared mental model. While both explicit and implicit coordination can be used effectively, explicit coordination requires both time and accessibility between members whereas implicit strategies are often more efficient. As a result, it is important to understand the factors that enable a group to coordinate implicitly. As I shall now discuss, one primary factor that leads to a group being able to successfully use implicit coordination is how group members make inferences through observation. The factors that influence such assessments ultimately determine how a group makes their coordination strategy decisions and in turn how effectively they are able to work together as a unit.

2.0 Para-linguistic Cues and Knowledge Inferences

Having discussed the importance of partner knowledge inferences for a group's ability to efficiently coordinate, I now turn to the factors that shape how group members perceive each other's knowledge. As mentioned, in some circumstances individuals working together can rely on their past experiences with one another or on explicitly provided information (such as information about someone's education history) to form an assessment of what others might know. However, such information may not always be accessible, particularly in contexts in which strangers contribute to the same project while widely dispersed in time and place (as often occurs in collaborative sites like GitHub or wikis). Also, regardless of how well group members know each other, the knowledge provided from these explicit sources may not always be directly relevant for task coordination. Additionally, inconsistencies between team members with respect to their task situation models may not be apparent if group members rely primarily on static, explicitly given information. For example, a screenwriter's current uncertainty about whether a scene should be cut to create slow-burning tension would likely go unnoticed if the rest of the group *only* had access to knowledge that this person has successfully completed similar work in the past.

So, what bridges the gap between the need for knowledge about one's collaborators and the information that is accessible within the collaborative situation? In many situations, group members may often use indirect information, such as cues derived during the task from each other's behavior, to draw inferences about what everyone knows. That is, rather than relying on static or pre-existing information about other members of a collaborative group, individuals may develop an understanding of other collaborators by attending to their behaviors and making inferences on the basis of these observations (Argyle, 1972). This leads to the question: What kinds of behaviors are helpful for supporting inferences about another person's knowledge? The answer may depend, in part, on the context and the accessibility of particular behaviors. The actions that individuals might rely on to draw inferences about others are likely to take one form when team members are co-located and familiar to one another, and another form in virtual groups using video calling, or groups engaged in asynchronous writing.

Regardless of the context, however, potentially informative behaviors include more than large-scale actions prior to the interaction (such as how one chooses to organize task-relevant materials) or explicitly communicated ideas. Significant relevant information about a person's knowledge, intentions, or mood can be inferred through implicit signals in the *way* that they communicate (Abercrombie, 1968; Crystal, 1969; Hinde, 1972). In face to face contexts these signals can include gestures, facial expressions and intonational patterns of spoken utterances. In online contexts, this can include the use of typographical modifications like repeated letters, mouse movements, and chronemics (or response timing). I will be referring to these signals collectively as *para-linguistic cues*. This term is intended to encompass the broad range of informative signals whether they originate from speech, body movement, or computer-mediated behaviors such as mouse movements.

As the ultimate aim of this proposal is to present ideas relevant to collaborative writing, my eventual focus will be on cues that are most relevant to text-based settings. One of big questions about text-based collaborative work, though, is whether it is inherently limited as a collaborative medium relative to face-to-face contexts. It has been suggested that because para-linguistic cues within face-to-face contexts play such a large role in the ability to make accurate inferences, no longer having access to those specific cues would lead to an inability to make accurate knowledge inferences (Kiesler, 1986). To address this idea, as well as to better understand the factors that as a whole influence a group's ability to successfully work together, I will discuss the relationship between knowledge inferences and para-linguistic cues in both online and in-person settings.

2.1 Face to Face Cues

In face to face interactive contexts people have access to a variety of paralinguistic cues, including facial expressions, co-speech gestures, bodily movements, and of course particular features of spoken utterances like intonational contours and disfluencies. In various ways, these cues potentially provide information about relevant aspects of the speaker and what he or she is trying to say, include his or her communicative intentions, mood, and level of confidence in what is being said.

Among the most-studied types of speaker information revealed by paralinguistic cues is emotional state (Haase & Tepper, 1972; Sauter, Eisner, Calder, & Scott, 2010; Scherer, Banse, & Wallbott, 2001). The reason for this focus is clear, as cues such as facial expression are universally tied to emotional states (Ekman, 1979). But, research centered on emotion-related inferences based on paralinguistic cues can also provide insight into knowledge-level inferences. Within the broad umbrella of 'emotion' are emotional states that relate to a person's attitude toward what they are saying, including how confident or certain they are. For instance, the perception of positivity and high interest as cued through speech patterns may prompt listeners to infer high competence on the part of the speaker (Nguyen & Gatica-Perez, 2009). Cue-based knowledge inferences are also discussed independently of emotion, of course, particularly in research on impression formation, where they are referred to as *competence judgments* (Biancardi, Cafaro, & Pelachaud, 2017; Fiske, Cuddy, & Glick, 2007).

One of the richest visual sources of such information is facial expression. Using facial cues, observers are able to distinguish between complex mental states, even those which could be viewed as falling under the same category. For instance, certain facial movements have been consistently tied to mental states such as disagreement, concentration, or interest (el Kaliouby & Robinson, 2005). Our ability to make these distinctions suggests that we can derive inferences from facial expression about mental states much more abstract than emotion. Trichas and Schyns (2011) also found that perception of a person's ability to lead a group was shaped by facial expression. When the facial expressions of presented faces matched participants' described prototype of a good leader, the presented faces were rated as better leaders. The participant's responses revealed that their preferred expression signaled other abstract qualities that they considered most valuable in a leader, including intelligence, determination, and sensitivity. The facial expressions that were consistently rated as high in leadership, and thus often were rated as higher in intelligence, included specific features such as pulling the eyebrows together, whereas faces with the eyebrows raised upwards showed the opposite effect, being associated with lower leadership ratings. A later study by Trichas, Schyns, Lord, and Hall (2017) confirmed this pattern, with faces presenting more apparent displays of negative emotions, such as stress or anxiety, being associated with decreases in perceived competence. However, other work has shown that when emotional displays occur in a context that justify the emotion (e.g. displaying sadness after viewing a sad video), this can lead to higher ratings of perceived competence by observers (Gardner, Fischer, & Hunt, 2009).

Another visually-available cue that has been shown to support inferences about competence or knowledge is manual gesture. The role of gesture in conveying added information to speech has been heavily explored, both in terms of production and comprehension. Different types of gestures have been shown to play a variety of roles from aiding the speaker's ability to formulate their thoughts into words, to clarifying or illustrating abstract or spatial information to attempt to improve listener understanding (Bilous, 1992; Gerwing & Bavelas, 2004; McNeill, 1992; Pietandrea, 2002). For example, while providing directions, a speaker might make a circular gesture with their arm to make a description of where to exit a roundabout more concrete. It has been shown that the use of complex or symbolic gestures that relate directly to content being spoken about (such as doing a spinning motion while explaining how tumble dryers work) can lead to higher competence ratings by observers, while the presence of gestures unrelated to the spoken content (such as tapping fingers against each other) can lead to perceived lower competence (Maricchiolo, Gnisci, Bonaiuto, & Ficca, 2009). It appears that gestural cues that seem intentional and communicative lead to higher perceived competence, while cues that originate from unplanned responses lead to lower competence ratings. In contrast, Chawala and Krauss (1994) showed that when actors were given a script to perform either spontaneously or after reading and rehearsal, their rate of symbolic gestures was significantly higher when speaking spontaneously, while their rate of non-symbolic or motor gestures, such as finger tapping, did not differ. Also, naïve viewers were able to watch silent videos of the actors and correctly identify which were speaking spontaneously, suggesting that they were able to recognize the actors' lack of familiarity with the script through their gestures and body language. Both of these findings illustrate the degree to which comprehenders can incorporate paralinguistic information into their perceptions of speakers. Given the importance of perceived knowledge on coordination, para-linguistic cues are undoubtedly an important factor in how individuals adjust how they interact with others based on these kinds of knowledge assessments.

Beyond visual non-verbal cues, we also rely on the manner of speech delivery itself, above and beyond the content of what is said, to guide inferences about the knowledge and mental state of one's conversational partner. For example, *contrastive stress* (or shifting the emphasis on different words or syllables within a sentence) can change perceptions of the speaker's communicative intentions (e.g. "you need to see *this*" in comparison with "you *need* to see this"; Bolinger, 1961). However, with speech-based cues in particular, the line between cues that are understood as intentionally produced and those that are a by-product of the speaker's mental state is ambiguous. For example, a rising intonation at the end of a statement can be used intentionally to indicate that the speaker is asking a question. Meanwhile rising intonation can also serve as a more general signal of uncertainty, even when not intended as such by the speaker (Tomlinson & Fox Tree, 2011). Smith and Clark (1993) illustrated this through an analysis of a corpus of spoken answers to difficult questions. They found speakers who ended their answer with rising intonation were incorrect significantly more often than when they had a falling intonation, suggesting that rising intonation may indicate uncertainty even when it is not intentionally produced to do so. Here, I will focus on cues that appear to be unintentional but that may still reliably convey information that can be used to support inferences about the speaker's knowledge, acknowledging that the degree of intentionality in such cues is up for debate (Bavelas & Chovil, 2006; Clark & Fox Tree, 2002).

Among the most salient features of spoken language generally viewed as unintentional are *speech disfluencies*. Spontaneous speech is rarely delivered completely fluently, and instead includes many kinds of disfluencies such as pauses, false starts, and mid-word changes. A relevant question, then, is whether listeners use the presence of disfluencies as reason to infer particular kinds of difficulties for speakers. The role of filled pauses, such as *uh* and *um*, on dialogue comprehension has been subject to particular debate, with some researchers arguing that speech disfluencies considerably alter the way both message and speaker are perceived (Clark & Fox Tree, 2002), and others arguing that listeners mentally 'edit out' disfluencies to clearly understand the intended message (Levelt, 1989).

A variety of evidence suggests that speech disfluencies can indeed shape comprehension (Corley, MacGregor, & Donaldson, 2007; Fox Tree, 2001). For example, Brennan and Schober (2001) showed that disfluencies can help listeners disambiguate references more quickly in a communication task. In this study, a pre-recorded speaker referred to one of two visually-presented items by either speaking fluently (*the yellow square*), by beginning to refer to one item, using a filled pause, then referring to the correct item (*the pur- uh yellow square*), or by correcting the error with a silent pause (*the pur- yellow square*). Brennan and Schober found that listeners were faster to select the correct item with the filled pause present than with either the fluent speech or a silent pause. The presence of the filled pause alerted the listener to the fact that an error had been made more directly than a silent pause or an immediate correction. This indicates that the filled pause itself provides information, in this case that an error has been made, that better enables listeners to interpret the speaker's intended message.

Similarly, listeners have been shown to alter their level of attention to upcoming speech depending on whether there is a preceding filled pause or not. Fox Tree (2001) demonstrated this by having participants listen for a specific word in a recorded spontaneous conversation. Filled pauses were either left in the recording or removed. They were directed to press a button once they heard each target word. It was found that hearing a speaker say "uh" led the listeners to more rapidly recognize the subsequent target word, whereas hearing them say "um" did not impact their speed. This suggests that not only do filled pauses provide listeners with cues that may shift the way they attend to speech, but that different disfluencies potentially serve distinct roles in how comprehension is shifted. For instance, as was suggested in Fox Tree (2001), a filled pause, particularly *uh*, can be viewed as an indicator that the speaker needs more time, or rather that they are going to continue speaking after a delay (Clark & Wasow, 1998; Christenfeld, Schacter, & Bilous, 1991).

Beyond affecting listeners' comprehension of what the speaker is saying, the presence of disfluencies has also been shown to impact perceptions of the speaker. For example, Brennan and Williams (1995) found that disfluencies influence judgments of a speaker's metacognitive state during a question-answering task. In this study, listeners had to make estimates about the speaker's "feeling of knowing," or whether that speaker *felt* that they knew the answer -- independent of whether the speaker actually answered the question correctly. Brennan and Williams found that the presence of speech disfluencies, in combination with other potential indicators of uncertainty such as rising intonation, predicted judgments of the speaker's feeling of knowing. Longer pauses, both filled and unfilled, led to higher ratings of feeling of knowing (meaning that the participants believed it was likely they felt they knew the answer and were experiencing difficulties accessing that knowledge), whether the question was eventually answered correctly or not. This was suggested to be due to listeners presuming a longer mental search was occurring when a disfluency was present. Evidence such as this moves past estimating a person's overall knowledge level, to show that viewers can use these cues to make assessments about the *nature* of a person's uncertainty, such as whether it stems from true lack of knowledge or retrieval difficulties (Williams & Holland, 1981).

Given the potential impact that disfluencies have on both interpretation of a speaker's intentions and on listeners' perceptions of the speaker's confidence, disfluencies should be expected to influence the process of knowledge estimation between conversation partners more broadly. Within the context of collaboration, and attending to a collaborative partner's perspective in particular, this leads to two broad questions. Is there a clear pattern between a speaker's state of mind or confidence level, and the rate and type of disfluencies that show up in their speech? Secondly, are such patterns consistent enough that disfluency expressions shape the way a listener interprets their conversation partner's degree of certainty?

To further parcel out the relationship between speech disfluencies and challenges with planning what to say, researchers have also looked to deceptive speech. In this area, there has been conflicting evidence, with some studies suggesting that the heightened cognitive load associated with maintaining a lie leads to frequent disfluencies (Vrij et al., 2000), while others have found that the reverse appears to be true, with filled pauses in particular occurring more in truthful speech than deceptive speech. After having participants produce a set of truthful and deceptive statements, Arciuli and Villar (2010) found evidence in support of the idea that speakers produce more disfluencies when they are being truthful than deceptive. Benus et al., (2006) proposed that deceptive speech may be more carefully monitored and pre-planned, leading to successful suppression of speech disfluencies. This may suggest that speakers are aware, to an extent, of the role disfluencies play in listener perception of speaker certainty and strive to alter their speech accordingly.

The concept that listeners receive disfluencies as a signal of uncertainty is further supported by Barr (2001), in which participants were directed to click on one of a pair of items based on instructions from a pre-recorded speaker. Over the course of the task, some items appeared multiple times, while a set of new items were continually introduced during the task. Barr found that when presented with displays containing one familiar item and one novel item, participants were faster to select the new item when the speaker began their description of that item with a filled pause (e.g., "Uh… the round one") This suggests that the relation between disfluencies and perceived confidence is strong enough to not only affect judgments about the speaker, but also expectations about content. This effect has also been shown with subtler forms of speech disfluencies, such as when the pronunciation of the definite article "the" is shifted from the standard pronunciation "thuh" to the elongated form, which rhymes with *tree*. Arnold et al. (2004) found that participants would more readily look towards a previously mentioned item when a speaker directing them spoke fluidly and used the standard form of "the." When the speaker used elongated "the," whether they included a disfluent expression afterwards or not, the listener was more likely to look towards a novel item. This suggests that beyond cues such as frequency of overt disfluencies, speaker uncertainty can be detected through disfluencies cues that do not fit the typical pattern.

2.2 Computer Mediated Cues

Given the importance of multimodal face-to-face cues such as speech disfluencies and gesture on face to face communication, it could initially appear that many forms of computer mediated communication (CMC) might create unique difficulties for interaction, to the extent that they remove access to most non-verbal cues and the information they provide. Early research in fact supported the idea that forming an impression of another person based on information accessible during the interaction itself would be inherently one-dimensional in text-based platforms (Kiesler, 1986). Due to this, "richer" CMC platforms, with access to voice and video, have generally been thought of as the easiest forms of online communicative tools to use (Daft & Lengel, 1984; Rice, 1992), primarily because they come closer to recreating the multimodal experience of face-to-face dialogue. Text-based interactive platforms, then, would presumably cause users to modify their natural communication and coordinative behaviors. As described in 'the cuelessness model' (Rutter & Stephenson, 1979), users would have to compensate by adopting a style of well-planned unspontaneous communication in order to be able to work with each other in such a depersonalized setting. In this view, the more cues a medium provides access to, the better users can develop an accurate understanding of one another and thus collaborate more effectively.

This proposed connection between media richness and a group's ability to successfully complete a task using said media was called into question by Dennis and Kinney (1998). They compared task performance between teams when working on four modified platforms. Participants had to answer a series of questions in groups, in which each member had only a portion of the information necessary to answer correctly. This then required group members to share information to be able to complete the task. The platforms allowed Dennis and Kinney to compare 'richness' in terms of both *amount* of cues accessible and *speed* that such cues became accessible. The first of these dimensions compared video calling, often seen as the richest CMC medium as it provides users with access to both visual and speech-based cues, to a text-only chat system. To manipulate speed of feedback on the video platform, users were assigned to a platform in which they could video chat synchronously or send video recordings to one another back and forth as they completed the task. A similar manipulation was used for the text only platform: in a synchronous version every keystroke was made immediately visible to all participants, whereas in an asynchronous version users could only see messages once they had been fully constructed and sent, and only one participant could send a message at a time. Dennis and Kinney found that feedback immediacy significantly improved performance, in both video and text-based contexts. However, contrary to media richness theory, participants did not perform better on the task when they used the video system than when using the text-based system. These findings suggest that the type of information that is accessible is not always the most important factor in determining how well groups collaborate. Instead, the findings about feedback immediacy indicate that having real-time access to a collaborative partner's work is more beneficial to team performance.

In a similar vein, other studies have shown users to be just as adept at working with one another in text-based contexts as in video-based contexts (e.g., Valacich et al.1994). One proposed explanation is that users readily adapt to text-based platforms, creating a set of text-based equivalents to face to face non-verbal cues that improve message comprehension. Hollingshead, McGrath, and O'Conner (1993) provided support for this idea through a comparison of task performance between face to face and CMC settings. They found that the negative impact on performance initially found when groups worked on a computer mediated platform dissipated as people spent longer amounts of time using said platform. CMC may not be inherently limited relative to face-to-face communication, given the ability of people to find meaning through any information source available, an idea referred to as *social information processing theory* (Walther, 1992). As users adapt to the lower-cue environments, they learn to identify what sources of potential information remain.

What users choose to attend to in such text-based settings has been a source of debate, with Walther (1992) suggesting that explicitly discussed information will hold higher weight than it might in face-to-face contexts. This view is distinct from the previous idea that CMC is inherently impersonal because users are less able to form impressions of each other based on the interaction itself. For instance, as users can choose how to present themselves in text-based contexts, what information they disclose or exclude can take on greater meaning that it might in face-to-face contexts, in which self-presentation is less under individual control. This is supported by the theory that any provided information about group member's social categories, such as gender and age, would become more heavily relied upon in text-based contexts (Lea & Spears, 1992) as there is less other information to derive inferences from.

Another possibility, which I am exploring in this dissertation, is that users rely on information from text-based cues that play a similar role as that played by traditional paralinguistic cues in spoken interactions. These information sources would be equivalent to face-to-face cues in the sense that they originate from the delivery of the written content, rather than from the content itself or from knowledge external to the interaction, such as information about other speakers' social categories. As is the case with face-to-face cues, text-based cues can be viewed as being intentionally produced or not. On the production side, users may adapt to text-based CMC contexts by intentionally including cues that would more often be unintentional in FtF contexts. Some of these cues might include the use of emoticons to mirror facial expressions and to express emotion, or purposeful spelling changes to mirror speech inflection differences (Kalman & Gergle, 2014). Whereas in speech, meaning can be shifted through vocal emphasis on certain words, on text platforms users might add emphasis through capitalization (e.g. "you NEED to see this"). The relationship between these intentional modifications and perceived knowledge has not been explored directly, and likely depends on the norms of the setting. In formal settings, for instance, cues such as using all upper-case or lower-case spelling, the presence of grammatical or spelling errors, and the use of emoticons, have all been shown to lead to lower perceived competence (Fuller & Thatcher, 2016).

As was the case with speech-based cues, the connection between less intentional text-based cues and knowledge inferences has been more thoroughly researched. One of the most explored unintentional cues within text-based settings is message timing, or *chronemics*, and how timing shapes the interpretation of both the content sent and its producer. Conversation with a fast, consistent pace is often interpreted as more successful, with readers interpreting their conversation partner as both more engaged and more confident and prepared in what they are discussing (Walther, 2002). More interesting are the inferences that people make based on message delays and changes of pace. Longer latencies between responses have been associated with reduced trust between the sender and the receiver (Feng, Lazar & Preece, 2004; Toma, 2010; Kalman & Rafaeli, 2011). Similarly, longer message latencies have been associated with perceptions of greater introversion (Kalman, Scissors, Gill & Gergle, 2013) and higher levels of cognitive load (Davenport & Beck, 2001). These connections mirror the inferences about a speaker drawn from pauses in spoken contexts. Interpretation of delays are also shaped by what information is known about message senders. For instance, it has been shown that receiving a delayed response from a lower-status person can lead to lower judgments of competence than the same delay from high status senders (Sheldon, Thomas-Hunt, & Proell, 2006). This illustrates how text-based non-verbal cues might play a similar role as their face-to-face analogs in modulating the recipient's model of the 'speaker,' in conjunction with other accessible information.

Alternatively, such results could be taken as support for the idea that certain kinds of information receives more weight when fewer communicative cues are available. Spears and Lea (1994) proposed the Social Identity Model of Deindividuation Effects (SIDE) to explain the observed tendency to over-attribute the behaviors or relatively anonymous online partners to their few known traits. For example, if, when interacting with someone online, the only information accessible about them is their posting history, their past statements would bear more weight on how their current behavior is interpreted than the same statements might have in a face-to-face context. Hancock and Dunham (2001) illustrated this by showing that participants make more extreme judgements of a person's personality (using a Likert scale rating of Big 5 personality traits) when interacting with over a text-based platform than when communicating face-to-face. If it is the case that CMC text-based contexts lead to stronger inferences based on fewer cues, then sender non-verbal behaviors may be imbued with a larger communicative role than equivalent behaviors in face-to-face contexts. This could potentially indicate that coordination decisions in text-based collaborative work would be more strongly impacted by subtler para-linguistic cues.

As is the case with face-to-face cues, combinations of available information can lead to the strongest levels of perceived certainty (or uncertainty). Collaborators may use inferences derived from multiple sources of information to determine how they ought to coordinate with each other. For instance, a person collaborating with someone who appears uncertain based on a combination of their contribution timing as well as static information such as their work history, might choose more explicit methods of coordination when attempting to work together. Marlow, Dabbish, and Herbsleb (2013) sought to identify, in a qualitative study of open source collaboration on GitHub, what type of information users prioritize to form an impression before collaborating with each other, as well as the ways their impression affect their behavior towards one another. Participants expressed that they based their willingness to work with an unknown person in part on inferences made about the person's competence and cooperativeness from their posting history in discussion threads. Additionally, the level of complexity of the code determined in part the extent that the participant would seek out more information on its submitter. The simpler the work itself, the less active users were in seeking out information to form an impression of its creator. Participants then used these judgements in combination with a preliminary evaluation of the submitted work itself to inform their decisions on whether to accept their input or not.

Though collaborative writing shares its text-based format with GitHub and wikis, as shown in the prior example, the impact of one's impression of their collaborators may differ significantly across formats due to elements unique to collaborative writing. Next, I will discuss the medium of collaborative writing specifically, and how the nature of writing tasks may shape the interplay of partner knowledge inferences and coordination decisions.

3.0 Collaborative Writing

As discussed, inferences (whether accurate or inaccurate) about one's fellow group members influences the group's coordination strategy. Members of a group with a significant level of shared knowledge, including the awareness that this knowledge is shared, will be able to communicate with fewer misunderstandings and with less time wasted discussing irrelevant information. They will also be in a better position to coordinate without explicit communication, as their ability to correctly anticipate one another's actions will be strengthened by their accurate understanding of one another's knowledge. As I outlined previously, developing an accurate understanding of other's knowledge may often involve para-linguistic cues that serve as indicators of mental states such as uncertainty or confidence. By being attentive to what other people are doing and how they are acting, people can coordinate and communicate with them more efficiently.

The view of the role that the communicative setting plays in a person's ability to stay attentive to these paralinguistic cues has shifted over time. While face-to-face settings are generally thought of as providing the richest set of communicative cues, text-based settings are now viewed has having their own sets of cues available to users. As suggested by Dennis and Kinney (1998), the most important factor is not the type of cues available, but the level of immediate access to these cues. For collaborative writing, accessibility to para-linguistic cues and as well as forms of direct communication can vary significantly. For instance, the cues accessible to a group working on separate documents while physically co-located would differ significantly from a group using a shared document asynchronously. This accessibility has shifted with the introduction of shared workspace platforms (such as Google Docs). Factors that are unique to collaborative writing (such as whether a group is writing synchronously and whether a workspace is shared) as well as factors that impact collaborative behavior generally (such as familiarity of group members or how similar group member's skill or areas of expertise are) likely shift the ways that partner knowledge inferences influence behavior within this specific setting.

Much of the research on remote collaboration centers on dialogue, whether that be general conversation or task-based dialogue. Although findings from this work can undoubtedly apply to the understanding of collaborative writing, there remains questions unique to collaborative writing that cannot be answered with the study of CMC spoken dialogue platforms. Before getting into the differences within the overall frame of collaborative writing, I must acknowledge the aspects specific to writing that may lead to differences in coordination strategy. Writing can be described as an open-ended design process (Neuwirth, Kaufer, Chadok, & Morris, 2000). This means that collaborative writing involves the creation of some form of shared output under circumstances in which the process of design is open and unstructured. This is in contrast to other shared output group work such as software design or building a specific object, in which the specifications for making design choices are often more pre-defined. This degree of openness, together with a requirement for cohesion, means that differences between group members, whether in terms of their background or their situation model of the task, can be more impactful than in tasks with concrete specifications (Erkens et al., 2005). Due to this, the more differences that are present among members involved in a collaborative writing task, the more difficulty they will have in coming to a common interpretation of the task (Gabarro, 1987).

As is the case with other forms of collaboration, co-authors must potentially come to a shared understanding at three different levels: 1) the content or current state of the project, 2) the organization of both the work itself and the division of labor between members, and 3) the relationship between collaborators. The last of these is important as it determines how group members interact with the work that others have created (Fish, Kraut, & Leland, 1988). All else being equal, members of equal status may be able to freely edit one another's work, while unequal members (such as an employee and their boss) or group members avoiding conflict may not. The type of information needed at each of these levels also differs depending on the stage of the task. For instance, collaborative writing media that provide greater access to immediate direct communication (typically through the inclusion of a separate chat function) are often preferred by users during planning stages, which may require a higher degree of coordination to develop a mutual understanding between collaborators (Kraut, Galegher, Fish, & Chalfonte, 1992).

Much of the research on collaborative writing has explored ways to make coordination easier through design modifications of online writing platforms, experimenting with the addition or removal of platform features that shape different collaborative concerns. For example, Fish et al. (1988) designed a platform called Quilt that included features for more accessible messaging and annotation. Other researchers, such as Kim and Eklundh (2001), have then tested these theory-based platforms to better understand how real users might interact with these features. Through this, they have identified gaps that exist between theoretical frameworks for how coordination within these environments *should* work and what users actually value (Cerratto, 1999; Dillon & Maynard, 1995). For instance, some research has suggested that the use of hypertext to link together ideas that are mentioned repeatedly within a document would benefit collaborative teams. Trigg and Suchman (1989) proposed that hypertext would improve a group's ability to arrive at a shared understanding of one another's work, as well as how each idea fits together. However, it was found that hypertext went widely unused when included in writing platforms (Dillon, 1993).

Additionally, the ways that each of the three factors mentioned above—group relations, the degree of access to the work in its most current state, and task distribution—affect a group's coordination decisions is determined in part by *when* the collaboration takes place. Specifically, coordination is shaped by whether groups are co-writing independently then proceeding to send one another their independent work (asynchronous writing) or co-writing in real time (synchronous writing). There can be asynchronous writing on a shared document by group members unknown to one another (as is the case in wiki contributions), or asynchronous writing can take place without a shared document (in the case of member's sending one another their own work or edited versions). Synchronous writing, however, nearly always involves a shared document (Kim & Eklundh, 2001; Mitchell, Posner, & Baeker, 1995). These different circumstances will lead to different coordination strategies between group members. In turn, the difference in coordination needs may lead to shifts in the importance of accurate knowledge inferences and paralinguistic cues.

The common thread between these different situations is the group members' need to maintain *situational awareness* – that is, to keep track of the state of the task along with their collaborative partners' actions and intended actions (Endsley, 1995). This concept has also been referred to as *group awareness*, defined by Dourish and Bellotti (1992) as maintaining an understanding of other member's actions in a way that forms a context that determines one's own task actions. Vertegaal (1997) broke down the broad concept of group awareness even further as applied specifically to collaborative writing. The two major categories he identified were *workspace awareness* and *conversational awareness*. Workspace awareness includes any form of information derived from the content, whether on screen in a shared workspace or information provided later. For instance, questions such as "Where are the other authors located within the document?" or "Which areas of this workspace can others see?" would fall under the category of workspace awareness. Meanwhile, conversational awareness might include similar types of information but from the angle of group interactions, such as "Which group members have interacted with each other directly?" or "When do I expect other members to communicate in the future?" Next, I will discuss different types of collaborative writing settings in terms of which aspects of group awareness are most relevant and how this informs a group's chosen coordination behaviors.

3.1 Asynchronous Writing

Asynchronous writing raises questions about collaboration that stem from the process of integrating individually constructed work. As changes to the written document are not transmitted to all group members immediately as they are made, the aspect of group awareness most relevant in these contexts is *change awareness*, or how easy it is for users to notice and keep track of changes to the document made by other members. This is particularly relevant in circumstances that do not include shared workspaces, as users then receive altered work to which they did not have access at the time it was being developed (Kraut, Galegher, Fish, & Chalonte, 1992). Due to this, the way a platform's design impacts the users' ability to detect change shapes the entire group's coordination strategy. Not having an immediate signal that a change has been made often leads to misunderstandings between group members, as critical details get made and overlooked (Tam & Greenberg, 2004). If the platform itself does not provide notice that changes have been made, group members may be forced to notify one another through explicit communication (Lowry & Nunamaker, 2003). In one study that looked at collaborative writing (Lowry, Lowry, & Curtis, 2006), teams using the same writing platform (which did not include change notifications) were either instructed to construct a detailed procedural script for how to complete the writing task or simply told to 'brainstorm' as a group before beginning. The teams who explicitly coordinated outperformed the other group on several measures, including less time spent drafting, higher ratings of one another as good collaborators, and greater satisfaction with the final writing content. This might suggest that the ideal way to facilitate collaborative writing would be to provide maximum transparency, with any edits always being highlighted to other users, to promote accurate and updated situational awareness between all members.

While explicit coordination may be the primary strategy used to coordinate asynchronous writing, Kittur and Kraut (2008) also analyzed how wiki editors use implicit means to integrate their work more effectively. They found that articles with little direct communicative activity (based on the discussion page) nevertheless show patterns of content convergence despite the lack of explicit discussion. Users' contributions might conflict with one another or vary in scope toward the beginning of an article's lifespan, but as other users continue to contribute, a unified consistent voice and scope often develops over time without discussion. This could be considered a form of implicit coordination based entirely on workspace awareness. Rather than communicating directly to develop a shared style between contributors, a single or some small number of users will cultivate a set of norms for the article, such that other users, upon viewing it will have enough information to accurately form expectations for how the article content should be modified or expanded upon. The information in the workspace itself is enough to provide unrelated users with a shared situation model that influences the way that they collaborate on the writing.

3.2 Synchronous Writing

Synchronous writing, in which multiple collaborators work on a shared document simultaneously, has been found to be less common than other forms of collaborative writing, at least in user surveys of academics who frequently write in groups (Kim & Eklundh, 2001). However, as commercial platforms that promote synchronous writing have become more common, new attention has been focused on this style of collaboration. Current research even suggests the process of real-time collaborative writing may have positive impacts on writing ability, particularly when introduced early in life (Krishnan, Cusimamo, Wang, & Yim, 2018). The previous discussion on change awareness was mostly relevant to asynchronous collaborative writing, given that noticing change is inherently more difficult when working separately. Within synchronous context, workspace awareness is generally viewed as the more relevant type of information for user coordination (Gutwin, 1997; Tam & Greenberg 2006). However, depending on the nature of the task and the number of contributors present there is potentially more information available in a dynamically changing workspace than one could attend to. As in face to face contexts, keeping track of all available information, particularly while working on the task itself, could be cognitively demanding (Endsley, 1995). In a survey of professionals who often write collaboratively, many expressed that they found synchronous writing distracting due to the amount of activity visible within the workspace (Wang, Tan, & Lu, 2017). Thus, three relevant sub-questions within workspace awareness are 1) to what extent it is beneficial to keep track of a co-collaborators' behaviors? and 2) what types of information present in the workspace might help users work together more effectively? Additionally, 3) how does an awareness of co-collaborators actions influence one's own task contributions?

Insight into these questions can be obtained through considering those aspects of synchronous collaborative writing that are most similar to face-to-face collaboration. As has been shown to be the case with both FtF and asynchronous collaboration, explicit coordination can improve a group's overall ability to successfully complete a task, even in synchronous writing. Access to real-time direct communication with other group members may help avoid the problem of distraction and overload, as users can coordinate their work strategy so that each member knows which aspects of the workspace are relevant for them to attend to. Yeh (2014) compared the quality of essays composed by groups of three individuals working simultaneously in a system based on Google Docs. In addition to collaborating on the essay itself, the groups used the chat option to communicate *about* their writing either heavily, moderately, or lightly. Groups who directly communicated more frequently produced essays that were more internally cohesive, had fewer grammatical errors, and were more accurate in terms of the content. Consistent with previously mentioned findings about the planning stage of writing (Neuwirth, Kaufer, Chandhok, & Morris, 2000), these results showed that the majority of the chat time was used to plan and organize the writing process, with less chat occurring during the writing or editing stages. The impact on quality suggests that a higher level of information is beneficial for the planning stages of synchronous writing.

It could be the case that explicit coordination causes users to focus their attention only on information within the shared workspace that is relevant to them. This in turn might lead to further segmentation and functionally independent work. Individuals may then work separately, despite being in the same nominal workspace. Gutwin, Greenberg, and Roseman (1997) proposed that an additional problem can originate from group member's actions not being visible during tasks in which a workspace is meant to be shared. During many longer-form tasks a partner's workspace may often be completely obscured due to length and the limits of screen size. To address this, Gutwin et al. explored alternative formats intended to help users readily notice important information in their shared workspace. For example, to keep the collaborator's behaviors directly accessible and changes easy to monitor, they added a "radar" display that consisted of a smaller schematic display of the overall workspace, alongside the larger yet narrower immediate workspace plus a smaller display of the collaborative partner's immediate workspace. They found that while the radar view did not provide access to details such as a partner's mouse movements, access to this global workspace led to better ratings from users and higher levels of interaction between collaborators.

Despite such findings, the more common format for collaborative writing remains a single shared workspace in which the actions of collaborators working on different sections of the task may or may not be visible. Many platforms do include a dialogue option for users to communicate directly about their task outside of the document workspace itself. The findings described above suggest that users benefit from having both localized access to their own portion of the task as well as a view of the common workspace. This allows them to direct their attention between their task and general task areas that become relevant, as learned through the use of the separate chat function. Both of these benefits can be viewed primarily as things that aid explicit coordination.

In contrast to work on face-to-face interaction and other forms of collaboration, the potential role of implicit coordination in synchronous writing has not been directly explored. To review, implicit coordination occurs when a group does not organize their actions with one another through direct communication, but rather are able to work in synchrony through having accurate expectations of one another's upcoming actions. A group's ability to implicitly coordinate is dependent on the degree to which each member's mental model of the task is consistent with the other members. Beyond the task itself, group members must also have an accurate understanding of one another to be able to coordinate implicitly. Based on these aspects of implicit coordination, it could be imagined that synchronous writing, as an environment in which group members can potentially observe one another's behaviors in real time, would be a setting in which groups frequently use implicit coordination strategies.

Decisions about how much work to take on individually as a member of a collaborative group is a form of implicit coordination, as such decisions are often made spontaneously and without explicit discussion amongst the group. The literature on social facilitation has discussed how the 'mere presence' of others shapes such decisions. In the context of synchronous writing, regardless of a group member's ability to effectively track the actions of other members, simply working within a shared document may influence their work strategy. Jun, Meng, and Johar (2017) illustrated the impact of social loafing within synchronous shared workspaces. In their study, participants were tasked with flagging a series of statements for whether or not they contained fact-based errors. When participants believed that other people were completing the task alongside them, conveyed by having the usernames of other participants appear on the document, they flagged fewer items as being likely to have a factual error, including implausible statement items. Solely viewing the names of others on a shared document, despite not being able to see anyone else’s contributions, appears to have been enough to lower task vigilance.

If synchronous writing does in fact rely heavily on *implicit* coordination strategies, it is natural to ask about the behaviors observable through synchronous writing that might function as sources of inferences about one’s partner. One obvious source is simply having access to the products of each partner’s work as it is being generated. Knowing exactly which aspects of the task are being addressed by whom would allow group members to organize their own roles without needing to explicitly discuss them. Additionally, though, synchronous writing – unlike many other forms of text-mediated interactions – also provides access to real-time information about the dynamics of users’ typing and cursor-movement behaviors. Shifts in typing patterns have been reliably shown to be indicators of different mental states, such as stress, cognitive load, and mood changes (Banerjee et al., 2014; Epp, Liphold, & Mandryke, 2011; Vizer, Zhou, & Sears, 2009). As such, typing dynamics could potentially function as a rich para-linguistic cue about the partner that is available in synchronous collaborative writing. This possibility suggests a potentially rich area for developing both an understanding of group coordination strategy generally and collaborative writing more specifically.

4.0 Conclusion

In summary, collaboration requires multiple people to work as a unit to complete a singular shared goal. The process of determining *how* to effectively work together as a unit is referred to as coordination. The coordination strategy a group chooses to use in a given moment depends on both the current circumstances (such as level of time pressure) and the degree to which group member's share a task understanding and are able to anticipate one another's behavior. In instances in which the group has the time and accessibility to do so, group members may choose to coordinate explicitly, through communicating about their intentions or the task state. They may also be required to do so if they lack the information necessary to be able to act in accordance with one another without prior discussion. Alternatively, groups can coordinate implicitly once group members have overlapping task mental models and accurate knowledge of one another's task model.

One major factor in a group's ability to successfully coordinate implicitly is each member's ability to maintain an updated sense of their partner's knowledge and task understanding. This requires that members base their concept of one another's knowledge not only on explicitly provided information, such as past experience, but on current incoming information. In writing tasks, the ability to do this is strongly related to whether collaborators are working within a shared document or working independently. As para-linguistic cues from a shared interaction drive a variety of inferences about one's partner, within shared writing workspace environments there should be greater potential for implicit coordination between collaborators, due to the greater availability of dynamic interaction-based cues relative to writing separately. While text-based computer mediated cues have been explored, some of these cues like emoticon use or intentional misspellings may be less relevant for collaborative writing tasks.

Other text-based cues, such as chronemics, have primarily been discussed in asynchronous contexts, despite the possibility that users in a shared workspace might make inferences about one another based on delays or timing shifts in their partner's real-time typing behavior. As such, there remains possibilities to further identify information available within a shared workspace that might be influential on group members' perception of one another and thus their coordination decisions.

A common thread illustrated in each of these ideas is that people have the capacity to adapt across disparate settings and task types. This is demonstrated not only by how people switch between implicit and explicit coordination depending on the current needs of the group, but also by their ability to readily make mental state inferences based on text-based cues. The fact that people are able to derive rich information from text-based cues, which bear little surface resemblance to the kinds of communicative signals present in traditional face-to-face interactions, leads to the question of what other aspects of text-based interaction might be informative or beneficial to include within collaborative platforms?

The range of potential answers to this question may be fairly broad and would likely include cues that are less direct than those available in face-to-face settings. As described by Lea and Spear's (1992) SIDE theory, in collaborative settings with less access to multimodal information (such as collaborative writing), the importance of para-linguistic cues may shift such that viewers attribute more meaning to subtler information. This was further illustrated in Birnholtz and Ibara (2012) in which students made sweeping inferences about their collaborative partner based on detailed aspects of their editing traces, such as inferring personal criticism based on the timing of other group member's edits to their work. The implication of this shift is that within online settings involving collaborative writing, users may be especially likely to attend to less overt cues and make inferences about a partner based on these cues. As new collaborative writing platforms change and develop, particularly as synchronous writing becomes a more common form of collaboration, access to indicators of a collaborator's mental state may become more broadly defined. Potential cues such as viewing another person's mouse movements, which has been shown to be informative but is not accessible in any shared writing platform, may be beneficial to include. As new cues are identified and explored, the question of how to maintain balance between providing access to informative cues and avoiding information overload becomes even more important. Moving forward, testing the impact the inclusion of potential cues, like a collaborator's typing in real time, on the inferences users make, can further determine how collaborative writing can be made easier.

Experimental Studies

In a series of experiments, I tested the hypothesis that visual access to a collaborative partner's typing pattern can influence the kinds of inferences that people make about that partner, as well as how people contribute to the collaborative task, similarly to how paralinguistic cues such as speech disfluencies shape face-to-face interactions. As a first step toward examining the impact of real-time text-based cues on estimates of others' knowledge, *Experiment 1* involved a typing production task to identify specific visible typing patterns associated with task uncertainty. *Experiment 2* followed up on this by presenting recordings of both certain and uncertain typing collected in Experiment 1 to naïve observers, who then rated each clip's typist on several confidence measures based on their typing patterns alone.

In *Experiment 3a* I similarly explore how observations of another's typing fluency can influence perceptions of the typist. However, in this case participants work directly in a pseudocollaborative editing task in which they were presented with pre-recorded typing from a partner. Importantly, the partner's typing was designed to be either fluent or disfluent. After working with this partner in a joint editing task, participants were then asked to edit the confederate typist's work independently, and I examined whether the extent to which they edit the partner's work was influenced by the partner's previous level of fluency. *Experiment 3b* featured the same paradigm of participants engaging with a typing script in the same task, however in this case I directly probed their perception of their collaborator by having them rate the quality of the revised sentences. This enabled me to assess the impact of fluency on viewer's perception of a collaborator's work quality. Together, these experiments will allow me to draw conclusions about the role visual access to typing plays in coordination when groups collaborate synchronously.[1]

[1] Finally, *Experiment 4* attempted to test how typing fluency shapes viewer's immediate expectations about what a partner might refer to. In this experiment, participants observed brief clips of typists describing an ambiguous shape and select from two options of such shapes which they expect the person in the clip to be referring to, as quickly as they feel they know which item is being described. The exception is that typing fluency served as an early disambiguating cue, with which they could use to make their selection. This experiment was not completed due to a series of technical issues which are elaborated upon in the appendix.

Experiment 1

In this experiment, I sought to identify how individual typing patterns would be impacted by how confident or secure a person feels in what they are writing. This is a necessary first step to establish that a person's mental state can indeed lead to differences in typing that could potentially be observable in a shared workspace. Based on the literature on typing behavior, I expected that fluctuations in typing speed, rate of errors, and rate of pauses would be impacted most by user uncertainty, as those particular features are the most impacted by changes in a user's mental state (Wobbrock, 2007). I aimed to understand both the extent that these patterns are influenced by uncertainty as well as whether the *nature* of one's uncertainty might affect these patterns differently. Specifically, I tested how typing behaviors might change depending on the typist's degree of familiarity with what they are typing versus the general difficulty of knowing what to type.

As discussed, the current literature on remote text-based collaboration (and specifically the role of visual communicative cues during remote collaboration) has heavily focused on cues based on the content being typed, such as letter repetition (Darics, 2013), mouse movements (Freeman & Ambady 2010), and chronemics (i.e., the relative timing of character or word production; Kalman, Scissors, Gill & Gergle, 2013). In domains beyond remote collaboration, however, other aspects of typing behavior have been more thoroughly explored. For example, research on 'biometrics' has focused on individuals' unique typing behaviors as a security user authentication measure (Joyce & Gupta 1990). On this view, typing behavior can provide substantial information about the user producing it, both in terms of durable personality traits and temporary cognitive or affective states such as mood. Although much of the research in this domain has focused exclusively on the typing behaviors themselves, and not necessarily on the potential use of typing behavior as a communicative cue, some of this research has addressed questions related to how mental states shape these behaviors, from increased cognitive load (Rheem, Verma, & Becker, 2018), emotional state (Epp, Liphold, & Mandryk, 2011), to having to produce deceptive statements (Banerjee et al 2014).

While some of the typing pattern differences in these studies could be visually detectable to an onlooker (such as general speed), many of the typing features identified in this work would not be visually observable to an outside observer, such as the degree of pressure the typist used to press each keystroke or duration of individual key presses. Given my primary interest in establishing typing as a useful cue in interactive settings, I chose to focus primarily on typing behaviors that would be visible within a shared workspace, such as typing speed and error corrections. To manipulate these behaviors, I provided participants with task materials intended to make it either easy or difficult for participants to compose adequate messages. In the experimental task, they were asked to type in descriptions of sets of images as if they were identifying them for a partner with access to the same set of images. As has been illustrated in experiments manipulating typists' mental states (Khanna & Sasikumar, 2010), I predicted that, when writing descriptions for complex or ambiguous images, participants would exhibit typing behaviors similar to those associated with states caused by stress or cognitive load. Specifically, I expected participants who were describing difficult items to display more frequent changes in typing speed, longer and more frequent delays (particularly delays that occur within the middle of words), and more backspaces, including both low-level "typo" correction backspaces as well as backspaces associated with phrasing and conceptual reformulations.

I also gave participants the opportunity to become more familiar with a subset of the images as the repeated over multiple rounds of the task. I expected disfluent typing to decrease as participants became more familiar with the items, mirroring participant behavior in verbal communication studies with similar paradigms (Lysander & Horton 2012), where description lengths and verbal disfluencies are reduced with repeated experience, even as speaking rate increases. To examine this, I recorded the participants' keystrokes as they carried out the task. In this way, Experiment 1 allowed me to both examine whether visible aspects of typing are impacted by uncertainty and to identify the degree of impact that can be expected.

Method

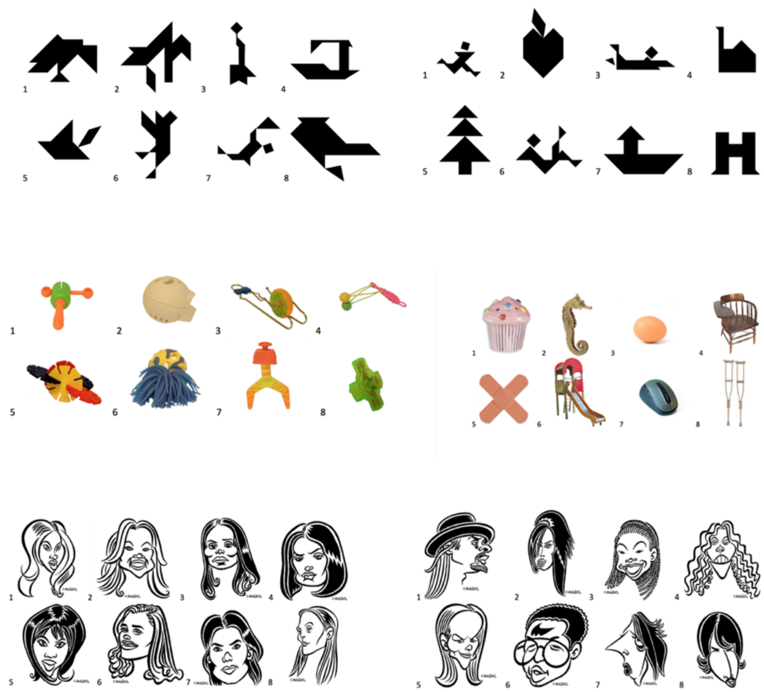
Participants

Twenty-eight Northwestern undergraduates aged 18 to 22 participated in this study for partial course credit. All were native English speakers.

Materials

I selected 48 unique images for the image description task. These images included three types of items: black and white abstract "tangrams," black and white facial caricatures, and color photographs of objects. For each item type, the images were further divided into Easy and Difficult subsets based on the complexity and ambiguity of the image. The complete sets of images for all three item types are shown in Figure 1.

Figure 1.



Item Selection

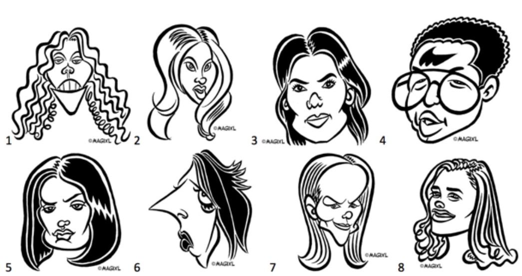
Tangrams. The tangrams were chosen from several sources, including past research on referential communication (Hupet, Seron, & Chantraine, 1991; Lysander & Horton, 2012) and web-based resources (e.g., https://www.tangram-channel.com). We distinguished between "Easy" and "Difficult" tangrams based in part on the degree to which they could be identified with a clear label (i.e., *codability*; Hupet et al. 1991). The Easy tangrams consisted of eight images that could be easily recognized as human-like figures or simple objects, such as a man running or a tree. The Difficult tangrams consisted of eight images—some human-like, some not—that were more abstract or that did not lend themselves as readily to an obvious interpretation.

Objects. The objects were all color photographic images of realistic objects. For the Easy objects, we selected eight images of familiar, easily identifiable objects, like a cupcake or chess piece, from stock photo sources. The eight Difficult objects were selected from the Novel

Objects and Unusual Names database (Horst & Hout, 2016), which includes sets of three-dimensional brightly colored images specifically chosen to be challenging to describe or identify. Horst and Hout (2016) demonstrated a lack of consensus across both adult and child participants in how they referred to these difficult-to-name objects.

Faces. The items in the face category were selected from an online collection of celebrity caricatures (http://www.magixl.com/cliparts/pop.php). For the Easy faces, I selected eight caricatures that included at least one highly distinctive feature, with clear variation across images. For instance, one caricature included a salient hat, while another included glasses. For the Difficult faces, we selected eight caricatures that lacked a single distinguishing feature and were overall more similar across the set. In general, these difficult caricatures were all images of women with small or unexaggerated features, differing primarily in terms of hair style and color, as shown in Figure 2.

Figure 2.



Set Creation

For the image description task, I organized these items into sets of eight images. For the initial "Novel" blocks, each item appeared as part of a homogenous set involving a particular category and difficulty level (e.g., all eight Easy tangrams). Then, half of the items in each image category were included as part of additional "Repetition" item sets for the subsequent blocks of the task. These repeated item subsets always included four Easy and four Difficult items from a given category. For example, the repeated facial caricature set included half of the Easy faces and half the Difficult faces, as shown in Figure 2. These critical subsets repeated twice in two Repetition blocks, with the same images in a different presentation order for each block. In this way, participant "certainty" about how to describe each item was manipulated in terms of image difficulty as well the participants' accumulated experience with the images.

Task Assembly

I implemented the image description task through a computer-based survey constructed within Qualtrics (https://www.qualtrics.com). Each task section presented eight images belonging to a particular item set, in two rows of four near the top of the screen. Each image was accompanied by a number from 1 to 8 (as shown in Figure 2). Below this were two text entry boxes where participants could type in their responses. The first text entry box asked participants to enter the number of the item being described, while the second entry box was for the actual item descriptions.

Each survey page elicited a description of a single image. To obtain descriptions for all eight images in each item set, eight copies of the same Qualtrics page were arranged in succession, each presenting the same eight images but asking for a single description at a time in image order. This constituted one block of the task. After the eighth page, the item set changed for the next task block and eight different images were presented for the next eight Qualtrics pages. The six homogenous item sets (each containing all eight images of one item type and difficulty level) appeared first, in six successive Novel blocks. This was followed by six Repetition blocks, in which half of the items in each category and difficulty level were presented twice more in six successive mixed-difficulty sets.

We created two versions of the Qualtrics survey to counterbalance Novel set order across participants. In both versions, the first six Novel blocks alternated between homogenous sets of Easy and Difficult items, but image type was staggered in a fixed fashion. Version A presented the Novel item sets in this order: Easy Object, Difficult Tangram, Easy Face, Difficult Object, Easy Tangram, Difficult Face. Version B presented Novel item sets in this order: Difficult Object, Easy Tangram, Difficult Face, Easy Object, Difficult Tangram, Easy Face. This was followed in both versions by the same six Repetition blocks presenting mixed-difficulty item sets in this order: Mixed Objects Repetition 1, Mixed Tangrams Repetition 1, Mixed Faces Repetition 1, Mixed Objects Repetition 2, Mixed Tangrams Repetition 2, and Mixed Faces Repetition 2. Qualtrics randomly assigned each participant to one of these two task versions.

Procedure

Participants were tested individually in a sound-attenuating booth. To continuously record participants' typing throughout the session, we used the RUI-Recording User Input keystroke recorder (Kukreja, Stevenson, & Ritter, 2006). This keylogger records the identity and timing of each individual keystroke (specifically, when a key is pressed, not when it is released), including nonprintable function keys like backspaces, tabs, and hard returns, along with the timing and screen locations of mouse button presses. I also used a screen recorder (https://www.ispringsolutions.com/) to record the visible activity on the participant's computer screen during the task. At the beginning of each session, after explaining the image description and obtaining informed consent, the experimenter initiated both the screen and keystroke recorders, which then ran in the background for the entirety of the session. The screen recordings were collected for use in Experiment 2 and will be discussed further then.

For each item set, participants were presented with all eight images in the set, in numbered order. The instructions informed participants that their task was to type a description of each item, one by one, as though they were communicating with another person who had access to the same images in a different order. This was intended to provide participants with a sense of the degree of detail necessary for identifying each image. Participants typed their description in the main text entry box. Before each description, they were instructed to enter the number of the item they were describing in a separate text box above the main box. Additionally, after participants finished describing a complete set of eight items, they were instructed to enter a specified symbol (such as three percentage marks: "%%%") before moving to the next item set. This procedure inserted identifiable markers into the keylogger file that could be used later to subdivide the raw typing output into the data associated with each item set and block.

All 48 items were presented once in their respective homogenous sets before any items repeated. Once the participants had completed all six Novel blocks, they immediately continued to the first three mixed difficulty Repetition blocks (Repetition-1). After this, they completed the same three mixed difficulty Repetition blocks a second time (Repetition-2). At the end of the task, they left the testing booth and were debriefed.

Design

Participants described 48 unique images. Twenty-four of those images (four from each item type/difficulty combination) were described only once, during the initial Novel blocks of the task. The remaining 24 critical images (again, four from each item type/difficulty combination) were each described a total of three times: once during the Novel block, and twice more in Repetition blocks 1 and 2. Our analyses will focus on these critical items that repeated across the entire task. Item Difficulty (easy, difficult) was a within-subjects and between-items factor, while Task Block (novel, repetition-1, repetition-2) was manipulated within-subjects and within-items. Item Type (tangram, face, object) was also manipulated within-subjects, but will not be considered as a separate factor in the analyses.

Data Preparation and Analyses

As mentioned previously, the keylogger recorded information about each keystroke, including its identity (i.e., the particular letter, number or punctuation key pressed, plus nonprintable characters such as SHIFT or SPACE) and the precise timing of the keypress, measured in milliseconds from the point at which the keylogger was initiated. The keylogger also recorded screen coordinates and timestamp information for each mouse click on the screen. For one participant, a technical issue caused the keylogger to stop recording halfway through the first Repetition-2 block, resulting in the loss of data from the final 20 items for this individual.

Before carrying out any analyses, I prepared the datafile in several ways. I first identified all keypresses used to mark the start of each item description (i.e., the numeric characters 1-8 entered by participants to indicate which item they were describing) and the keypresses used to mark divisions between item sets (i.e., special character keys like “%%%”). I used these markers to add item-specific coding information (e.g., for individual items, item types, and blocks) to the data file. I also used the SPACE keypress to identify the ends of individual words, while periods were used to identify the ends of sentences.

I then carried out initial trimming by removing all keypresses that were part of managing the task and therefore not part of the actual item descriptions. Specifically, this involved removing those keypresses used to segment the data into separate item descriptions, as well as all data associated with mouse clicks (which usually appeared before or after a given description, when participants moved the cursor to start typing in a text box or to go to the next item). In total, this removed 4.5% of the raw keypress data, leaving 160,954 individual keypresses across all participants and items.

All statistical analyses in Experiments 1 and 2 were carried out by estimating mixed effect models using the *lme4* package (version 1.1-23; Bates, Mächler, Bolker, & Walker. 2015) in R (version 4.0.3; R Core Team, 2020). For Experiment 1, all models included Difficulty, Block and their interaction as fixed effects. Item difficulty was sum coded (difficult = +0.5, easy = -0.5). For item block, we defined two contrasts that allowed us to compare behaviors in the Novel task block to each of the subsequent Repetition blocks: Novel vs. Repetition-1 and Novel vs. Repetition-2. I did not include Item Type as a factor in the analyses because, again, I did not have any *a priori* reason to be interested in behavioral differences across image categories. For standard linear mixed models, estimated *df* and *p* values were obtained via the *lmerTest* package (Kuznetsova et al., 2017; version 3.1-2) using Satterwaithe's approximation. For generalized linear mixed models, I report significance values based on Wald's Z.

All models included by-participant and by-item random intercepts. Note that including the by-participant random intercept term, in particular, accounts for variance in typing behaviors specific to individuals. When possible, I fit the maximal random effect structure justified by the design (Barr et al., 2013). For models that did not converge, I simplified the random effect structure until a successful model fit was achieved. In the results below I report the final model for each analysis.

Results

Keypress Count

First, I report an analysis of simple keypress counts, computed as the number of keypresses per item description. This is intended to be analogous to measures of description length traditionally reported for referential communication tasks (Clark & Wilkes-Gibbs, 1986). In the current context, average keypress counts necessarily included *all* keypresses, including those additional keypresses related to error correction (which we consider next). As such, these total counts do not perfectly correspond to the lengths of the final descriptions. Even so, we should expect more difficult items to elicit longer descriptions and therefore more keypresses on average. Also, average keypress counts should decline over the course of the task, reflecting shorter descriptions as participants accumulate experience in describing these items.

Figure 3 presents a boxplot of the number of keypresses per item description, by item difficulty and task block. For this and all subsequent boxplots, the dark bar within each box represents the median (50^{th} percentile) and the upper/lower boundaries of each box represent the interquartile range. As expected, participants generally engaged in more typing for difficult items (M= 118.9 keypresses per description, SD = 105.9) than for easy items (M=69.8 keypresses per description, SD = 67.8), and also typed more for descriptions during the first Novel block (M =

142.3, SD = 126.0) than they did in successive blocks (Repetition-1: M = 82.5, SD = 58.2; Repetition-2: M = 56.8, SD = 46.8).

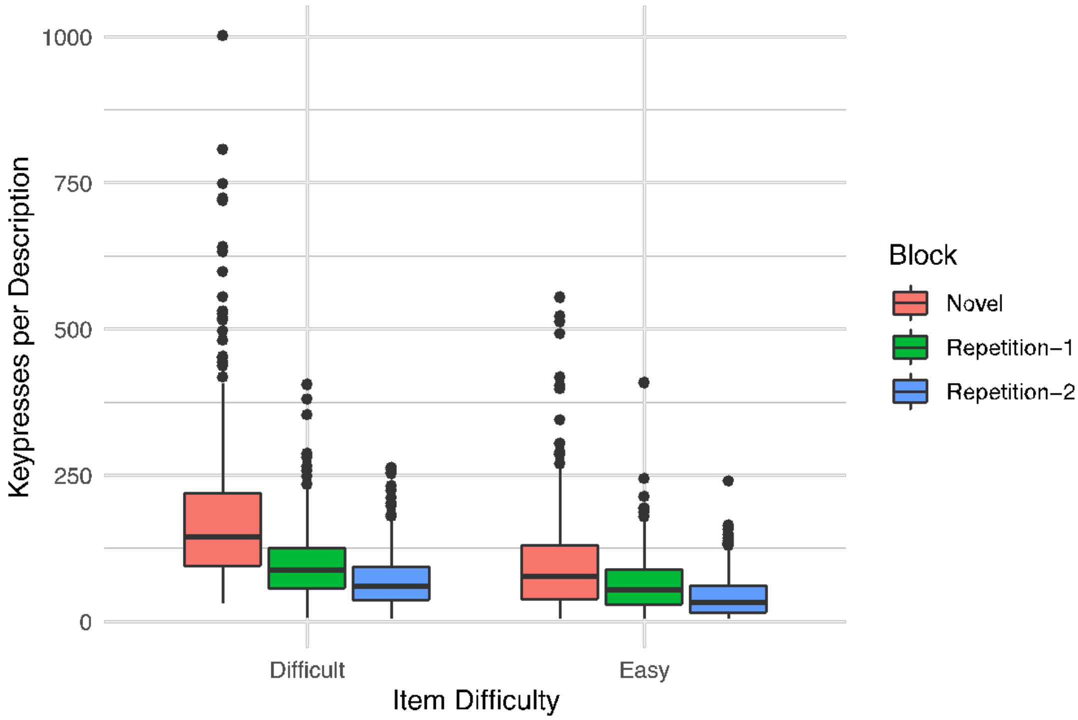


Figure 3. *Boxplot of the number of keypresses per item description, by item difficulty and block.*

To analyze these patterns, I fit the keypress count data to a generalized linear mixed effect model with a Poisson link function. The final model was maximal, having by-participant random slopes for difficulty, block and their interaction, and the by-item random slope for block. This model revealed a significant main effect of difficulty ($\beta = 0.575$, $Z = 5.328$, $p < .001$), with more typing for difficult items than for easy items. For task block, the Novel vs. Repetition-1 ($\beta = 0.234$, $Z = 6.193$, $p < .001$) and Novel vs. Repetition-2 ($\beta = 0.461$, $Z = 7.120$, $p < .001$) contrasts were both significant as well, confirming that participants typed less during repetition blocks. Finally, item difficulty interacted with the Novel vs. Repetition-1 contrast ($\beta = 0.088$, $Z = 2.40$, $p < .02$), indicating that the change in typing amount from the initial Novel block to the first Repetition block was larger for difficult items than for easy items. However, item difficulty did not interact with the Novel vs. Repetition-2 contrast (β = 0.035, Z = 0.800, p = .42).

In terms of the sheer quantity of typing, then, participants displayed patterns typical of referential communication tasks, with more typing for hard-to-describe items, and less typing as items repeated over the course of the task. Against this background, I turn to an examination of participants' more specific typing behaviors.

Backspace Proportion

As noted, the previous analysis included all keypresses related to the item descriptions. In addition to routine typing, this included keypresses associated with the deletion and retyping of some portion of a description. Because these sorts of interruptions generally involve removing one or more characters, they can be readily identified in the keylogger data through the presence of "backspace" keypresses (in our data, participants rarely made corrections by using the mouse to move the cursor to the location of text to be corrected). As a measure of the prevalence of message correction I identified each instance of a backspace keypress in the keylogger data and compared that to the total number of keypresses for each item description. Note that this total necessarily includes both the original typing as well as the keypresses associated with the correction. Figure 4 presents a plot of the proportions of description keypresses that consisted of backspaces, grouped by item difficulty and task block.

The patterns in Figure 4a suggest that manipulations of item difficulty and task block may have had a modest impact on how much participants engaged in revisions to their descriptions. On average, the proportion of reformulation backspaces was slightly greater in descriptions for difficult items (M = 0.071, SD = 0.256) than descriptions for easy items (M = 0.062, SD = 0.241). Also, the overall proportions of such backspaces decreased from the Novel block (M = 0.074, SD = 0.261) to Repetition-1 (M = 0.065, SD = 0.246) to Repetition-2 (M = 0.056, SD = 0.229). However, it is clear there was substantial variability in the proportions of reformulation backspaces within each group.

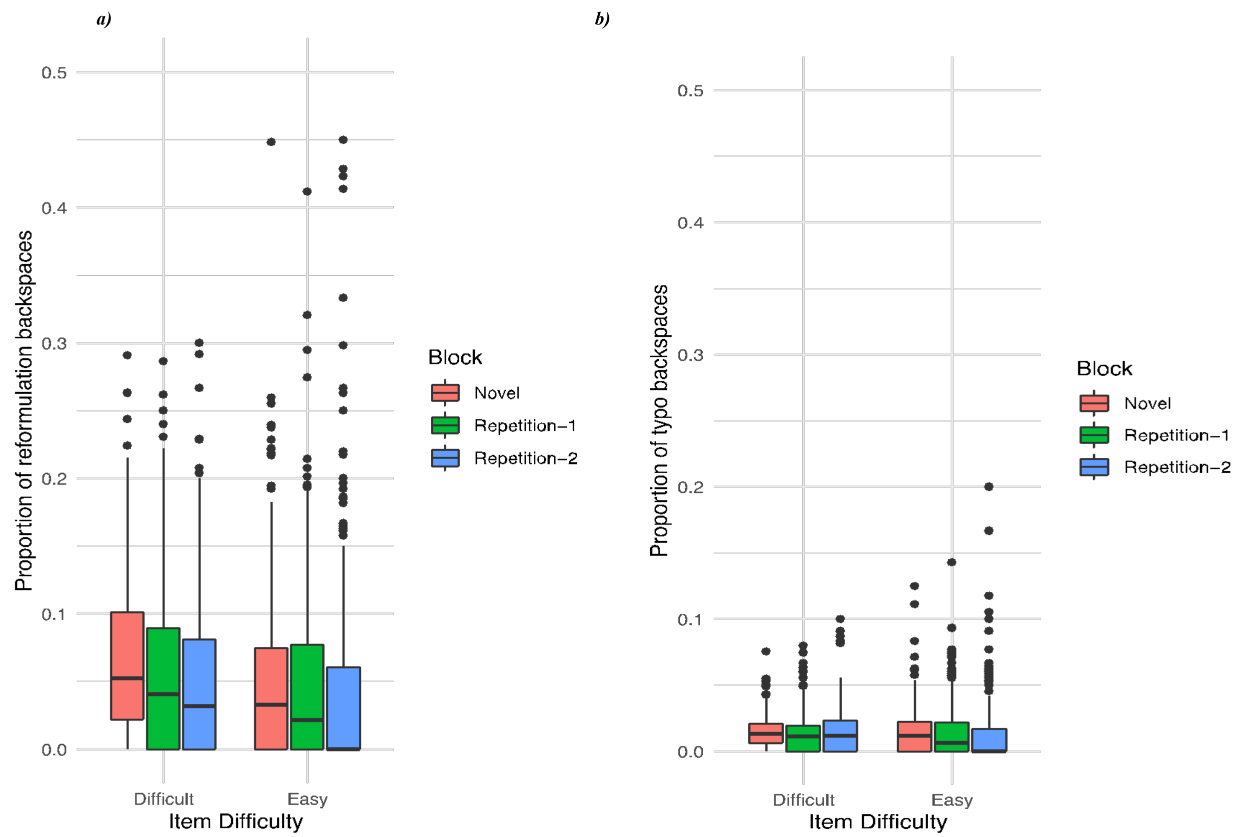


Figure 4. *Boxplots of a) the proportion of reformulation backspaces per item description and b) the proportion of typo backspaces per item description, by item difficulty and task block.*

To analyze these patterns, I fit a generalized linear mixed effect model to the binomial keypress data (backspace = 1, other keypresses = 0) using a logit link function. The final model included item difficulty, task block, and their interaction as fixed effects, along with the maximal random effects structure. This model revealed no effect of item difficulty ($\beta = 0.093$, $Z = 1.227$, $p = .22$), and no effect for the Novel vs. Repetition-1 block contrast ($\beta = 0.034$, $Z = 0.740$, $p = .46$). However, there was a significant effect for the Novel vs. Repetition-2 block contrast ($\beta =$

0.135, Z = 2.239, $p < .03$), indicating a reduction in backspacing from the first to the last block. Neither interaction between difficulty and the task block contrasts reached significance (Difficulty X Novel vs. Repetition-1: β = 0.050, Z = 0.97, p = .32; Difficulty X Novel vs. Repetition-2: β = -0.077, Z = -1.306, p = .19). In general, then, there is some evidence that the frequency of edits to the descriptions declined from the beginning to the end of the task, but this was not significantly affected by item difficulty.

Typing Speed and Pauses

Next, I consider patterns in typing speed, measured using the relative timing of each individual keypress. To compute this timing, I subtracted the timestamp associated with the *preceding* keypress from the raw timestamp of the current keypress. This provided a measure of the inter-keypress interval (IKI; Conijin, Roeser, & van Zaanen, 2019), with smaller values corresponding to faster typing. It is important to note that this IKI calculation was carried out before any of the non-description related keypresses were trimmed (as described previously). Thus, the relative timing of description-relevant keypresses was always computed with respect to adjacent keypresses, even if those keypresses were subsequently removed.

The measure of the IKI ranged initially from 0 s to 55.6 s (Mean = 0.245s, Median = 0.141s, SD = 0.60), and was positively skewed toward longer intervals. Based on inspection of the raw data, keypress interval values at or very close to 0 s nearly always represented instances in which the keylogger recorded two near-simultaneous keypresses. At the other end of the range, the longer keypress intervals in the data reflected instances in which people appeared to momentarily hesitate or stop typing. Such typing stoppages may be considered the equivalent of silent or unfilled pauses commonly found in spoken production, and thus represent another kind of typing disfluency (Wengelin, 2006).

Before conducting analyses based on these inter-keypress intervals, the data was prepared in several ways. First, I log-transformed the IKI values to reduce the extreme positive skew. Applying the log-transform required removing 184 keypresses with an inter-keypress interval of zero, or 0.1% of the data. I also identified typing pauses, represented by IKI values greater than 3SDs of each individual participant's average IKI. Across participants, this 3SD threshold ranged from 1.20s to 3.51s. Notably, these values fall around the 2-second threshold that has been frequently used to identify pauses in the literature on typing production (e.g., Wengelin, 2006). Although how to definitively identify so-called "cognitive" pauses in analyses of written production has a been a topic of some discussion (e.g., Chenu et al., 2014; Hall, Baaijen, & Galbraith, 2022), our current focus on typing behavior as a possible communicative cue guided our decision to use pause thresholds that would be detectable to potential observers, even if this risked undercounting other instances of brief (< 2 s) hesitations during message formulation. Additionally, computing pause thresholds separately for each participant allowed us to take each participant's overall typing speed into account. Using 3SD of each participant's mean IKI as a working definition of the threshold for a typing pause, I identified 1696 pauses in our data, or 1.1% of the total IKI values.

Based on these values I first report an analysis of typing speed after removing the IKI values identified as pauses. This measure reflects whether routine typing was subject to influence by our task manipulations. For clarity, I report mean typing speeds based on the untransformed data, although our plots will be in log units. Second, I report an analysis of pause behavior. For this analysis, I counted the number of inter-keypress intervals classified as pauses in each description and divided that by the total number of keypresses in that description. Analogous to the analysis of backspaces, higher pause proportions should reflect greater typing disfluency.

Overall typing speed. Figure 5 presents boxplots of the log IKI values by item difficulty and task block. These data include longer IKIs associated with pauses. On average, IKIs were larger (representing slower speeds) for difficult items (M = 0.284 s, SD = 0.81) than easy items (M = 0.257 s, SD = 0.66). Also, IKIs decreased from the Novel task block (M = 0.274 s, SD = 0.76) to Repetition-1 (M = 0.219, SE = 0.36) to Repetition-2 (M= 0.207, SD = 0.60).

To analyze overall typing speed, I fit a linear mixed effect model with an identity link function to the log of the inter-keypress intervals. The final model included fixed effects for item difficulty, task block, and their interaction, by-participant and by-item random intercepts, and the by-participant random slopes for difficulty and task block and their interaction. This model revealed a significant main effect of item difficulty ($\beta = 0.032$, $t(62.3) = 2.86$, $p < .01$) and significant effects for both the Novel vs. Repetition-1 ($\beta = 0.037$, $t(22.8) = 7.11$, $p < .001$) and the Novel vs. Repetition-2 ($\beta = 0.049$, $t(22.8) = 7.83$, $p < .001$) block contrasts. In addition, there was a significant interaction between difficulty and the Novel vs. Repetition-2 contrast ($\beta = -0.017$, $t(24.3) = -2.97$, $p < .01$). Compared to the Novel block, the decline in IKI values in the Repetition-2 block was greater for easy items than for difficult items.

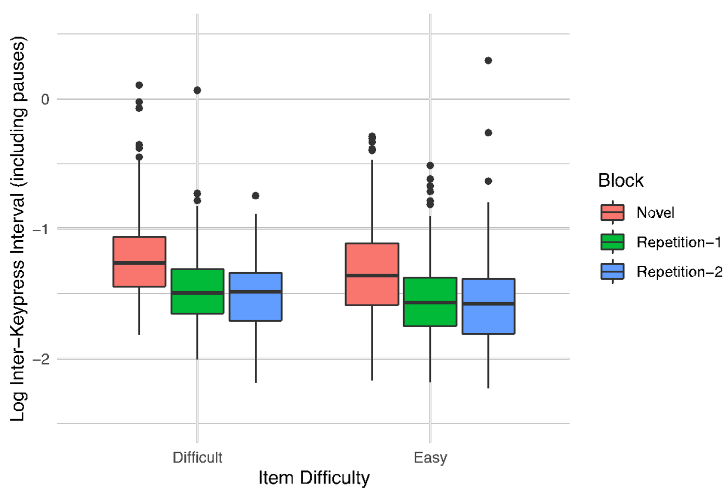


Figure 5. *Boxplot of the log-transformed inter-keypress intervals (IKI) including pauses, by item difficulty and task block.*

Routine typing speed. Figure 6 presents a boxplot of log IKI values after removing IKIs representing pauses, by item difficulty and task block. Again, average IKIs were larger for difficult items (M = 0.205 s, SD = 0.22) than easy items (M = 0.200 s, SD = 0.21). Also, IKIs showed a decrease from the Novel task block (M = 0.211 s, SD = 0.23) to Repetition-1 (M = 0.198, SD = 0.21) to Repetition-2 (M= 0.191, SD = 0.20), representing faster typing over time.

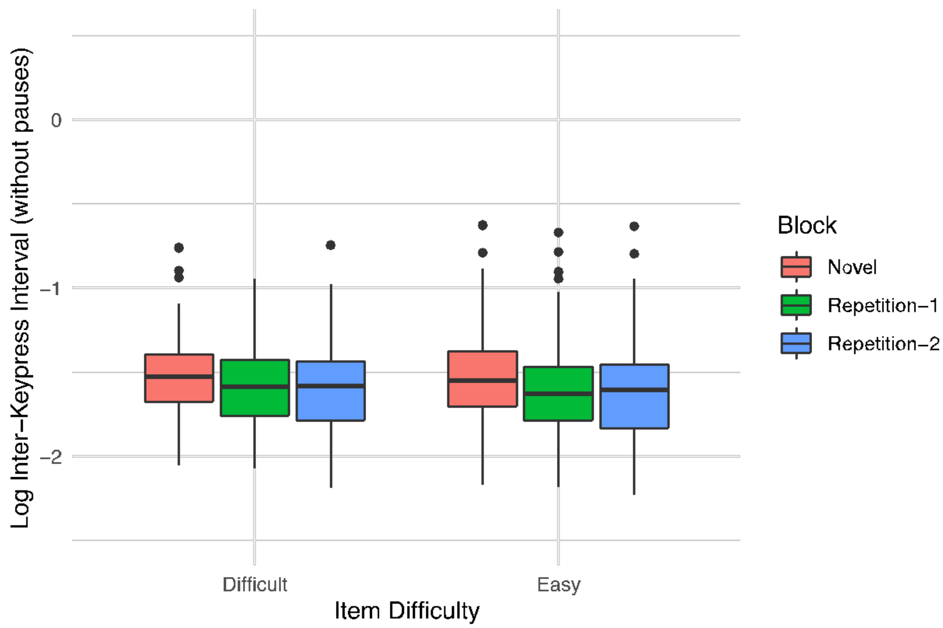


Figure 6. *Boxplot of the log-transformed inter-keypress intervals (IKI) not including pauses, by item difficulty and task block.*

To analyze routine typing speed, I fit a linear mixed effect model with an identity link function to the log of the inter-keypress intervals without the pauses. The final model included fixed effects for item difficulty, task block, and their interaction, by-participant and by-item random intercepts, and the by-participant random slopes for block and the difficulty by block interaction. This model revealed an effect of item difficulty (β = 0.023, $t(55.7) = 2.38$, $p < .03$) and significant effects for both the Novel vs. Repetition-1 (β = 0.024, $t(23.0) = 5.17$, $p < .001$)

and the Novel vs. Repetition-2 (β = 0.035, $t(22.9) = 6.75$, $p < .001$) block contrasts. In addition, there was a significant interaction between difficulty and both the Novel vs. Repetition-1 contrast (β = -0.012, $t(25.6) = -2.28$, $p < .04$) and the Novel vs. Repetition-2 contrast (β = -0.026, $t(23.0) = -3.51$, $p < .002$). The change in IKI values across both block contrasts was larger for easy items compared to difficult items, consistent with increases in typing speed due to both ease of description and task block.

Pause proportions. Recall that here typing pauses are classified as all inter-keypress intervals longer than 3SDs above each participant's mean log IKI. In doing so, I made no attempt to distinguish between different pause locations – e.g., whether the pause occurred between words, between sentences, or mid-word (each of which could reflect different sorts of planning difficulties; Spelman Miller, 2006). Each keypress interval identified in this way was coded as either a pause or normal typing (pause = 1; other = 0). From this coding, I computed the proportion of keypress intervals in each description that involved pauses.

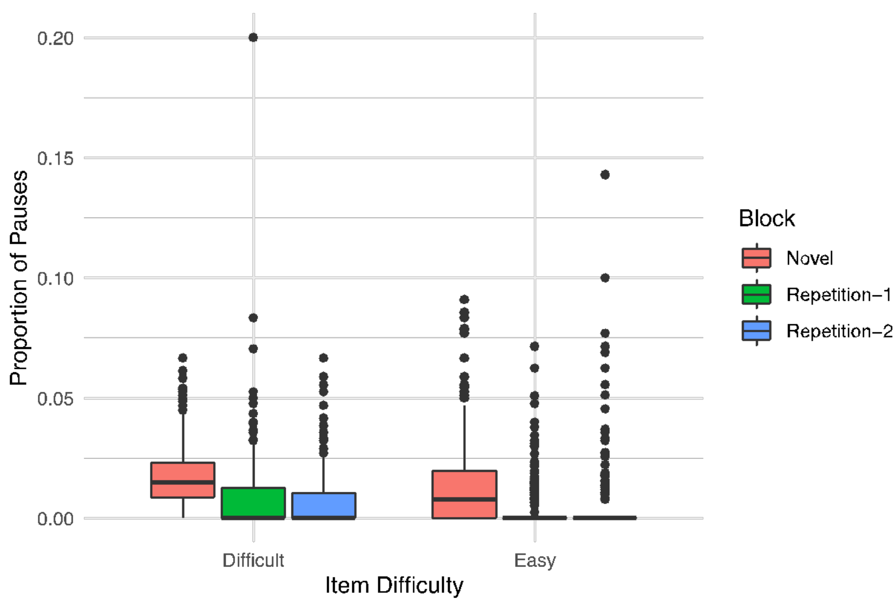


Figure 7. *Boxplots of the proportions of pauses per item description, by item difficulty and task block.*

Figure 7 shows the average proportions of pauses per description, by item difficulty and task block. Overall, the proportion of substantial pauses in the item descriptions was low, at around 1.1% of keypress intervals. Even so, there was a greater proportion of pauses in descriptions of difficult items (M = 0.012, SD = 0.11) than easy items (M = 0.009, SD = 0.09), and proportionally more pauses in the Novel task block (M = 0.015, SD = 0.12) than in Repetition-1 (M = 0.006, SD = 0.08) or Repetition-2 (M = 0.005, SD = 0.07) blocks.

The binomial pause data were fit to a generalized linear mixed effect model with a logit link function. The final model included item difficulty, task block, and their interaction as fixed effects, by-participant and by-item random intercepts, and the by-participant random slope for task block only. This model revealed a significant effect of item difficulty (β = 0.354, Z = 3.781, $p < .001$) as well as significant effects for the Novel vs. Repetition-1 contrast (β = 0.441, Z = 9.827, $p < .001$) and the Novel vs. Repetition-2 contrast (β = 0.598, Z = 8.888, $p < .001$). Neither difficulty by block contrast interaction was significant (Difficulty X Novel vs. Repetition-1: B = -0.016, Z = -0.229, $p = .82$; Difficulty X Novel vs. Repetition-2: β = -0.134, Z = -1.398, $p = .16$). In general, then, these results confirm that participants produced proportionally more pauses in their descriptions for difficult items and in the Novel block compared to the rest of the task.

Discussion

In Experiment 1, I examined the effects of two different sources of "uncertainty" on the real-time typing behaviors of individuals carrying out a written picture description task: 1) uncertainty arising from the challenge of composing descriptions for ambiguous, hard-to-describe images, and 2) uncertainty arising from simply being relatively unfamiliar with the images and how to describe them. Our results show that both sources of uncertainty had effects on participants' typing. For item difficulty, when describing relatively ambiguous images participants engaged in not just more typing but their typing was also slower with more pauses. For item familiarity, compared to their initial descriptions, participants' subsequent descriptions of the same items in Repetition blocks involved less typing and faster typing with proportionally fewer pauses, as well as a smaller proportion of typing revisions. Finally, these types of uncertainty interacted for the measures of keypress count and typing speed, with a decline in typing amount and an increase in typing speed over the course of the task that was greater for easier-to-describe items.

These patterns confirm that typist uncertainty can be reflected in the dynamics of *how* individuals produce written messages, in ways that likely run parallel to the content of *what* they type. When people are unsure about what they are trying to write, this lack of certainty bleeds into their typing in multiple ways, resulting in slower, more disfluent production. I believe this is directly equivalent to work on spoken language production showing that speaker uncertainty is frequently associated with halting, disfluent speech (Brennan & Williams, 1995; Krahmer & Swerts, 2005). Just as some paralinguistic features of spoken utterances (like speech rates or filled pauses) reflect underlying challenges in message formulation, the speed and fluency of typing can directly reflect similar challenges.

Given that typing behaviors appear to be reliably impacted by uncertainty in a way that leads to observable patterns, it is now possible to ask how viewers interpret typing with frequent pauses and errors. In Experiment 2, I sought to answer this question by using video clips of fluent and disfluent typing collected as part of Experiment 1 and presenting these clips to naïve viewers. I was interested in the kinds of inferences these viewers would make about the typists based on patterns of visible fluency.

Experiment 2

Experiment 1 confirmed that message uncertainty can be reflected in the output visible onscreen, providing insight into the kinds of typing behaviors that would be most impacted by uncertainty. Given these findings, the next relevant question was whether *observers* of these behaviors would draw relevant inferences about user certainty. In Experiment 2, I explored this question by presenting clips from the screen recordings collected as part of Experiment 1 to a new set of naïve viewers, who were asked to make judgments about the level of confidence, familiarity, task effort, and task difficulty experienced by the original typists. I also asked them two open ended questions to allow them further explanation of what kinds of judgements they made about the typists and what aspects of the videos led them to make such judgments. I was interested in how viewers watching a screen-recording of someone's real-time typing behavior would interpret these behaviors, and whether typing fluency and speed, in particular, would shape viewers' perception of typists' confidence and task familiarity.

Method

Participants

Four hundred and eight participants were recruited for this study from Amazon Mechanical Turk via CloudResearch (https://www.cloudresearch.com). Using the CloudResearch screening options, we restricted participation to adult native English speakers located within the United States, Canada, Australia, the Caribbean, and the United Kingdom. To further ensure participants were native English speakers, our survey included questions about language experience. All participants self-reported as native speakers of English. One participant did not complete the study, and so their data was excluded.

The remaining sample consisted of 190 women, 215 men, and 2 individuals who chose not to report their gender. The majority of participants were between the ages of 26 and 45 (n=287), with seven participants between the ages of 18 and 25 and 113 participants age 46 and older. Of this sample, 277 of 407 participants reported having experience with online collaborations, with the largest number (n=214) indicating that this involved “document writing,” followed by “creating presentations” (n=137), “video editing” (n=34), and “other format” (n=80); participants could choose more than one of these options. Participants reported spending 7.65 hours online per day on average (SD = 3.6; min=1; max=20).

Materials

Video materials for this study were selected from the computer screen captures collected as part of Experiment 1, which recorded the key-by-key typing of participants as they completed an image description task. From these screen recordings, I created separate video clips of typed descriptions for single items. Each selected clip began when the participant first visibly clicked the text description box to start typing and ended when they moved the mouse to click the arrow at the bottom of the screen to move to the next item.

As described previously, Experiment 1 employed six image sets: tangrams, facial caricatures, and objects, with eight images in each category selected to be “Easy” and eight images selected to be “Difficult” to describe (48 unique images total). For the current study, we focused our selection of video clips on the 24 items (eight from each category) that were presented multiple times, as this gave more source descriptions from which to choose.

For each item, I selected multiple clips from the video captures, choosing clips that contained relatively fluent or disfluent typing. Specifically, each video clip was classified as either “Fast” or “Slow” in typing speed which was measured by the average keypress speed with

Fast items averaging .25 seconds between keypresses and slow items averaging .58. The second dimension was based on error frequency, with either “Frequent” or “Infrequent” errors measured as proportion of backspace keypresses out of total keypresses for each description, a proxy for overall degree of error correction. “Frequent” error items included an average of 16 backspaces for every 100 characters typed, while “Infrequent” and an average of 5 for every hundred. These classifications were based on the measurements collected as part of Experiment 1. Together, they allowed me to examine whether certain typing behaviors would have more influence on viewers’ inferences about the typist, as well how combinations of features might shift the viewer’s interpretation. To reduce the degree to which participants would be influenced by the content of what was being typed, I attempted to select clips in which the description for a given item was similar in length across examples. I also tried as much as possible to ensure that the content of each clip mentioned similar aspects of the image. For instance, descriptions of ambiguous tangrams ranged from holistic interpretations (e.g. “a man running”) to analytic shape-based formulations (e.g. “the triangle on top connected to a square”). In these cases, I selected clips for each disfluency category that included similar description types. Similarly, for descriptions of facial caricatures, I selected clips in which the participants referenced the same set of features. These considerations resulted in a final set of 45 unique clips: 7 Slow speed + Frequent backspacing, 11 Fast speed + Frequent backspacing, 11 Slow speed + Infrequent backspacing, and 14 Fast speed + Infrequent backspacing. The Fast+Infrequent item category can be considered maximally “fluent” (i.e., the typing was relatively quick with minimal pauses, error corrections, or editing), while the other three combinations were each “disfluent” in different ways (either slow typing, frequent backspacing, or both).

Task Questionnaire

These video clips were embedded into a Qualtrics questionnaire, which presented each respondent with a single randomly chosen video out of the full set. Following the video, an initial block of five items asked about the participants' perceptions of the typist. Participants responded to each of these questions on a 0-100 continuous sliding scale. These questions were:

1. Confidence: *How confident do you think this person is about how to describe the image?*

   [0 = Very unconfident; 100 = Very confident]

2. Familiarity: *How familiar is this person with the image they're describing?*

   [0 = Very unfamiliar; 100 = Very familiar]

3. Effort: *How much effort do you think this person is putting into their description?*

   [0 = Little effort; 100 = Significant effort]

4. Difficulty: *How much difficulty is this person having with their description?*

   [0 = Little difficulty; 100 = Significant difficulty]

5. Described before*: How likely is it that this person has described this image before?*

   [0 = Very unlikely; 100 = Very likely]

A second block of items prompted participants to explain their answers to the previous questions. Here, they were asked two open-ended questions. The first question asked: "*What specific aspects of the video informed your answers to the previous questions? That is, please briefly describe features of the video that most clearly allowed you to judge the respondent's familiarity or confidence in what they were typing*." The second question asked: "*Do you feel like you were able to get a clear sense of what the typist was thinking as they carried out this task?*

Why or why not?" For these questions, we were interested in how often participants would spontaneously mention aspects of the typing such as speed or disfluency.

Finally, a third block of six items asked for estimates of specific typing behaviors present in the previously-viewed clip (which participants were prevented from re-watching). With these items, we probed how salient different typing behaviors were to viewers and whether disfluent typing was generally seen as containing more of certain behaviors than fluent typing. Specifically, we asked for categorical estimates of 1) hesitation frequency, 2) frequency of simple typos, 3) frequency larger typing errors, 4) how often the typist appeared to change their mind about what to type, 5) overall typing speed, and 6) relative variability in typing speed.

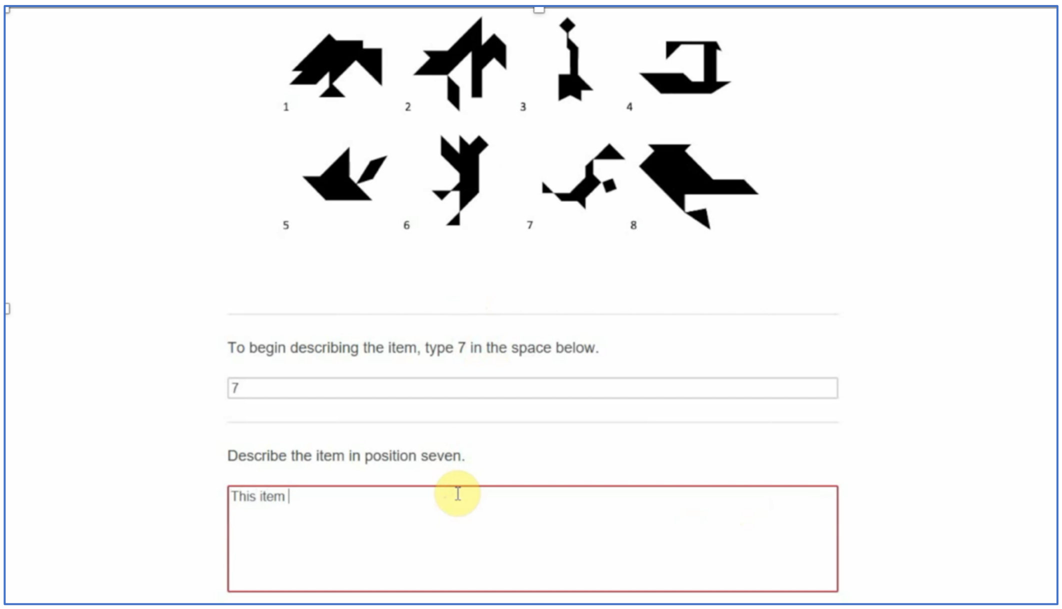
All participants responded to the complete set of questions in the same fixed order. The Qualtrics survey ended with several items asking about the amount of time the participants spend online in general and their experiences with various types of online collaboration.

Procedure

After accessing the survey through the Qualtrics link, participants read instructions informing them that they were being asked to carefully watch a short video in order to answer some questions about it. They were told that the video was a screen capture taken from a previous image description task, and that they would watch as someone typed in a description of a visual image, which would also be visible in the screen recording. Figure 8 presents a sample static screenshot of what participants saw. The video clip was embedded directly within the Qualtrics survey, and participants were asked to carefully watch the video, and were not allowed to click through to the next page until they remained on that screen long enough to view the entire video at least once. On the next screen, participants were presented with the first set of questions and were given the opportunity to watch (and re-watch, if necessary) the video again while responding. Participants then answered the remaining questions without viewing the clip again. Importantly, each participant was presented with only a single video clip and answered questions specifically about that one clip.

Figure 8

Screenshot from video capture of previous image description task.



Design

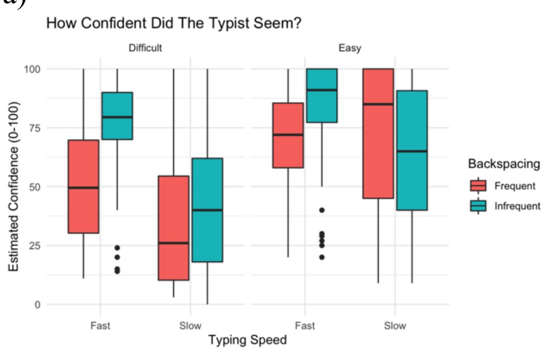
Each participant viewed and answered questions about one video clip showing a single item description. Across all 400 participants, from 4 to 16 (M=9.7) participants ended up answering questions about one of each of the 45 clips. The clips varied along four fully crossed independent variables: Item Category (tangrams, objects, faces), Item Difficulty (easy, hard), Typing Speed (fast, slow), and Backspace Frequency (frequent, infrequent). In the present analyses, I collapse across Item Category, considering instead only the effect of item difficulty, typing speed, and backspace frequency on participants' judgments.

Results

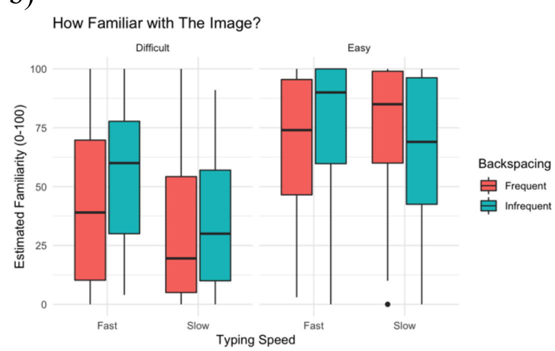
Perceptions of the Typist

To start, I will first focus on the results of the first set of questions asking about participants' perceptions of typist's confidence, familiarity with the item, degree of effort, difficulty with the description, and likelihood of having described this item before. Participants gave their responses to each of these questions on a 0-100 continuous scale. Figures 9a-e present boxplots of the ratings for these five questions organized by the difficulty of the item being described (easy, difficult), visible typing speed (fast, slow), and visible backspace frequency (frequent, infrequent). To analyze the data for each question, I used *lmer* in R to fit separate linear mixed effect models with an identity link function to standardized values of the 0-100 scale ratings. Here I report the raw means for clarity.

a)

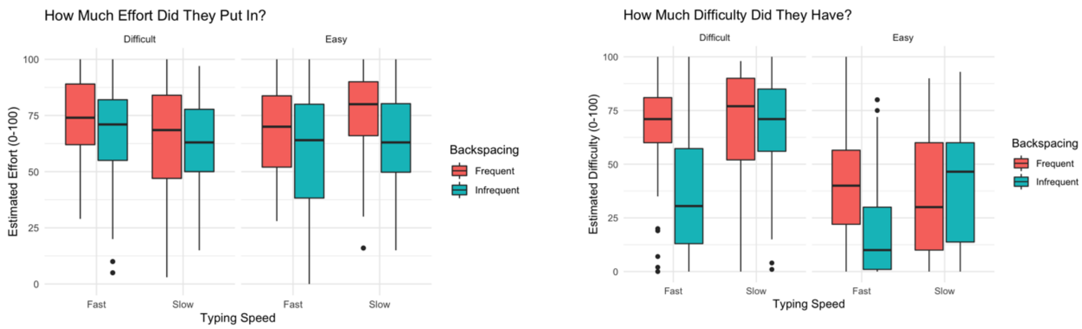


b)



c)

d)



e)

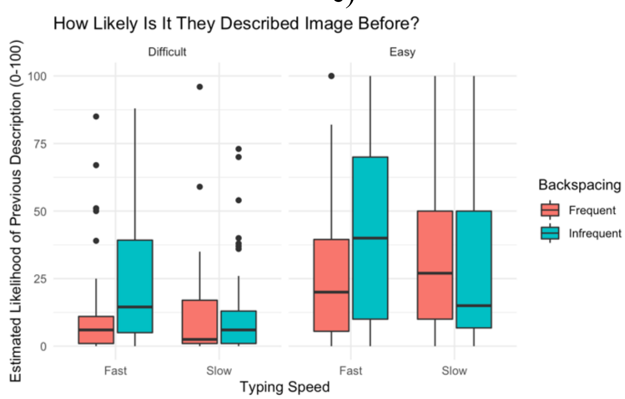


Figure 9. *Boxplots of participants ratings from Experiment 2 for a) typist confidence, b) typist's familiarity with the image, c) typist's degree of effort, d) typist's difficulty with the description, and e) the likelihood of the typist having described the image before, grouped by item difficulty, typing speed, and backspace frequency.*

Confidence

Figure 9a presents a boxplot of the judgments of typist confidence. Typists producing descriptions for easy items were judged as more confident (M = 71.99, SD = 26.5) than those producing descriptions for difficult items (M = 53.78, SD = 29.0), β = -0.61, $t(11.6) = 3.41$, $p <$ .01. Descriptions with fast typing were also judged as being produced by more confident typists (M = 71.12, SD = 25.1) than descriptions with slow typing (M = 51.69, SD = 30.7), β = -0.65, $t(20.6) = 6.73$, $p < .001$, and descriptions with infrequent backspacing were judged as being produced by more confident typists (M = 65.98, SD = 28.9) than descriptions with frequent backspacing (*M* = 58.16, SD = 29.1), β = -0.36, $t(21.6) = 3.70$, $p < .002$. There was also a significant speed X backspacing frequency interaction, β = 0.53, $t(21.3) = 2.56$, $p < .02$. For descriptions with slower typing, backspace frequency did not affect judgments of typists' confidence, while for descriptions with faster typing, frequent backspacing was associated with perceptions of less confidence.

Familiarity

Figure 9b presents a boxplot of the judgments of how familiar the typists seemed with what they were describing. Here, descriptions for easy items were judged as involving greater familiarity (M = 70.38, SD = 30.5) than descriptions for difficult items (M = 42.19, SD = 30.8), β = -0.84, $t(11.8) = 3.76$, $p < .003$. There was also an effect of typing speed, with fast typing associated with judgments of greater familiarity (M = 60.98, SD = 32.8) than slower typing (M = 49.95, SD = 34.0), β = -0.33, $t(20.7) = 3.60$, $p < .002$. Finally, there was a significant effect of backspace frequency, with less frequent backspacing prompting judgments of greater item familiarity (M = 58.2, SD = 33.1) compared to clips with more frequent backspacing (M = 53.3, SD = 34.5), β = -0.22, $t(21.8) = 2.33$, $p < .03$. None of the interactions between factors was significant.

Effort

Figure 9c presents a boxplot of the judgments of the degree of typist effort. Here, descriptions with more frequent backspacing were seen as involving more effort (M = 69.48, SD = 20.7) than descriptions with infrequent backspacing (M = 62.68, SD = 23.9), β = 0.29, $t(23.3) = 2.36$, $p < .03$. Additionally, there was a marginal item difficulty X backspacing interaction, β = -0.48, $t(21.9) = 1.98$, $p = .06$. For clips that involved descriptions of easy items, slower typing was perceived as involving more effort (M = 66.8, SD = 23.3) than faster typing (M = 62.7, SD =

24.9), while for clips that involved descriptions of difficult items this pattern reversed, with fast typing associated with perceptions of greater effort (M = 69.4, SD = 20.8) than slower typing (M = 62.6, SD = 22.1).

Difficulty

Figure 9d presents a boxplot of the judgments of how much difficulty typists had in formulating their descriptions. Not surprisingly, descriptions for easy items were judged as involving less difficulty (M = 32.29, SD = 26.8) than descriptions for difficult items (M = 56.94, SD = 28.7), $\beta = 0.80$, $t(11.7) = 4.25$, $p < .002$. Descriptions with fast typing were also judged as less difficult for typists (M = 38.10, SD = 29.2) than descriptions with slower typing (M = 53.43, SD = 29.7), $\beta = 0.51$, $t(18.9) = 5.23$, $p < .001$, and descriptions with infrequent backspacing were judged as involving less difficulty (M = 40.23, SD = 30.4) than descriptions with frequent backspacing (M = 51.27, SD = 29.1), $\beta = 0.43$, t(19.5) = 4.37, $p < .001$. There was also a significant speed X backspacing frequency interaction, $\beta = -0.679$, t(19.5) = 3.26, $p < .005$. For descriptions with slower typing, perceptions of description difficulty did not differ across clips with frequent (M = 50.6, SD = 32.5) and infrequent (M = 54.7, SD = 28.3) backspacing, while for descriptions with faster typing, infrequent backspacing was associated with perceptions of less difficulty (M = 21.2, SD = 26.1) than frequent backspacing (M = 51.6, SD = 27.3).

Described Before

Figure 9e presents a boxplot of participants' judgments of how likely it was that the typist had described the image before. Participants judged it more likely that participants had previously described easy items (M = 32.0, SD = 30.0) than difficult items (M = 15.7, SD = 20.7), $\beta = -0.63$, $t(11.6) = 2.63$, $p < .03$. Judgments of previous experience were also seen as marginally greater for descriptions with faster typing (M = 27.0, SD = 28.3) than descriptions with slower typing (M = 19.6, SD = 24.5), β = -0.22, $t(18.0) = 2.09$, $p = .051$, and greater for descriptions with infrequent backspacing (M = 26.0, SD = 29.3) than descriptions with frequent backspacing (M=20.6, SD = 24.5), β = -0.27, t(18.8) = 2.62, $p < .02$. These effects were qualified by a speed X backspacing frequency interaction, β = 0.471, t(18.3) = 2.13, $p < .05$. Participants perceived the greatest likelihood of a previous description for clips involving both fast typing and infrequent backspacing (M = 33.0, SD = 30.9) compared to clips with disfluent typing (infrequent backspacing, slow speed: M = 18.12, SD = 23.4; frequent backspacing, fast speed: M = 19.5, SD = 23.5; frequent backspacing, slow speed: M = 22.7, SD = 26.4).

Spontaneous Mentions of Typing

After making these metacognitive judgments, participants were asked a pair of open-ended questions about the aspects of the video clip that influenced their responses and whether they felt like they knew what the typist was thinking. First, for the question "*What specific aspects of the video informed your answers to the previous questions?*" 63% of participants spontaneously mentioned, in some manner, specific features of the visible typing behavior. Specifically, 26% mentioned typing speed, 31% mentioned hesitations or pauses, 26% mentioned editing or changes to the typing (and 17% mentioned more than one of these), while only 37% of participants gave nonspecific responses that simply mentioned ease or difficulty or gave other sorts of answers. Thus, participants frequently indicated that the nature of the typing informed their judgments of the typist's epistemic state.

For the question "*Do you feel like you were able to get a clear sense of what the typist was thinking as they carried out this task? Why or why not?*" 82% of participants responded in the affirmative, giving explanations such as "watching them type gave me a fairly good indication of how clear they were with their thought process" and "because it seemed deliberate the way they typed." For the 18% of participants who responded "no" to this question, their responses sometimes mentioned other explanations for the observed typing behaviors, such as "the delay could be due to being a mediocre typist" or they expressed doubt that it was possible to infer anything about the typist at all—e.g., "It was just a video of some typing. No emotions showed and it all seemed rather mundane." Generally, though, most participants' subjective intuition was that they could discern something about the typist's mental state from observing how the typist went about the item description task.

Discussion

Taken together, the results from Experiment 2 demonstrate that observers are sensitive to variation in real-time typing patterns and can reliably make judgments about a person's degree of confidence based in part on visible evidence for the timing and fluency of written message production. In particular, both typing speed and typing correction (backspacing) contributed to judgments of typist confidence and difficulty, with slower typing and frequent backspacing associated with judgments of lower confidence and greater difficulty with the task. Frequent backspacing also led raters to judge the typist as putting forth greater effort, while slower typing was associated with the belief that the typist was less familiar with what they were typing. Overall, judgments about the typist's epistemic state were influenced by characteristics of the "delivery" of the image descriptions. This is consistent with the idea that when access to other cues is limited, as they often are in online collaborative contexts, people may be able to adapt by attending to more subtle sources of information to draw inferences about their partner and how their partner is managing the task (Lea & Spears, 1992; Walther, 1992).

Additional confirmation that participants spontaneously attended to the visible typing behaviors comes from their open-ended responses, which frequently mentioned speed, pauses, and character or word editing as reasons for determining when the typist was confident or experiencing difficulty. Notably, these responses were elicited before the task called specific attention to anything about the manner of typing. Indeed, when participants didn't mention typing they often mentioned other aspects of the image descriptions, such as whether the typist was able to succinctly identify the image in question. Of course, watching an isolated clip of a single item description did not give participants a lot of information to draw upon when making these judgments, which may have increased attention to typing behavior. Even so, it is clear that the dynamics of *how* the original typists formulated their descriptions were salient to participants.

Further evidence of this salience comes from responses to the final set of questions, which asked participants to provide retrospective categorical estimates of the typing behaviors observed previously. Participants consistently demonstrated sensitivity to both the speed and fluency of typing in the original clips. In particular, judgments of the prevalence of hesitations, typos, larger errors, and description reformulations each exhibited a typing speed by backspace frequency interaction, which reflects a general tendency to view these behaviors as occurring more often in clips with slower typing and frequent backspacing (and less often for clips with faster typing and infrequent backspacing). Similarly, estimates of typing speed appeared to be influenced not just by the actual speed of typing but also by the frequency of backspacing. The fact that participants, despite being presented with only a single clip to judge, were able to reliably distinguish fast from slow typing (not to mention more frequent from less frequent typing disfluencies) suggests that they may have been relying on implicit expectations regarding the kinds of behaviors that characterize fast, fluent typing, even in an unfamiliar task context. In text-based communicative contexts that allow users to observe messages being formulated character-by-character, these kinds of expectations are likely an important prerequisite for inferences about a partner's epistemic state.

Experiment 3

In Experiments 1 and 2, I established that typing patterns can not only systematically vary based on typist uncertainty, but that viewers are able to derive meaningful inferences about the typist from viewing isolated clips of these typing behaviors. Here in Experiments 3a and 3b, I expand on these findings to address the following questions: When given the opportunity to observe a partner's typing while *engaged* in a collaborative task, do individuals form the same sorts of judgments about their partner as they do when viewing text disfluencies in an isolated video (as illustrated in Experiment 2)? And what might be the result of inferring that a collaborative partner is uncertain about their work?

My expectations for both questions are informed by the literature on collaborative work that has focused on groups with members who are not equally equipped for the task. Unlike the current experiments, much of the research examining this question has involved asynchronous contexts. That is, in these studies judgments of another's contributions stem not from real time cue-based inferences, but rather from observations of the *content* of another's work. For instance, Storch (2004) analyzed verbal interactions[2] between pairs during an editing task (finding and resolving grammatical and lexical errors) in terms of equality (the amount of each member's

[2] An additional important distinction is that much of the work on collaborative writing allows collaborators to speak with each other in person to plan their writing strategy. While this will not apply to the present study, there is information to be gained about how people respond to working with uncertain group members based on what groups talk about during these sessions.

contributions) and mutuality (the degree to which each member engaged with their partner's contributions). Pairs in which one member was dominant in their contributions but did *not* accept the other member's feedback were categorized as low in both equality and mutuality. Meanwhile, pairs in which one member was an expert and the other a novice tended to fall into the pattern of low equality (because one member contributed the most) and high mutuality (because the expert engaged with the novice's work in an instruction role, and the novice engaged with the expert's work by accepting their changes and feedback). This suggests that, for groups in which members display skill inequality based on the content of their work so far, stronger group members may naturally adjust their level of contribution to overcome their partner's areas of weakness. In Storch (2004), this was shown not only in stronger members' contribution levels, but also in the fact that they provided unsolicited instruction towards their weaker group member's contributions.

Li (2013) similarly analyzed the discussion patterns of groups of individuals contributing to Wiki pages, this time in asynchronous writing rather than verbal conversation. In her analysis, she identified three common collaboration patterns based on each contributor's rate of initiating discussion and responding. These three interaction dynamics were referred to as "contributing/mutually supportive," "authoritative/response," and "dominant/withdrawn." That is, groups were either well-distributed in their contributions across members and mutually responsive to each other's work (contributing/mutually supportive), or they featured a subset of domineering members who contributed the most while other members who either engaged with the dominant member's work (authoritative/response) or accepted another's changes passively rather than discussing (dominant/withdrawn). The fact that Li (2013), much like Storch (2004), could *not* identify a passive/passive group could be taken to suggest that individuals adapt to other members that are passive or uncertain by becoming a more dominant contributor. Daiute and Dalton (1993) found a similar pattern in a study where they paired novice students with ‘expert’ students of the same age in a collaborative writing task. They found that these expert/novice pairings yielded higher amounts of planning conversation than groups matched on knowledge. This indicates that group members respond to inequalities in task knowledge not only by providing a higher level of attention or instruction to their weaker group member’s work, but that when possible, they also adjust their coordination strategy before their work begins. The combination of each of these findings ultimately suggests that when a person is paired with someone who provides evidence of being less knowledgeable, they commonly respond by coordinating with them more explicitly and paying a higher degree of attention to that group member’s work, including providing them with feedback and instruction.

The current study expands on these ideas by testing whether these adaptations that occur between novice and expert contributors also occur when cues present in their partner’s *delivery* provide indicators of certainty, independent of the content contributed. This study will also provide insight into whether such patterns occur based on real-time interactions, which has not been tested within a written collaboration context. To test this, I asked participants to engage in a sentence by sentence editing task in a shared document alongside another (sham) “participant.” In actuality, the contributions of this second participant were conveyed through a series of prerecorded keystroke files, which I used to make it seem as if this partner was also engaged in the collaborative editing task. Importantly, these recordings showed the partner revising critical sentences (which were the same across conditions) in a manner that was either fluent (relatively fast with few pauses or errors) or disfluent. Although the participant worked on revising other sentences within the document at the same time, the task was constructed to ensure that participants had the opportunity to notice the partner's editing activity.

In Experiment, 3a, I sought to test the influence of the partner's typing fluency by examining the extent to which participants would subsequently choose to edit their partner's revisions. Thus, after working through the document once alongside this "partner," participants were then presented with their partner's revised sentences and asked whether they would like to make any additional improvements. If participants edit their partner's contributions more substantially when their typing is disfluent, this suggests that not only is onscreen typing an informative window into a typist's task confidence (as illustrated in Experiment 2), but that it can have a broader impact on the collaborative process as a whole. This design allowed me to test whether disfluent typing would not only prompt further scrutiny of work with content that is obviously insufficient (i.e., poorly edited sentences), but also prompt participants to revise content that does not require further revision. Conversely, I was also interested in whether fluent typing would lead to a higher level of acceptance of the partner's edits, even for work which might require further editing. To do so, I examine how often participants chose to edit their partner's work or accept their partner's edits of their work in both the fluent and disfluent conditions. Based on the discussed patterns observed in work groups, both online and in person, I expect that participants will be less willing to accept their partner's contributions and place more scrutiny on their partner's changes when their prior typing is disfluent.

Experiment 3a

Method

Participants

Thirty-eight Northwestern undergraduates aged 18 to 22 participated in this study for partial course credit. All were native English speakers. The participants carried out their completion of the study remotely.

Materials

For the editing task, I constructed a series of paragraphs based on items taken from an online SAT preparation website (cracksat.net), which I modified to ensure that each line of every paragraph would require some degree of editing. In total, I modified four paragraphs, each ten sentences in length. The types of edits required of these sentences were not exclusively grammatical, but rather stylistic and structural in nature. For instance, a sentence might include excessive repetition, passive voice, or confusing word order. I also included minor grammatical errors, such as misuses of punctuation, in a portion of sentences. Some examples of the sentences are shown in Table 1. Each of the original sentences was modified to include a similar number of errors and/or poor writing choices. Specifically, starting with the base sentence from the SAT preparation website, I added two to four "errors" by inserting unnecessary words, introducing tense inconsistencies, and rearranging segments of the sentence to make the meaning less clear.

Table 1. *Original Sentences (to be Revised by False Participant)*

Globalization, or the integration of cultures across nations, can be to some, a prominent yet controversial topic. (Line 1)
Some defend globalization for its benefits greater creativity and appreciation of heritage. (Line 2)
Now, many nations gets to enjoy booming domestic film markets that can even compare with or surpass the United States in production and popularity. (Line 5)
Certainly, many individuals prefer films that reflect their own cultural identities, but trends indicate rising popular interest, even in Hollywood, in multicultural and cross-cultural movies. (Line 10)

For each paragraph, half of the original sentences were assigned to the participant for the editing task, and half were assigned to be edited by the "partner." For the critical sentences assigned to the partner, I created a second set of pre-scripted "revisions" designed to be the partner's contributions to the editing task. To compare the impact that visible typing disfluency might have over a broad range of contribution "quality," these revisions were constructed to be either *correct* or *insufficient*. Correct revisions were intended to be generally acceptable edits of the target sentences and were based in part on the original crackset.net suggested corrections to the source sentences. In particular, the additional errors that I introduced to the original sentences were fixed in these correct revisions. For the insufficient revisions, I corrected at least one included error but inaccurately corrected the remaining errors, either by shifting the word order in a way that did not improve the sentence structure, changing tense errors to an additional incorrect or inconsistent tense, or the substitution of unnecessary words with different yet equally unneeded words. These insufficient revisions are intended to represent stylistically or grammatically unacceptable changes to the target sentences. Examples of these scripted revisions are provided below in Table 2.

Table 2. *Scripted Revision Sentences*

Correct	Insufficient
Globalization, or the integration of cultures across nations, has become an increasingly prominent yet controversial topic. (Line 1)	Due to it's benefits of greater creativity and appreciation of heritage, some defend globalization. (Line 2)
Despite many individual's preference in films that reflect their own cultural identities, trends indicate rising popular interest in cross-cultural movies. (Line 10)	The domestic film markets in many nations are now booming and can even compare with or surpass the United States in production and popularity. (Line 5)

Designing the Shared Workspace Task.

For the joint editing task, the paragraphs were then embedded in a Google Doc to be presented to participants. Each sentence was presented in as a separate numbered line with an empty text box below the original sentence, as shown in Figure 1. This text box is where both the participant's revisions and the recorded revisions are to be entered. This format was used for both the shared document in which the first round of revisions took place, as well as a second document to which only the participant had access.

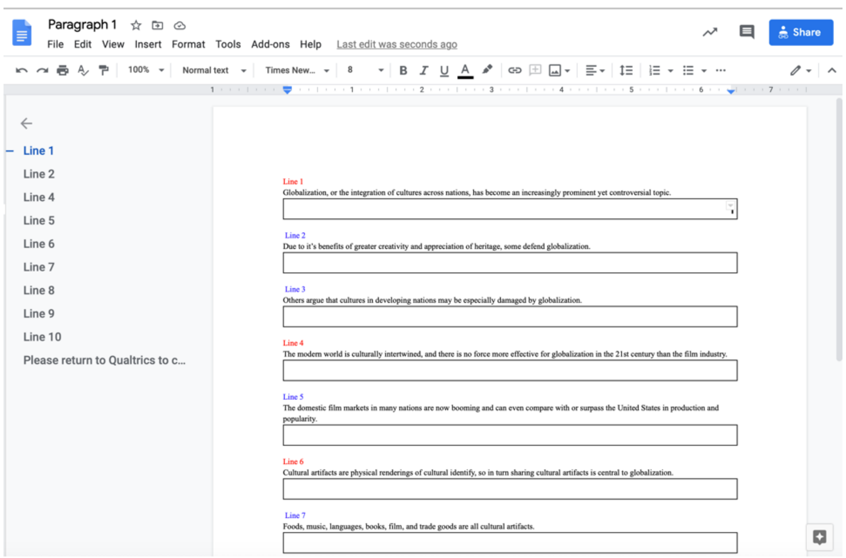


Figure 1

A screenshot showing the shared workspace for the joint editing portion of the task.

For each scripted revision sentence, I created two prerecorded clips to mimic the partner's typing during the editing task: a fluent version and a disfluent version. These typing clips were recorded using Mouse Recorder (https://www.mouserecorder.com), which records keystrokes so that they can be played onto any text-input area (whether in a writing program such as word, a search engine, or anywhere else that one can type), exactly as they were recorded. The timing between keystrokes, as well as any backspaces and changed content are preserved when the recording is replayed. When viewed within the shared document used for the editing task, the replay of these recordings is indistinguishable from real typing.

The recorder used to create the partner's edits additionally records data for each keypress, which allowed me to ensure that the properties of each pre-recorded sentence fit the fluent/disfluent distinction as defined by our Experiment 1 results. Specifically, the macro recorder provides the ability to compute the average typing speed after recording is finished. To ensure appropriate typing speeds, I tracked the keystroke speeds of each scripted revision and selected the final recording set based on the computed speeds, as well as how naturally all other features were depicted in each particular clip. Based on the keypress data obtained in Experiment 1, fluent versions of the scripted revisions included average raw keypress speeds below 0.3 seconds (the average unadjusted speed for items that had been repeated twice), no more than two pauses over 200 ms, and at most 1 to 3 low-level typing errors. The fluent recordings also did not include any significant edits that include over three repeated backspaces or keypresses. Disfluent recordings were created to have at least two content edits (i.e., backspace events that include over three consecutive keypresses), 1 to 5 low-level typing errors ("typos"), and an average keypress speed ranging from 0.3 to 0.5 seconds between keystrokes.

Procedure

For the main collaborative editing task, participants were asked to work with a partner to jointly edit a series of paragraphs. The participants were informed that they and their partner would each be assigned specific sentences within each paragraph to re-write as needed. After completing an initial Qualtrics survey used to obtain informed consent and to present the task instructions, participants then followed a link to a shared Google Doc where they edited their assigned sentences for each paragraph. While working within this Google Doc, they could also view the partner's "editing" contributions in real time. Participants were told they were completing this task with another participant. However, the typing of the other 'participant' was conveyed through a series of pre-recorded typed macros. Because this task was carried out in Google Docs, both the participant's and the pre-recorded typing cursors were followed by a Google Doc name-tag. The pre-recorded typing was identified across both conditions with the name "Taylor Lee", which was intended to be gender and racially neutral. The scripted edits were inserted at predetermined locations in the shared document and were designed to appear to come from the other person actively editing the document in real time.

Half the sentences in the paragraph were assigned to the participant to edit, while the rest were assigned to the partner. The participants were informed that they were randomly assigned a color (red or blue) and that they should edit all sentences in that assigned color; however, the participant was always assigned red. Sentences assigned to the partner were interspersed amongst the items assigned to the participant, ensuring that both the participant's edits and the pre-recorded typing were equally distributed throughout each paragraph. Also, to encourage participants to attend to areas of the shared document where the partner's pre-recorded edits appeared, the order of assignments followed an alternating pattern with no more than two sentences assigned to the same individual occurring back-to-back.

The task instructions explained that the participants would be editing a series of short paragraphs in conjunction with a partner, and that half of the lines of each paragraph will be assigned to them, while the other half will be assigned to their partner. The participants were instructed that they should enter their edits into the text box below each assigned sentence in the

Google Doc, and that they should not leave the shared document until both partners completed their entire set of revisions for that paragraph. This was to prevent participants from moving through the task too quickly and outpacing the pre-scripted partner.

Once the participant completed their re-writes of each line they were assigned for a given paragraph (and all the pre-scripted edits by the partner were entered as well), they were instructed to return to Qualtrics, where they were given a link to a separate document. In this document, participants were provided with all of their partner's final edits from the previous paragraph. Although participants saw these edits being produced by the partner either fluently or disfluently, the final set of sentence revisions were always the same for all participants – specifically, they were the static scripted sentence revisions created for this purpose. Within this separate document, the participants were then asked to make any additional edits they believed would be necessary to improve the sentences further, or, if they believed the sentence required no further edits, they could simply re-type it "as is" in the box below each sentence. Participants alternated between the two tasks through all four paragraphs. They were led to believe that their partner was also editing their revisions in another document at the same time.

Finally, after completing all four paragraphs, participants were given a series of post-task questions that were intended to probe the participant's perception of the task and of their partner in particular. The complete set of post-task questions is presented in Appendix A. First, participants were asked to estimate what percentage of the original sentences required revision and what percentage of the partner's edits required further revision. Then, they provided ratings of how difficult it was to edit the original sentences as well as the difficulty of editing the partner's revisions. They also responded to a series of questions about both their impression of the partner and the experience of working in a shared document. These questions were:

- "How *effective* do you believe your partner was in revising their original sentences?"
- "How heavily did your partner edit their original sentences?"
- "How *confident* do you feel your partner appeared in making their revisions?"
- "How much did you *notice* your partner's editing as you worked on each original paragraph?"
- "Do you believe you would have edited their sentences more thoroughly if you had not previously worked with them in the shared document?"

Finally, participants were asked if they ever questioned whether their partner was a real participant. Following the post-questions was a debriefing page explaining the nature of the task and informing them that they did not work with a real participant.

Results

Edit Distance

My primary interest is in the degree to which exposure to partner's fluent or disfluent typing in the joint editing task would influence how much participants chose to further revise each of the partner's sentence edits. As an index of how much correction participants felt the partner's sentences needed, I operationalized the extent of additional editing that participants engaged in by computing the Levenshtein distance between the pre-scripted edits and the participants' modified versions. The Levenshtein distance is a metric of the "edit distance," which quantifies the difference between two strings of text based on the number of text insertions, substitutions, and deletions. For instance, the Levenshtein distance between the words "forward" and "informal" would be 5 as the difference between them includes five-character modifications, two insertions and three substitutions. An additional example, if the source string is "take" and the target string is "back," to transform "take" to "back," would require three substitutions without additional deletions and insertions. Thus, the Levenshtein distance between "take" and "back" would be 3. To calculate edit distance between the participant's edit and the original 'revised' sentences, I used the *stringdist* package within R, specifying calculation of the "Damerau-Levenshtein distance" for each comparison, which takes into account character transpositions in addition to deletions, insertions and substitutions.

To analyze the impact of partner fluency on how much participants chose to edit the partner's revisions, I fit a linear mixed effect model to our computed measure of edit distance, with sentence correctness, partner fluency and their interaction as fixed effects. Because the Levenshtein distance produces a count of the number of changes between the original and revised sentences, this model used a Poisson link function. Our model also included by-participant and by-item random effect terms. This model revealed a significant effect of sentence correctness, with a greater edit distance when the sentences were 'insufficient' or included errors (M= 25.05; SD=6.49) than for 'correct' sentences (M = 17.91; SD = 3.97), t(38) = -2.29 p = .03. However, partner fluency did not have a significant effect on editing rate, with edits to sentences produced disfluently showing an edit distance (M = 21.59; SD = 2.23) similar to edits to fluently produced sentences (M =19.40; SD = 4.35), t(38) = -.78 p = .44. The interaction between correctness and fluency was not significant as well, t(38) = -1.09 p = .28.

The effect of sentence correctness, as expected, indicates that participants did indeed engage in more editing for the sentences that contained errors. The lack of an effect for partner fluency, however, fails to support the prediction that the extent of participants' edits to the partner's task contributions would be shaped by whether the partner appeared to be confident or not in their typing.

Post-task Questions

To further explore participants' impression of their partner and the editing task, we also examined the questions that participants completed at the end of the experiment. However, due to an error with Qualtrics during data collection, some participants were not able to move through each stage of the experiment as designed. As a result, these participants instead had to be manually sent each document as they progressed through the task. This included the final portion of the experiment that included the post-task questions, and not every participant chose to return the post-task questionnaire. While the primary experiment had 38 participants, 32 completed the post-task questions in full. Here, I report the responses from those 32 participants, 17 of whom were in the fluent condition and 15 were in the disfluent condition.

First, I report the results for the question that asked participants if they questioned whether their partner was a real participant. Responses were given on a three-point scale with 3 indicating "yes," 2 indicating "somewhat," and 1 indicating "not at all." Of the participants who responded to this question that worked with the disfluent partner, 10 of 18 (55%) responded that they did question at least 'somewhat' whether their partner was real, in comparison to only 5 of 19 (26%) participants with the fluent partner. Although participants with the disfluent partner were more likely to report thinking this, this is not a significant difference ($t(35)= .18$ $p = 0.86$). Participants were also given the opportunity to expand on why they did or did not question the authenticity of their partner, which produced a range of responses. Some responses included:

- *They worked very quickly and some of their revisions made the content more confusing* (responding "Yes" about a disfluent partner)
- *I assumed we'd been paired off as we were told, I had not thought twice about it* (responding "Not at all" about a fluent partner)

- *Because I did not know the person, nor had I even heard of them, so I questioned whether they were real or not* (responding "Somewhat" about a fluent partner)
- *I heard it was a partner study so assumed my partner was also another student* (responding "Not at all" about a disfluent partner)

Next, we turn to the questions asking for participants' judgments of the difficulty of the task. On a five-point Likert scale, participants were asked to rate how difficult both the original sentences and their partner's revisions were to revise. On this scale 5 was extremely difficult and 1 extremely easy. The average rating of the difficulty of revising the original sentences was similar for participants in the disfluent condition (M= 2.72; SD=3.1) and participants in the fluent condition (M= 3.26; SD=2.48), t(35)= .58, p =.56.

However, their ratings for the difficulty of revising the partner's edits was significantly lower in the disfluent condition (M= 2.05; SD=0.52) than the fluent condition (M=2.47; SD=0.68), t(35)= 2.04, p = .048. This could indicate that working with the disfluent partner made the errors in the revised sentences more salient. Typing error frequency could then lead to a higher expectation of errors in the revised sentences, thus leading to an easier correction experience for participants who had previously observed typing errors.

Participants were also asked to rate how much they noticed their partner's edits during the task, with 1 being they did not notice their partner and 5 meaning that they were always aware of their activity. The average rating for this measure for participants in the disfluent partner condition (M = 2.83; SD = 1.12) was similar to participants in the fluent partner condition (M = 3.0; SD = .78), t(35) = 0.53, p = .59. Likewise, in response to whether they would have changed their editing behavior had they not known they would return to a shared document with their partner (1 indicating 'definitely would not have' to 5 or 'definitely would have'), the average response for participants in the disfluent condition (M = 1.94; SD = 1.77) did not differ from the average from participants in the fluent condition (M = 2.79; SD = 1.47), t(35) = 1.61, p = .12.

Participants were then asked how much they thought their partner revised the original sentences. For this question, they responded on a 5-point Likert scale with 1 indicating they thought that the partner edited less than half the sentence and 5 indicating they though the partner edited most of each sentence. Participants in both conditions (disfluent condition: M = 3.42; SD = .80; fluent condition: M = 3.55; SD = .72) gave similar ratings of the partner editing amount, t(28) = 0.42 p = .68. Participants also rated on a Likert scale, with 1 meaning least effective and 5 very effective, how effective their partner's edits were. For this question, ratings from participants in the disfluent condition (M = 3.17; SD = 1.58) did not differ from those from participants in the fluent condition (M = 3.68; SD = 4.11), t(35)=.43 p=.67. Both of these patterns are consistent with the fact that the final edited sentences were always identical across conditions. For their rating of how confident their partner seemed (1 being very unconfident and 5 being very confident), although participants in the disfluent partner condition gave ratings (M = 3.22 SD = 0.80) that were lower than those from participants from the fluent partner condition (M = 3.58; SD = 1.34), this was not significant, t(35) = 0.96 p = .34.

Finally, we asked participants to judge the quality and effectiveness of the partner's edits. In general, the response to the question "what percentage of your partner's edits required further revision" appeared somewhat consistent with the prediction that disfluency would lead to a higher awareness of errors. On a percentage 0-100 scale, participants in the disfluent condition rated their partner's initial edits as requiring more substantial revision (M = 49.4; SD = 26.0) than participants who worked with a fluent partner (M = 44.0; SD = 22.6). However, this distance was not significant, t(35)= 0.69, p = .49.

Experiment 3A Discussion

In Experiment 3A, I sought to identify how participants' editing strategy within a shared workspace would be influenced by the nature of the collaborator's contributions, and specifically by how confident their typing appeared to be. Although I found that participants edited more heavily when the partner's revised sentences included errors, they did not edit significantly more when the partner had been disfluent during the joint editing task. While the effect of revision correctness suggests that participant's level of editing was impacted by sentence quality, the lack of an effect for partner fluency has a few potential interpretations.

A first possibility is that our measure of edit distance (as computed using 'stringdist') may not be sufficiently sensitive to capture nuances in how people responded to evidence of disfluency. For example, imagine two distinct editing strategies that could have been employed by participants, one in which the sentence content is rearranged and unnecessary words are removed but the sentence is left fundamentally intact, while another involves re-writing phrases in different terms or making significant changes to a portion of the sentence while leaving another portion as is. A sentence edit that paraphrased or re-worded content might be expected to represent a greater degree of change than a revision that maintains unaltered portions of the sentence and reduces the amount of smaller filler words. In other words, the degree of distance calculated based simply on character-by-character deletions or substitutions does not differentiate based on content changed, but by number of characters changed. The degree of possible variations in editing behavior that could occur even if the amount of characters changes remain comparable, poses a challenge for quantifying the difference between such changes.

To make this clear, consider Table 3 below, which presents three sample participant revisions. Determining which qualifies as more "substantial" is not clear-cut. With automated measures of text similarity, such as string distance, these differences can get lost, as demonstrated in these examples. Characters being removed are considered "equal" to characters being introduced, for instance, although it is not clear whether one signals a greater degree of editing than another. While the second sample revision summarizes and re-states the original sentence it receives a similar edit distance score as sample 3, which maintains unchanged portions of the sentence in a modified order. This illustrates that character changes that result from moving portions of the sentence, such as the change in sample 3 of moving the phrase 'under their care,' results in the addition of novel characters 'along with having them' that are counted individually.

Table 3

Sample Participant Revisions

	Sample Sentences	Levenstein Distance Score
Original Sentence	They teach at the unique level of the children under their care, continually working to prepare them for many coming years of formal education.	0 (original)
Revision Sample 1	ECE educators teach their children on a unique level, for their work prepares them for future formal education.	79
Revision Sample 2	They prepare their students for the upcoming years of formal education, tailoring their teaching to the child's specific needs.	100
Revision Sample 3	They teach children at a unique level. They prepare these children for many years of formal education along with having them under their care.	95

An additional issue could be that the delay between the participant's exposure to their partner's typing in the shared document and the individual editing portion of the task muted the potential influence of fluency. Partner fluency may have affected editing more strongly if the task, like a true collaborative writing task, had been structured so that participants could more directly intervene during the shared document phase. The post-question responses suggest that fluency did not generally impact their perceptions of the effectiveness of the sentence revisions and may have been less important for determining whether they edited the sentences than the 'correctness' of the edits themselves. This possibility is supported by participant's responses to the open-ended question about whether they questioned the authenticity of their partner. Some responses indicated that the frequency of errors in the corrected sentences, as half of the sentences were designed to include errors, made them question their partner. This suggesting that our method for testing the likelihood of being 'proactive' in collaboration with an uncertain partner was flawed.

Finally, there of course remains the possibility that participants were not significantly impacted by their partner's fluency, or they were able to ignore the presence of another typist. Importantly, while the responses to some of the post-questions indicate that typing fluency was likely noticed by the participants, it is not clear whether this led to differing impressions of the same set of sentences, and of the partner, across fluency conditions. While some of the post question results trend in the direction of our hypotheses, further exploration is needed into how disfluency could affect participants' perceptions of the partner's output quality. While the post-questions provided some exploratory insight into participants' experience of the task and their partner, this study wasn't able to achieve the goal of identifying how disfluency might influence their specific attention to errors within the sentences.

Recall that in Experiment 2, I probed participants' interpretations of typing fluency based on watching a non-interactive video of a person's typing. In contrast, Experiment 3a was my first attempt to assess the impact of fluency on behavior within a pseudo-interactive editing task. While my goal of determining how fluency affects work engagement led to inconclusive results, implementing a shared workspace interaction nevertheless prompted more engagement for insufficient partner edits. And, while the difference between fluency conditions did not strongly prompt distinct reactions from participants, many of their post-question responses patterned in the direction of perceptions of less effective, less confident editing from disfluent partners.

Interestingly, one instance where participant's post-task responses did show a difference was that those who worked with a disfluent partner rated the process of revision as less difficult than participants who worked with a fluent partner. As such, visible typing fluency by the partner could still be having an effect upon people's perceptions of the partner and their work by drawing attention to particular errors. Alternatively, participants may have attributed higher levels of proficiency to the fluent partner's revisions, leading them to perceive the revision process as more challenging. Conversely, participants working with a disfluent partner may have perceived the revision process as less difficult due to their lowered expectations of their partner's abilities.

These results suggested that further exploration using this method of a shared workspace interaction between participants and our fluency manipulated editor could bring insight into how fluency impacts an interaction. However, instead of testing how participants engage with the partner's output directly, I wanted to probe participants' impressions of their partner's work in a more fine-grained way. I explored this possibility further by carrying out Experiment 3B.

Experiment 3B

In Experiment 3A, the post-task questions were intended as a 'holistic' check of the participants' impressions at the end of the experiment. Although participants clearly were engaged in the shared editing task, the results for both the edit distance results and the post question responses were inconclusive. Thus, in Experiment 3B, I chose to implement another version of the shared editing task that would let me probe participants' evaluations of the partner's edits more directly, on an item-by-item basis. That is, rather than asking participants to re-edit the partner's revisions following the joint editing task, I asked participants to directly evaluate the quality of the partner's edits for each sentence. This is more likely to reflect participants' immediate impressions of each sentence revision and may be provide a window into the consequences of the fluency manipulation that is more sensitive than the general post-task questions used in Experiment 3A. If this is the case, I expected that participants who work with the disfluent partner would rate the sentences as less effectively edited than those who work with the fluent partner, despite the final edited sentences being the same.

Participants

Thirty-eight Northwestern undergraduates aged 18 to 22 participated in this study for partial course credit. All were native English speakers.

Materials

The paragraphs and revision scripts from Experiment 3a were used again here.

Procedure

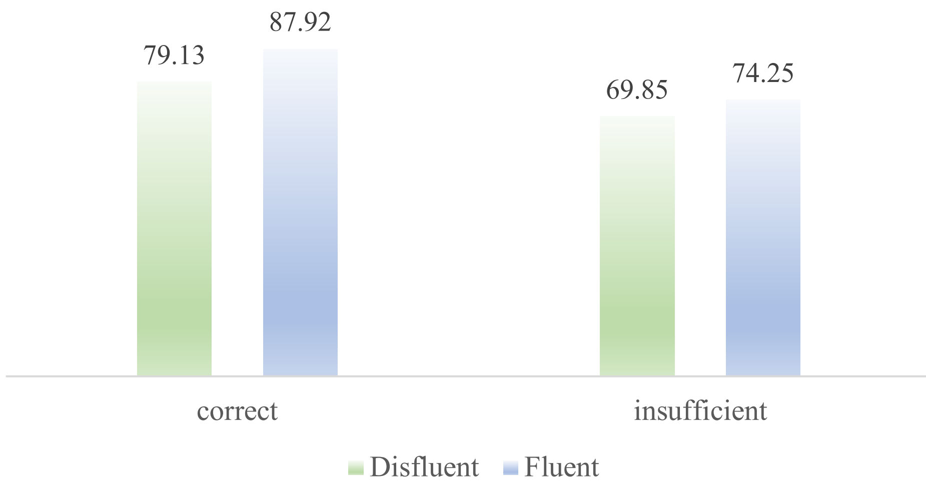
As in Experiment 3A, each participant was told they would be working with another student to edit a series of four short paragraphs within a shared Google Doc, and that each paragraph would potentially contain a number of grammatical and stylistic errors. After carrying out the joint editing task for each paragraph with the partner, participants were again presented with the partner's edited sentences for a secondary judgment task. However, in Experiment 3B, instead being asked to revise the partner's edits in the second document, participants were asked to independently judge the quality of *each* of their partner's revisions to the previous paragraph. Specifically, participants were asked to "consider the overall readability, clarity, and grammatical correctness" of each of the revised sentences produced by the partner (which, again, were always pre-scripted and identical across the fluent and disfluent typing conditions). Participants made this judgment on a 0-100 scale, where "0" was labeled as "fundamental errors and stylistic issues remain; needs complete rewriting" while "100" was "free of grammatical or stylistic errors, no further editing needed." Participants were led to believe that their partner was separately rating the quality of the participants' revisions as well. After providing these ratings, participants were then instructed to return to the main shared Google Doc to continue the collaborative editing task for the next paragraph. The collaborative editing and independent rating tasks alternated until all four paragraphs were complete.

Results

Figure 2 presents the average partner revision quality ratings computed across both revision sufficiency and partner fluency. Again, confirming the success of our manipulation of revision sufficiency, participants gave significantly higher ratings for sufficient revisions ($M=83.52$; $SD=16.86$) than insufficient revisions ($M=72.05$; $SD=21.91$), $F(1, 35)=67.41$, $p<0.001$.

Figure 2.

Edit quality rating scores



Importantly, though, there was also a significant main effect of partner fluency. Participants also gave significantly higher ratings for revisions produced by a fluent partner ($M = 81.08$; $SD = 20.97$) than for revisions produced by a disfluent partner ($M = 74.49$; $SD = 19.19$), $F(1,35) = 22.30$, $p < 0.001$. However, these two factors did not significantly interact, $F(1,35) = 2.46$, $p = .117$.

Experiment 3B Discussion

In this experiment, I shifted the focus to a more direct measure of the extent to which disfluency impacted the participants' perceptions of the partner's revisions. Rather than centering our exploration on the work that participants produce and on the global, post-task impressions, Experiment 3b directly captures participant's perception of the partner's edits while the task is ongoing. Participants who worked with the disfluent typist gave the set of sentences lower ratings than participants in the fluent condition, and this pattern held consistent across error conditions. Sentences that included more errors were rated lower overall than sentences without errors, confirming that participants were sensitive to the actual sentence quality, while disfluency led to perceptions of more errors in both conditions than people who worked with the fluent partner. This suggests that observing typed disfluency primes viewers to expect to find errors in the output, just from having observed errors during its creation. Importantly, this effect of fluency occurred despite the same final sentence output being present across the fluent and disfluent conditions.

Like spoken conversations, the dynamics of production in text-based shared workspaces can influence people's impressions of one another and their contributions. This unique finding highlights the significance of typing dynamics as cues that shape social interactions within collaborative contexts. It is noteworthy because previous research on typing and text-based shared workspaces has often overlooked the potential of typing dynamics to provide specific and contextually-relevant cues. In Experiment 3B, I demonstrated that people can not only intuitively pick up on text-based cues, but that they allow these cues to influence their perceptions of the written output.

General Discussion, 3A and 3B

In a shared workspace that allows access to character-by-character typing input, the real time dynamics of message formation are present onscreen. As is the case with speech, disfluencies in typing can provide indication, as demonstrated in Experiments 1 & 2, that a typist is currently uncertain about what they are producing. The original goal of Experiment 3 was to explore the impact that onscreen disfluencies have on a collaborative task, and specifically whether they shift the way a collaborator might interact with the content that was created. During Experiment 3A, though, the presence of visible disfluencies had an inconclusive impact on the way that participants engaged with the partner's output in terms of edit distance, whereas editing behavior was more influenced by the actual correctness of the output. And, while participants afterwards reported finding it less difficult to spot errors when their partner's real-time editing process contained onscreen disfluencies, many of the remaining post-question responses showed only trends consistent with the prediction that participants with disfluent partners would view their work as requiring more editing. In previous research on disfluencies (Brennan & Schober, 2001; Corley, MacGregor, & Donaldson, 2007) in speech, listeners are more likely to perceive disfluent speech as less credible or less informative. In Experiment 3B, the focus was shifted to a direct report of the participants' view of the partner's edits on a sentence-by-sentence basis, more similar to Experiment 2. This provided a more direct insight into what kind of impact the disfluencies could be having. Participants in Experiment 3B confirmed the patterns present in Experiment 3A's holistic post-question responses, as participants who worked with the disfluent partner gave overall lower revision quality ratings than those who worked with a fluent partner.

Overall, these findings provide evidence concerning the impact of visible typing production on the dynamics of collaborative editing tasks. The use of shared workspaces can play a crucial role in fostering engagement between collaborators by providing them with a rich set of informative cues about their partner's work process as it unfolds. This not only enhances the collaborative experience but also influences individuals' perceptions and interactions within the task. While we observed a significant effect of fluency in the context of perceptions of partner's edits (3B), this effect was not apparent in the participants' actual revisions of the same edits (3A).

This discrepancy could likely be explained by differences in the nature of the task they were given in each of these experiments. As shown in the results of both experiments, the sentences which were designed to include embedded errors continued to receive significantly lower ratings of quality than the sentences that were intended to represent correct revisions. When designing this experiment, I believed it necessary to include a range of revision quality to better explore the impact disfluency had on participant's perception of the revisions. However, perhaps the range of quality caused the task of further revising to be too dependent on the actual quality of the sentences. In this case, it would make sense that the participant editing behaviors would be comparable across fluency conditions given the unavoidable objectivity of the task of correcting the sentences. In the case of Experiment 3B however, in which they were simply rating the sentences rather than editing them further, the more subjective influence of their experience working with the fluency scripted partner could have more potential impact. In other words, assessing quality and assigning a score rating may allow for more of an impressionistic interaction than further editing which involves directly thinking about the sentences themselves.

General Discussion

Generating appropriate inferences about an interlocutor's confidence is an important part of successful interaction. Among other things, discerning when someone is uncertain about what they are saying is often helpful for knowing how to respond – e.g., whether to accept their contributions at face value or to seek further clarification. In spoken interactions, speakers routinely convey certainty not only linguistically, in terms of *what* they say (message content), but also paralinguistically, or *how* they say it (message delivery). There is a wealth of evidence from spoken contexts demonstrating that speakers routinely rely on both linguistic epistemic markers and various paralinguistic devices to express epistemic state, and that comprehenders can use even subtle cues like filled pauses and certain gestures to infer speaker certainty (Brennan & Williams, 1995; Roseano et al., 2016; Swerts & Krahmer, 2005).

Increasingly, however, people communicate and work together in circumstances that rely almost entirely on written messages. As is well-documented, text-based platforms generally have particular affordances for facilitating interaction, especially when compared to platforms that enable multimodal communication (Brennan & Clark, 1991; Dennis, Fuller, & Valacich, 2008). Among other things, text-based media are often less immediate and place a greater focus on the products of communication rather than on the processes of how messages are constructed. Consequently, communicator (un)certainty in text-based settings is conveyed primarily via linguistic markers and other devices encoded as part of the message-as-product.

However, in shared, fully synchronous platforms like Google Docs and versions of two-way real time chat, users have real-time access to the dynamics of one another's written contributions, enabling other kinds of potentially informative cues concerning communicative certainty. In the present work, I explored the transmission of epistemic state in a pseudo-communicative context intended to mimic this sort of real-time setting, in which messages were conveyed exactly as typed, character-by-character. Compared to more traditional quasi-synchronous chat contexts, dialogue in such "supersynchronous" CMC environments has been characterized as being high in "orality," or bearing a strong resemblance to spoken conversation, with greater numbers of brief, unplanned utterances written in a close, interpersonal style (Jonsson, 2015).

From the analysis of keystrokes in Experiment 1, both typing speed and fluency reflect the ease with which individuals were able to formulate adequate image descriptions. When items were conceptually challenging to describe, typing was generally slower and less fluent, with frequent pauses and backspacing. But repeated experience with the same images allowed participants' typing to became faster and more fluent over time, presumably due to increased confidence in how to describe these items. Notably, these typing dynamics mirror analogous patterns observed in spoken referential communication tasks in which speakers, confronted with the challenge of describing novel images (e.g., tangrams) for a partner, typically start by producing verbose, disfluent utterances with many qualifiers, but become more concise and efficient with repetition (Clark & Wilks-Gibbs, 1986). Similar to how paralinguistic features of spoken utterances may reflect speaker certainty (Brennan & Williams, 1995; Smith & Clark, 1993), these results show that the speed and fluency of written (typed) message production can reliably reflect the epistemic state of the typist. These findings also add to the body of work documenting various ways that typing behaviors are indicative of a typist's mental or emotional state (e.g., Allen et al., 2016; Epp et al., 2011; Vizer et al., 2009). Thus, even beyond the content of written messages, the keystroke dynamics of message production can indicate the sender's confidence concerning what they are trying to communicate.

To explore the possible role of visible typing behavior as a meaningful communicative cue, though, I sought to expand on experiment 1 by demonstrating not just that typing speed and fluency can vary along with the typist's degree of certainty, but also that these behaviors are salient to others in ways that could reliably inform inferences about the typist's epistemic state. Thus, in Experiment 2, I presented brief video clips of image descriptions recorded as part of Experiment 1 to a separate group of participants, who were asked to make a series of judgments about the original typist and about the behaviors observed in the clip. Although each participant judged only a single clip, each clip was selected to ensure the presentation of the full spectrum of image descriptions that independently varied in both typing speed and typing fluency across the full set. Clips with slower typing speeds and more frequent backspacing were judged as being produced by typists experiencing less certainty, more difficulty with their descriptions, and less familiarity with the object being described.

Broadly, these results show that individuals are able to make accurate epistemic judgments based on observations of how someone types. These metacognitive judgments are similar to other demonstrations in spoken language contexts involving “feeling of another’s knowing” (FOAK), in which listeners take speech disfluencies and other hesitation phenomena as evidence for judging when someone might have less knowledge of the topic at hand (Brennan & Williams, 1995; Swerts & Krahmer, 2005). Just as verbal disfluencies, intonational shifts, and manual gestures can signal uncertainty in face-to-face communicative contexts (Roseano et al., 2016), typing dynamics can inform epistemic inferences in text-based communicative media that make processes of message formulation visible in real time.

One limitation of the first two experiments, is that I did not fully control the content of the image descriptions, which were selected from the descriptions produced as part of Experiment 1. This choice was intended to elicit epistemic judgments based on observations of naturalistic message production, rather than, for example, taking a base image description and attempting to artificially insert (or remove) hesitations or typing disfluencies at key moments. It is possible that participants in Experiment 2 based their judgments at least in part on the content of the descriptions rather than on the nature of the typing alone. Even so, when asked which aspects of the descriptions influenced their judgments, participants frequently mentioned specific features of the visible typing. This provides confidence that participants were using the typing behavior to draw inferences about of the typist’s epistemic state.

A feature that separates the first two experiments is the decision to examine the dynamics of message production in Experiment 1 separately from aspects of message reception in Experiment 2. This enabled direct control over when individual typists might express uncertainty in their image descriptions and over how particular features of these production dynamics would be conveyed to potential recipients. In experiments 3a and 3b I sought to combine these aspects to create a collaborative context to observe whether the patterns in experiments 1 and 2 hold consistent. While the results of experiment 3b were as expected, as in experiment 1 & 2 disfluency led to lower ratings of effectiveness, experiment 3a had points of inconsistency. Contrary to the Experiment 3a hypothesis, fluency did not measurably influence participants' editing patterns towards the content produced by their collaborator. Notably, in the post questions for experiment 3 participant ratings for their partner's confidence were not significantly impacted by fluency. As opposed to experiments 1 & 2, in which there were no right or wrong ways to complete the image description task thus the content of the task bore less influence on rater's perceptions, experiment 3a included a second condition of item correctness, featuring content that either included embedded errors or was fully correct. To create a realistic yet controlled collaborative task, the potential influence of the task content itself was difficult to avoid. The expansion into testing how disfluency impacts an ongoing collaborative task, led to the demands of the task informing participant's editing behavior enough to overpower the potential effect of their partner's fluency. Experiment 3b, which examined self-reported perceptions within the same context, indicated that typing fluency still influenced participants' impressions of their partners' revisions, with disfluency again being taken as an informative cue that shaped their global perception of work quality. This suggests that impact of disfluency may be most observable in how it shapes collaborator's assessments after the fact, rather than in the moment of working on the task.

Perhaps instead of testing the impact of fluency on editing behaviors, a prior step may have been to explore other aspects of the group dynamic in the constructed supersynchronous workspace. While in this project, the outcome I have primarily centered here is inter-collaborator assessment, the role of disfluency in collaboration may be equally relevant in how it informs immediate communicative decisions. Studies of collaborative writing and computer-assisted language learning within online groups have mostly focused on the nature of the collaborative activity as an aid to task or learning outcomes (e.g., Abrams, 2019; Li, 2018; Woodrich & Fan, 2017). Yet, when multiple users are simultaneously logged into and entering text within the same document or chat application, with the ability to see each other's contributions in real time, this has the potential to shape not only their interactions and interpretations of one another, but how they manage the task at hand. For example, studies looking at "supersynchronous" two-way chat contexts have shown that, despite the increased prevalence of simultaneous overlapping turns across individuals, conversational coherence may be maintained in part by the ability of users to monitor one another's typing (Anderson et al. 2007). Throughout this project, I suggest that access to visible character-by-character typing could likewise give individuals in online contexts the ability to respond more immediately to moments in which another user appears to be struggling in some way (or, alternatively, appears fully confident). Moreover, the implicit nature of typing behavior means that this sort of epistemic information can be available without the need for explicit discussion. This idea mirrors findings in the literature on eye gaze and collaboration, which suggests that groups engage in a predictable pattern of observation of each other's gaze based on one another's contribution timings to coordinate implicitly and maintain a shared task model (Bavelas, Coates, Johnson, 2006).

Further exploration with a more varied task under the same paradigm used in experiment 3, a shared workspace interaction with a pre-recorded typist design, could reveal additional ways in which disfluency in shared workspaces impacts group dynamics. For instance, a more realistic free format essay structure could be an optimal task to observe, as it would allow for more advanced editing choices on the part of the participants. While individuals engaged in forms of fully synchronous CMC can generally be expected to attend to the contributions of other interlocutors – thus making it more likely that these kinds of typing dynamics would be a potentially salient feature of interpersonal communication – an open question is how frequently users in other kinds of collaborative contexts would be in a position to attend to these sorts of typing-based cues. Studies of collaborative writing and editing practices, in particular, have observed that users often show a preference for "additive" contributions, with different individuals working on separate sections of a common document, over "simultaneous" contributions, in which different individuals jointly work on the same parts at the same time (e.g., D'Angelo et al., 2018; Wang, Tan, & Lu, 2017).

Thus, although visible typing speed and fluency can function as an informative communicative cue concerning a user's epistemic state, the utility of such cues for shaping collaboration depends on a variety of factors. These factors include the level to which task behavior is interdependent across group members and of course the degree to which other users have the opportunity to notice fluency patterns. Further research is needed in more realistic collaborative environments to determine when and under what circumstances individuals attend to the ways others are formulating their contributions to the task at hand for the purpose of generating appropriate epistemic inferences.

Theoretical Implications

As mentioned previously, the results from this project suggests that keystroke behaviors may hold an analogous role to speech disfluencies and other paralinguistic features of spoken language, both in terms of their meaning at a production level and their interpretation by viewer's/listeners. At the level of production, spoken language and written typing differ in a number of ways, stemming from their purpose and function. Most notably, typed language lacks the auditory and prosodic cues present in spoken language, such as intonation, pitch, and volume. These non-verbal cues play a crucial role in conveying emotions and emphasis during spoken communication. However, typing allows for editing and revising. This can lead to more polished and precise written expressions, as typists have the opportunity to refine their messages nonlinearly over time. Due to these differences, observers of typed language may draw insights from pace and flow of message production, as well as the presence of disfluencies, pauses, and other cues that are also characteristic of spoken language. Some of the similarities between spoken language and typed language lie in their shared goal of conveying meaning and facilitating communication. These fundamentals lead to functional similarities, or features that hold an analogous role to both the speaker/producer and the interpreter. Regarding the expression of uncertainty or knowledge in particular, both typing and speaking can exhibit similar cues. In typed language, just like speaking rate in spoken language, typing speed in written communication can vary based on uncertainty. Backspaces during typing are akin to speech editing or correction in spoken language. In typed communication, individuals can delete and revise their text before finalizing their message, similar to how speakers might correct themselves or rephrase their statements during spoken discourse to clarify their thoughts.

Further evidence for similarities between the nature and role of disfluencies across speech and typing modalities comes from the domain of second-language learning. Second language learning provides a rich context for evaluating the patterns and purpose of spoken disfluencies, as second language learners are often uncertain about upcoming speech, so the opportunity to observe natural speech disfluencies are especially abundant. The categorization of types of disfluencies present in second language speech production closely maps onto the typing disfluency features identified in Experiment 1. For instance, Segalowitz (2010) divides utterance fluency into the subcomponents of speed, breakdown, and repair fluency. Speed here being analogous to our keypress count, as it refers to the density of syllables spoken per minute. Breakdown refers to pause lengths and frequencies. Repair fluency includes repetitions, false starts, or any other form of quick changes while speaking, and is comparable to backspaces within a text-based context (De Jong & Bosker, 2013).

Similarly, these features have been analyzed in terms of listener perceptions of L2 speakers, with findings that a speaker's second language perceived competence is most impacted by speed (or content density), pause frequency, and pause location, and is less influenced by repair frequency (Saito et al., 2018). These results pattern similarly to the results of experiment 2 in which backspacing was the least influential factor in participant's responses to all questions, except for "How much effort is the typist putting in", in which case it was the most influential factor. The implications of the finding that backspaces signal a higher degree of effort, but not a lower degree of general confidence, could suggest that listeners and evaluators are particularly sensitive to the fluency and smoothness of communication, whether it is in spoken language or typed messages. Factors like speed, pause frequency, and fluidity seem to play a substantial role in shaping perceptions of competence or familiarity in communication tasks. On the other hand, errors corrections, or backspacing, might be perceived as more intentional and provide more of an indication into user attention, as opposed to speed which may be viewed as an automatic byproduct of certainty. Just as hesitations and fillers in speech can convey uncertainty or the need for careful consideration, the presence of backspacing in typed text highlights instances where the typist is actively evaluating and revising their written expression. This decision-making process during typing suggests that backspacing could additionally serve as a parallel to pauses in speech, representing moments of cognitive engagement and reflection in the act of communication. While it has been suggested that one reason for pauses in speech is to provide the speaker with more time to search for their next phrase, their communicative intentionality, that is whether speakers intend to convey meaning through fillers and in what circumstances, has been frequently debated, particularly in the case of filled pauses like ‘uh’ and ‘um’ (Schachter et al., 1991; Clark & Fox Tree, 2002). Like spoken language, where fillers and hesitations are may arise naturally in the flow of speech, backspacing straddles a similar line between being a feature that is deliberately or consciously produced and something that arises automatically as a reaction to uncertainty about how to proceed.

The presence of similar features in both spoken and typed language suggests that people rely on consistent cognitive and communicative strategies across modalities. It also implies that individuals can effectively adapt their understanding of these cues to make inferences about speakers' confidence and other psychological attributes in various communication contexts, reinforcing the flexibility of human communication.

Appendix A

Experiment 3a Post Questions

These questions are about the initial document editing task:

What percentage of the *original* sentences did you believe required substantial revision?
0-100 scale

How difficult did you find it to revise the *original* sentences?
Scale 1 (very easy) -5 (very difficult)

These questions are about the subsequent task to examine and revise your partner's edits:

What percentage of your *partner's edits* did you believe required substantial revision?
0-100 scale

How difficult did you find it to revise your *partner's edits?*
Scale 1 (very easy) -5 (very difficult)

These questions are about your impressions of your partner:

How *effective* do you think your partner was in revising their original sentences?
1-5 (very effective)

How heavily did your partner edit the content of their half of the original sentences?
1-5 (very heavily)

How *confident* do you feel your partner appeared in making their revisions?
1-5 (very confident)

How much did you *notice* your partner's editing as you worked on each original paragraph?
1-5 (a great deal)

Do you believe you would have edited their sentences more thoroughly if you had not previously worked with them in the shared document?
1-5 (definitely yes)

Did you ever question during the task whether your partner was a real participant?
1 (no, not at all) – 3 (yes)

Why or why not?

Appendix B

In my proposed Experiment 4, I sought to identify how cues from real-time typing could have immediate effects on a partner's interpretation of typist's communicative intentions. In doing so, my goal was to expand the understanding of what *kinds* of information can be deduced from having visual access to a collaborator's real time typing. As I describe in this Appendix, however, technical difficulties and other issues prevented me from carrying out a successful version of this experiment as originally proposed. Here, I describe the experimental logic and method of the study as design and attempt to give a characterization of some of the issues that prevented the study from being completed. We believe that a future version of this study that addresses these issues could provide valuable additional evidence regarding the interpretation of real-time typing behavior.

Experiment 4

Experiments 1-3 mostly focused on how typing cues provide *broad* evidence of a typist's knowledge and certainty. Participants' interpretations of the typing cues were primarily about the cues themselves, rather than the context surrounding them. Indeed, much of the literature on the types of information that can be drawn from typing patterns has been centered around global inferences about the typist's state of mind, such as their emotions (Epp, Liphold, & Mandryk, 2011) or use of deception (Banerjee et al., 2014). However, it is also possible to ask whether the cues that indicate uncertainty on the part of the typist might more specifically influence a viewer's expectations about the *source* of that uncertainty. In Experiment 4, participants were intended to have the opportunity to use the visible typing during a referential communication task to infer something about the typist's degree of uncertainty. When the communicative context includes two referents, one familiar and one unfamiliar to the typist, I was interested in whether participants would take the presence of disfluency as an immediate cue that the typist is attempting to refer to the unfamiliar item.

Barr and Seyfeddinipur (2010) tested a version of this question in the context of spoken language comprehension, examining whether listeners take spoken disfluencies as a cue about speaker intentions, even in a situation in which listeners were not asked directly to attend to speech fluency. Participants viewed pairs of items on a computer screen and heard instructions from one of two speakers directing them to click on one of them. A subset of items repeated across trials, while a set of new items were continually introduced during the task. The two speakers provided instructions in alternation, which meant that certain items were new to one speaker but had been described in an earlier trial by the other. Participant's mouse movements were tracked as they selected between each item. This enabled the comparison of participant's movement trajectories, which provided insight into their expectations during each moment of the audio recording. The researchers found that when viewing item pairs that consisted of an item familiar to the current speaker plus an item new to that speaker, participants were faster with a more direct selection trajectory to choose the *new* item when the speaker began their description with a filled pause like "uh." Participants apparently used the presence of the filled pause as evidence that the speaker was about to identify the novel item rather than something already familiar. This suggests that the relation between spoken disfluencies and perceived confidence is strong enough to not only affect judgments about the speaker, but also can impact expectations at a level that influence the listener's behavior.

In Experiment 4, I attempted to replicate this design using typing disfluencies. In doing so, I wanted to test the extent to which typing disfluencies could prompt similar expectations about the typist's intentions, broadening an understanding of the types of inferences that can be made from typing patterns alone. To do so, I generated video recordings of typed item descriptions, similar to those used in the typist judgment task in Experiment 2. However, the content being typed, as well as the locations of pauses and other disfluencies, was directly based on the materials used in Barr and Seyfeddinipur (2010). In their study, each experimental item included a pause at the beginning of the sentence after a word or two had been spoken. For instance, a transcribed example of an item with an included disfluency would read "(pause) um the cake with the candles on top" whereas the same item without the disfluency included a pop of static audio to replace the 'um'. These pauses were covered by the sound of static to create the impression that the pause was not caused by the speaker, but rather a minor sound glitch in the recording. The items that included an audible pause were considered the "disfluent" items, while the items in which the pause was disguised were "fluent" items. To reproduce this logic with written production, in the materials for Experiment 4 I included a visible pause in the typing at the moments where the corresponding vocal recordings featured audible pauses. However, where their original audio included a glitch covered pause, I instead included a visible loading "buffering" circle icon that appeared for the duration of the pause. Thus, in the disfluent condition the pause was intended to be attributable to the typist hesitating, while in the fluent condition the pause appeared to be caused by recording or playback error. I used the iSpring screen recorder to create typed versions of these vocal recordings, which included both fluent and disfluent items.

To facilitate the participant's awareness that a given item is new to a *particular* typist, the recorded typing from each of the two typists was identified with a unique name. The included names were distinct in gender and length, as was the case with the named speakers in Barr & Seyfeddinipur (2010). In their original study the audio recording implicitly emphasized the speakers identify through the contrast of a male and female voice recording; the present text-based format required explicit identification associated with each typist. To do so, I created separate named accounts to be used in Google Docs, so that as the typed content was recorded, the name of such accounts would trail beside the cursor (as shown in Figure B2, below). This allowed the distinction between typists to be salient even without an audio component, helping participants to keep track of which typist saw which items throughout the task. On critical trials, when one object was familiar and one object unfamiliar to the current typist, this in turn was expected to shape how disfluency in the item description would impact the participant's expectations of which item is likely to be the target. If people are sensitive to typing disfluencies in ways similar to how they interpret spoken disfluencies, I predicted that a visible disfluency early in the clip should lead the participant to expect that the typist will refer to the novel item.

Like Barr and Seyfeddinipur (2010), I used mouse movements as the primary measure of how quickly participant's formed expectations about which item was being referred to in the video clips. Mouse tracking has been used to capture real-time processing of information through the access it gives to a person's decision making before a selection is made (Spivey, Grosjean, & Knoblich, 2005; Freeman & Ambady, 2010). Previous research has shown that the trajectories a person makes when deciding where to click varies based on the ambiguity of the selection choice, such that the closer one's trajectory fits to an 'ideal' or straight path, the less conflict or uncertainty the person experiences in making the decision. Trajectories that vary or that include shifts in their placement on the screen occur more commonly when a selection is not obvious (Freeman, Pauker, & Sanchez, 2016). The timing and speed of selection can also be indicative of both *when* a person came to a conclusion and how strongly they believe that conclusion (Piovesan & Wengström, 2009).

Participants were instructed to quickly select the item they believed the person was describing as soon as they thought they knew which of two options was being described. I recorded their mouse movements, intending to use the latencies and paths of their item selection as a measurement of their interpretations of the descriptions. This would allow me to capture not only what cues in the video might have led them to choose the item (based on time and speed), but their level of perceived ambiguity between items based on their movement patterns. In other words, when items were described somewhat ambiguously (such as if the typist were to begin their description by referring to traits present in *both* items), the expectation is that their mouse movements would display a variable trajectory. However, if typing behavior can serve as an indicator of uncertainty, the early presence of disfluencies should disambiguate the description and may result in a straighter path to the novel item.

When an item is referred to with disfluent typing, we would expect participants to be faster to select novel or ambiguous items. If the selection trajectory and selection speed with which participants identify which object is being described varies depending on the level of disfluency, that will suggest that typing patterns, regardless of the content being typed, can facilitate expectations about communicative intentions.

Method

Participants

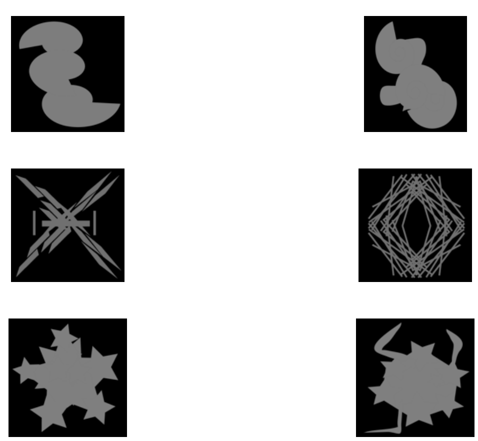
Fifty-six Northwestern undergraduates aged 18 to 22 participated in this study for partial course credit. All were native English speakers. The participants carried out their completion of the study remotely.

Materials

The images used as targets for the item descriptions presented to participants consisted of a set of black and white ambiguous shapes, selected from Barr and Seyfeddinipur (2010). I organized these shapes into pairs based on how many features they share in common, allowing the content of the typewritten descriptions to be potentially applicable to both items for the initial portion of the video clip. This was intended to prevent participants from immediately identifying the target based on content alone, potentially before the inclusion of any disfluencies. Examples of item pairs are presented in Figure B1.

Figure B1.

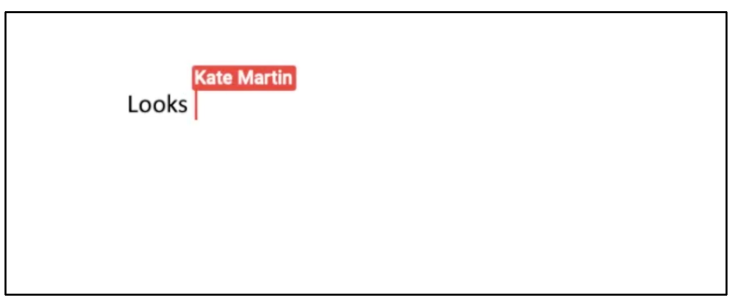
Example item set presented in pairs



Based on these item pairs, I created a selection of videos depicting, in simulated real-time, typed descriptions of individual target images. Aside from the description content, the typing also indicated the identity of the typist. When text is entered in Google Docs, the name of the account follows each keystroke, as shown in Figure B2. Having this name information visible allows the 'identity' of each typist to be salient for participants. There were two typists, each with a distinct name "Kate" and "Mark" (these same names were used in Barr and Seyfeddinipur, 2010). If participants keep track of each typist throughout the task, they should develop an awareness of which items are familiar or unfamiliar to each typist. This awareness is necessary for meaningful inferences about target items based on the presence disfluencies in critical descriptions.

Figure B2.

Sample image of the typist nametag from a trial video

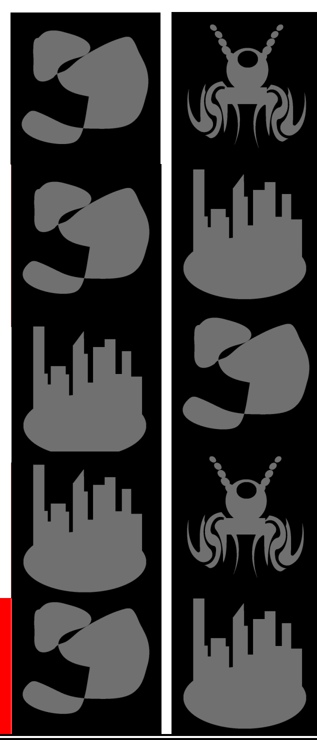


The script for these typed descriptions were based on the audio descriptions used in Barr and Seyfeddinipur (2010). For each target description, I created two video clips, one featuring fluent typing and the other disfluent typing. Typing disfluency was manipulated via the inclusion of a visible pause in the typing in the disfluent version versus a "disguised" pause in the fluent condition. The pause onset occurred within the first two seconds of all experimental trials. In both conditions, these pauses lasted for a five-second duration. The disfluent versions included an 'uncovered' pause that appeared as if the typist stopped typing (i.e., the cursor stayed in one place). The fluent versions included an interruption of the same length, with onset and duration held constant. However, the pause was replaced with a visible "loading" icon overlay that appeared over the screen during the duration of the pause, which was intended to create the impression that the typist did not pause, but rather the video clip froze. Both versions of these videos featured the same final description being typed following the critical pause.

Following B&S (2010), the items were organized into five trial blocks in which familiarization trials always appeared before their associated experimental trials. This was intended to allow participants to develop representations of which items being described by which typists as the block unfolded. The block arrangement was not identified to participants, who instead experienced each trial individually in a sequential stream. The set of familiarization trials were used to establish typists' familiarity with the non-target items. Half the descriptions of the familiarization trials were fluently typed, with the other half disfluently typed. However, unlike the target items, both items in these familiarization trials were always being described by the typist for the first time, so there was no expectation of a differing selection response from participants. The purpose of varying the level of fluency on these items is to keep the familiarization items consistent with the target items, so there will be no discernable difference between how these trials are typed.

To reiterate the logic of the stimulus design, on critical trials when one option is familiar to the typist and the other option is not, if the description is produced disfluently, I expect participants to initiate mouse movements to the unfamiliar item first and more quickly when a disfluency occurs that can disambiguate the initial description. Conversely, if the description is produced quickly without errors or changes, I would expect them to infer that the typist is more likely to be referring to the item they've described before, rather than the item that is novel to that typist. This may be especially true during later trials, at which point the participant would have observed enough disfluent typing to potentially form inferences about fluent typing, which all else being equal might be viewed as less informative than typing with errors. No trials in which both items are familiar were included, as that would not be informative in either direction.

Figure B3. *Sample trial order of images in relation to the typist describing them*



Trial	**Trial Type**	**Typist**
1	familiarization	mark
2	familiarization	mark
3	familiarization	mark
4	familiarization	mark
5	critical	Kate

In Figure B3 the experimental trial features a typist (Kate) referring to an item which is new to them but has been previously seen (and referred to!) by the other typist. In half of the experimental trials, the current typist referred to a novel item, while the other half the typist referred to an item they had described previously. Following B&S (2010), in the experimental blocks the typist who would appear in the fifth critical trial was always “Kate”. This allowed the typist familiarity to be manipulated based on which typist appeared in the familiarization trials. Half the blocks ended in a critical trial in which “Kate” appeared in the four familiarization trials and was thus familiar with the critical item in the experimental trial, having described this item previously. In the ‘unfamiliar’ condition, “Mark” was the typist for the four familiarization trials, such that in the fifth critical trial Kate would not have previously described any of the items in that block (as in the example in Figure B4). Additionally, we included a larger set of non-critical blocks in which Kate and Mark were equally represented in all five trials.

The use of the fluent or disfluent clip for each given experimental trial was counterbalanced across two versions of the experiment, with one version of the study assigning half the trials to include the disfluent clip (distributed throughout the set at random) and the other version switching which items were assigned to play the fluent/disfluent version. Participants saw only one version of each critical trial.

Procedure

Participants were presented with the experiment in Qualtrics. They were told that they will be presented with a series of brief screen recordings taken from a prior experiment that will present someone typing out an image description within a shared Google doc. For each description, the participants’ task was to quickly select, by clicking, which of two images on the computer screen they thought the person in the video was describing. They were encouraged to make their selection as soon as they think they know which item is being described.

I tracked participants' mouse movements within Qualtrics using an open-source modification from Mathur and Reichling (2019), which enables the collection of all speed and coordinate location data from the mouse movements of anyone responding directly within Qualtrics. Mathur and Reichling developed this mouse tracking modification using Javascript to enable Qualtrics to record the mouse's onscreen coordinates every 10 milliseconds. To foreshadow some of the current challenges with using this procedure, this modification was designed for a perception experiment in which participants were asked to rapidly select whether a presented image was of a human face or a robot. As a result, the mouse-tracking modification was originally designed for recording mouse trajectories that occurred within a 1 to 2 seconds. This was a substantial difference from the design of this study, as participants were expected to watch a video clip with an average duration of 10 seconds before making their selection. In an optimal case, participants would make their selection shortly after the disfluency occurred, which was always within the first five seconds of the video start. However, for familiarization trials or fluent items, it was often necessary for the whole video to be viewed before the item would be disambiguated, as was the case in the audio based original study.

Each typed description recording, and image pair was presented on its own page within Qualtrics. I embedded each video in between the two item options, and the participant clicked a control button to initiate to initiate playback of the video. This ensured that their mouse was in a central position relative to the two items as the video begins. The item descriptions always began with a 2 second pause during which both images were visible, before the video could be clicked for playback. This allowed the participant to take note of each image option before directing their attention to the typing video. In this initial still image of the video, the name of the typist was also visible, which ensured that all the necessary information (which typist they will be viewing and what both image options looked like) was available before the video clip began.

Data Issues

The combination of the time sensitive mouse-tracker and the experimental design in which participants had to watch a seconds-long video clip during their selection process led to several errors in data collection. More specifically, during the initial set of participants run through this study (N=35), approximately 70% of the data was lost or corrupted in what appeared to be an overflow of tracked mouse movements. As mentioned earlier, the Qualtrics mouse tracker modification records the position of the mouse cursor every 10 milliseconds, which in combination with the significantly longer duration of the video compared to the Mathur study the open source JavaScript modification was designed for, created a data overload. This resulted in data that was ultimately unusable because Qualtrics would stop writing the mouse-tracking coordinates to the datafile when this happened. This resulted in unpredictable data loss because this occurred at a different point for each participant, depending on how quickly they generally responded on each trial.

To address this problem, we attempted to introduce several modifications aimed at speeding up the amount of time participants spent on each item. We began by modifying the message participants received at the beginning of the experiment to emphasize that they should make their selection as quickly as possible. We also added a notice that the task was speed based and an explicit statement that participants should not wait until the end of the video to make their choice. We ran an additional 22 participants following these changes. However, given the length of the videos and the fact that the item description was often only completed towards the end of each clip, participants continued to make their selection too late for the data to record correctly. This modification reduced some of the data that was lost, though not significantly enough to reduce the overload recording error that interfered with the functionality of the mouse tracker.

Finally, we introduced a time-limit for each question to force participants to respond within ten seconds. However, to accommodate the length of each video clip the time limit could not be as low as was necessary to resolve the problem of the mouse tracking's recording capacity. Ultimately the nature of the task was fundamentally incompatible with the mouse-tracker addition that could be included into Qualtrics. Requiring participants to view a video (as opposed to responding quickly to a static image) generated too much data for the mouse tracker to record properly. Unfortunately, as a result no useable data was generated from this experiment.

References

Abercrombie, D. (1968). Paralanguage. British journal of disorders of communication, 3(1), 55-59.

Aldridge, M., & Fontaine, L. (2019). Using keystroke logging to capture the impact of cognitive complexity and typing fluency on written language production. In E. Lindgren & K. Sullivan (Eds.), *Observing writing: Insights from keystroke logging and handwriting* (p. 285-305). Leiden, the Netherlands: Koninklijke Brill NV.

Anderson, J. F., Beard, F. K., & Walther, J. B. (2007). Turn-taking and the local management of conversation in a highly simultaneous computer mediated communication system. *Language@Internet*, *7*, article 7.

Arciuli, J., Mallard, D., & Villar, G. (2010). "Um, I can tell you're lying": Linguistic markers of deception versus truth-telling in speech. Applied Psycholinguistics, 31(3), 397-411.

Arnold, J. E., Hudson Kam, C., & Tanenhaus, M. (2007). If you say thee uh you are describing something hard: The on-line attribution of disfluency during reference comprehension. Journal of Experimental Psychology: Learning, Memory and Cognition, 33, 914□930.

Argyle, M. (1972). Non-verbal communication in human social interaction.

Banerjee, R., Feng, S., Kang, J. S., & Choi, Y. (2014). Keystroke patterns as prosody in digital writings: A case study with deceptive reviews and essays. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 1469-1473).

Barr, D. J. (2001). Trouble in mind: Paralinguistic indices of effort and uncertainty in communication. Oralité et gestualité: Interactions et comportements multimodaux dans la communication, 597-600. Paris: L'Harmattan.

Bavelas, J. B., & Chovil, N. (2006). Nonverbal and Verbal Communication: Hand Gestures and Facial Displays as Part of Language Use in Face-to-face Dialogue.

Beattie, G. W., & Butterworth, B. L. (1979). Contextual probability and word frequency as determinants of pauses and errors in spontaneous speech. Language and speech, 22(3), 201-211.

Benus, S., Enos, F., Hirschberg, J., & Shriberg, E. (2006). Pauses in deceptive speech. Paper presented at the Third Speech Prosody Conference, Dresden, Germany.

Biancardi, B., Cafaro, A., & Pelachaud, C. (2017, November). Analyzing first impressions of warmth and competence from observable nonverbal cues in expert-novice interactions. In Proceedings of the 19th ACM International Conference on Multimodal Interaction (pp. 341-349). ACM.

Bilous, F. R. (1993). The role of gestures in speech production: Gestures enhance lexical access.

Birnholtz, J., & Ibara, S. (2012, February). Tracking changes in collaborative writing: edits, visibility and group maintenance. In Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work (pp. 809-818). ACM.

Blickensderfer, E., Cannon-Bowers, J. A., & Salas, E. (1997). Training teams to self-correct: An empirical investigation. In 12th annual meeting of the Society for Industrial and Organizational Psychology, St. Louis, MO.

Bolinger, D. L. (1961). Contrastive accent and contrastive stress. Language, 37(1), 83-96.

Bolstad, C. A., & Endsley (1999). Shared mental models and shared displays: An empirical evaluation of team performance. In Proceeding of the 43rd meeting of the Human Factors and Ergonomics Society (pp. 213-217). Santa Monica, CA: Human Factors and Ergonomics Society Press.

Borràs-Comes, J., Roseano, P., Vanrell, M. M., & Prieto, P. (2011). Perceiving uncertainty: Facial gestures, intonation, and lexical choice. In C. Kirchhof, Z. Maliszm, & P. Wagner (Eds.), In *Proceedings of the 2nd Conference on Gesture and Speech in Interaction* [GESPIN 2011]. Bielefeld, Germany: Bielefeld University.

Brennan, S. E. & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. Journal of Experimental Psychology: Learning, Memory, and Cognition, Vol 22(6). 1482-1493.

Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. Journal of Memory and Language, 44(2), 274-296.

Brennan, S. E., & Williams, M. (1995). The feeling of another′ s knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. Journal of memory and language, 34(3), 383-398.

Brown, S., Fuller, R., & Thatcher, S. M. (2016). Impression formation and durability in mediated communication. Journal of the Association for Information Systems, 17(9), 1.

Calhoun, C. (1991). Indirect relationships and imagined communities: Large scale social integration and the transformation of everyday life. In J. Coleman & P. Bourdieu (Eds.), Social theory for a changing society (pp. 95–121). Boulder, CO: Westview.

Cannon-Bowers, J. E., Salas, E., & Converse, S. (1993). "Shared Mental Models in Expert Team Decision-Making". In Castellan, J. (Ed.), Individual and Group Decision-Making: Current Issues: 221: LEA Publishers.

Cannon-Bowers, J. E. & Salas, E. (2001). "Reflections on Shared Cognition". Journal of Organizational Behavior, 22(March): 195-202.

Cerratto, T. (1999). Instrumenting Collaborative Writing and its cognitive tools. In Proceedings HCP 99 Conference, Human Centred Processes (pp. 24-26).

Chawla, P., & Krauss, R. M. (1994). Gesture and speech in spontaneous and rehearsed narratives. Journal of Experimental Social Psychology, 30(6), 580-601.

Christenfeld, N., Schachter, S., & Bilous, F. (1991). Filled pauses and gestures: It's not coincidence. Journal of Psycholinguistic Research, 20(1), 1-10.

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. Cognition, 84(1), 73-111.

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10-63). Cambridge, England: Cambridge University Press.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22,* 1-39.

Clark, H. H. (1996). Using language. Cambridge: University Press.

Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. Cognitive psychology, 37(3), 201-242.

Clifton, C., Jr., & Frazier, L. (2018). Evaluation of the epistemic state of the speaker/author. *Quarterly Journal of Experimental Psychology*, *71*, 1482-1492.

Conijn, R., Roeser, J., & van Zaanen, M. (2019). Understanding the keystroke log: The effect of writing task on keystroke features. *Reading and Writing: An Interdisciplinary Journal, 32*, 2353–2374

Corley, M., MacGregor, L. J., & Donaldson, D. I. (2007). It's the way that you, er, say it: Hesitations in speech affect language comprehension. Cognition, 105(3), 658-668.

Couture, B., & Rymer J. (1991): Discourse interaction between writer and supervisor: A primary collaboration in workplace writing. In M.M. Lay and W.M. Karis (eds.): Collaborative Writing in Industry: Investigations in Theory and Practice. Amityville, NY: Baywood Publishing Company, pp. 87–108.

Crowston, K., & Kammerer, E. E. (1998). Coordination and collective mind in software requirements development. IBM Systems Journal, 37(2), 227-245.

Crystal, D. (1970). Prosodic systems and language acquisition. Prosodic feature analysis, 77-90.

Dabbish, L.A., Kraut, R.E., Fussell, S., & Keisler, S. (2005). Understanding email use: Predicting action on a message. Proceedings of the SIGCHI conference on human factors in computing systems (CHI ’05), pp. 691-700. New York: ACM Press.

Daft, R. L., & Lengel, R. H. (1984). Information richness: A new approach to managerial information processing. Research in organizational behavior, 6, 191-233.

Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. Management Science, 32(5), 554–571.

Daft, R. L., Lengel, R. H., & Trevino, L. K. (1987). Message equivocality, media selection, and manager performance: Implications for information systems. MIS Quarterly, 11(3), 355–366.

Davenport, T. H., & Beck, J. C. (2001). The attention economy: Understanding the new currency of business. Harvard Business Press.

Dennis, A. R., & Kinney, S. T. (1998). Testing media richness theory in the new media: The effects of cues, feedback, and task equivocality. Information systems research, 9(3), 256-274.

DeSanctis, G., & Poole, M. S. (1994). Capturing the complexity in advanced technology use: Adaptive structuration theory. Organization science, 5(2), 121-147.

Dijkstra, C., Krahmer, E., & Swerts, M. (2006). Manipulating uncertainty: The contribution of different audiovisual prosodic cues to the perception of confidence. In R. Hoffmann & H. Mixdoff (Eds.), In *Proceedings of the Third International Conference on Speech Prosody* (1–4). Dresden, Germany: TUDpress.

Dillon, A. (1993). How collaborative is collaborative writing? An analysis of the production of two technical reports. In Computer supported collaborative writing (pp. 69-85). Springer, London.

Dillon, A., & Maynard, S. (1995). ‘Don't forget to put the cat out’—or why collaborative authoring software and everyday writing pass one another by!. New Review of Hypermedia and Multimedia, 1(1), 135-153.

Dourish, P., & Bellotti, V. (1992, November). Awareness and coordination in shared workspaces. In CSCW (Vol. 92, No. 3, pp. 107-114).

Druskat, V. U., & Pescosolido, A. T. (2002). The content of effective teamwork mental models in self-managing teams: Ownership, learning and heedful interrelating. Human Relations, 55(3), 283-314.

Ekman, P., & Oster, H. (1979). Facial expressions of emotion. Annual review of psychology, 30(1), 527-554.

El Kaliouby, R., & Robinson, P. (2005). Real-time inference of complex mental states from facial expressions and head gestures. In Real-time vision for human-computer interaction (pp. 181-200). Springer, Boston, MA.

Endsley, M. R. (1995). Measurement of situation awareness in dynamic systems. Human factors, 37(1), 65-84.

Endsley, M. R., & Jones, D. G. (2001, October). Disruptions, interruptions and information attack: impact on situation awareness and decision making. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 45, No. 2, pp. 63-67). Sage CA: Los Angeles, CA: SAGE Publications.

Englund Dimitrova, B. (2006). Segmentation of the writing process in translation: Experts versus novices. In K.P.H. Sullivan & E. Lundgren (Eds.), *Computer Keystroke Logging and Writing: Methods and applications* (pp. 189-201). Oxford: Elsevier.

Entin, E. B., & Entin, E. E. (2000). Assessing team situation awareness in simulated military missions. In Proceeding of the Human Factors and Ergonomics Society 44th Annual Meeting (pp. 73-77). San Diego, CA: Human Factors and Ergonomics Society Press.

Entin, E., & Serfaty, D. (1999). "Adaptive team coordination", Human Factors, 41, 312–325.

Entin, E., Serfaty, D., & Deckert, J. (1994). "Team adaptation and coordination training" (Report No. TR-648-1). Burlington, MA:ALPHATECH.

Epp, C., Lippold, M., & Mandryk, R. L. (2011, May). Identifying emotional states using keystroke dynamics. In Proceedings of the sigchi conference on human factors in computing systems (pp. 715-724). ACM.

Erkens, G., Jaspers, J., Prangsma, M., & Kanselaar, G. (2005). Coordination processes in computer supported collaborative writing. Computers in Human Behavior, 21(3), 463-486.

Feng, J., Lazar, J., & Preece, J. (2004). Empathy and online interpersonal trust: A fragile relationship. Behaviour & Information Technology, 23(2), 97-106.

Fiore, S. M., Salas, E., & Cannon-Bowers, J. A. (2001). Group dynamics and shared mental model development. In M. London (Ed.), How people evaluate others in organizations: Person perception and interpersonal judgment in industrial/organizational psychology (pp. 309-336). Mahwah, NJ: Erlbaum.

Fish, R. S., Kraut, R. E., & Leland, M. D. (1988, April). Quilt: a collaborative tool for cooperative writing. In ACM SIGOIS Bulletin (Vol. 9, No. 2-3, pp. 30-37). ACM.

Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. Trends in cognitive sciences, 11(2), 77-83.

Fox Tree, J. E. (2001). Listeners' uses of um and uh in speech comprehension. Memory & cognition, 29(2), 320-326.

Foushee, H. C. (1984). Dyads and triads at 35,000 feet: Factors affecting group process and aircrew performance. American Psychologist, 39(8), 885.

Fulk, J., Steinfield C. W., Schmitz J., Power J. G., "A Social Information Processing Model of Media Use in Organizations", Communication Research, vol. 14, pp. 529-552, 1987.

Gabarro, J. J. (1987). The dynamics of taking charge. Harvard Business Press.

Gardner, W. L., Fischer, D., & Hunt, J. G. J. (2009). Emotional labor and leadership: A threat to authenticity?. The Leadership Quarterly, 20(3), 466-482.

Gerwing, J., & Bavelas, J. (2004). Linguistic influences on gesture's form. Gesture, 4(2), 157-195.

Grice, H. P. (1975). Logic and Conversation. In D. Davidson & G. Harman (Eds.), *The Logic of Grammar* (pp. 64-75). Encino, CA: Dickenson, 64-75.

Gutwin, C., & Greenberg, S. (1997). Effects of awareness support on groupware usability.

Gutwin, C., Greenberg, S. and Roseman, M. Short Paper: Supporting Awareness of Others in Groupware. In Conference Companion of the Conference on Human Factors in Computing Systems - CHI'96 (Apr. 13-18, Vancouver. Canada). ACM. N.Y.. 1996. PP. 205.

Haase, R. F., & Tepper, D. T. (1972). Nonverbal components of empathic communication. Journal of counseling psychology, 19(5), 417.

Hancock, J. T., & Dunham, P. J. (2001). Impression formation in computer-mediated communication revisited: An analysis of the breadth and intensity of impressions. Communication research, 28(3), 325-347.

Hancock, J. T., & Dunham, P. J. (2001). Language use in computer-mediated communication: The role of coordination devices. Discourse Processes, 31(1), 91-110.

Harbers, M., Riemsdijk, M. V., & Jonker, C. M. (2012). Measuring sharedness of mental models and its relation to team performance. In Proceedings 14th International Workshop on Coordination, Organisations, Institutions and Norms (pp. 106-120).

Harbers, M., Jonker, C.M., Van Riemsdijk, M.B.: Context-sensitive sharedness criteria for teamwork. In: Proceedings of AAMAS 2014, pp. 1507–1508. ACM (2014)

Hinde, R. A. (Ed.). (1972). Non-verbal communication. Cambridge University Press.

Hinds, P. & Weisband, S. P. (2003). Knowledge sharing and shared understanding in virtual teams. In C. B. Gibson & S. G. Cohen (Eds.), Virtual Teams That Work (pp. 21-36). San Francisco, CA: Jossey-Bass.

Holler, J., Kendrick, K. H., Casillas, M., & Levinson, S. C. (Eds.) (2015). Turn-taking in human communicative interaction. Frontiers in Psychology, 6, 1919.

Hollingshead, A. B., McGrath, J. E., & O'Conner, K. M. (1993). Group task performance and communication technology: A longitudinal study of computer-mediated versus face-to-face work groups. Small Group Research,24, 307-333.

Hübscher, I., Esteve-Gibert, N., Igualada, A., & Prieto, P. (2017). Intonation and gesture as bootstrapping devices in speaker uncertainty. *First Language*, *37*, 24-41.

Hupet, M., Seron, X., & Chantraine, Y. (1991). The effects of the codability and discriminability of the referents on the collaborative referring procedure. *British Journal of Psychology*, *82*, 449–462.

Jehn, K. A., & Mannix, E. A. (2001). The dynamic nature of conflict: A longitudinal study of intragroup conflict and group performance. Academy of management journal, 44(2), 238-251.

Jonsson, E. (2015). *Conversational writing: A multidimensional study of synchronous and supersynchronous computer-mediated communication*. Frankfurt am Main: Peter Lang

Kalman, Y. M., & Gergle, D. (2014). Letter repetitions in computer-mediated communication: A unique link between spoken and online language. Computers in Human Behavior, 34, 187-193.

Kalman, Y. M., & Rafaeli, S. (2011). Online pauses and silence: Chronemic expectancy violations in written computer-mediated communication. Communication Research, 38(1), 54-69.

Kalman, Y. M., Scissors, L. E., Gill, A. J., & Gergle, D. (2013). Online chronemics convey social information. Computers in Human Behavior, 29(3), 1260-1269.

Kiesler, S. (1986). The hidden messages in computer networks (pp. 46-47). Harvard Business Review Case Services.

Kiesler, S., Siegel, J., & McGuire, T. W. (1984). Social psychological aspects of computer-mediated communication. American Psychologist, 39, 1123–1134.

Kim, H. C. E., & Eklundh, K. S. (2001). Reviewing practices in collaborative writing. Computer Supported Cooperative Work (CSCW), 10(2), 247-259.

Kittur, A., & Kraut, R. E. (2008, November). Harnessing the wisdom of crowds in wikipedia: quality through coordination. In Proceedings of the 2008 ACM conference on Computer supported cooperative work (pp. 37-46). ACM.

Kleinman, D. L., & Serfaty, D. (1989, April). Team performance assessment in distributed decision making. In Proceedings of the symposium on interactive networked simulation for training (pp. 22-27). Orlando, FL: University of Central Florida.

Klimoski, R., & Mohammed, S. (1994). Team mental model: Construct or metaphor?. Journal of management, 20(2), 403-437.

Kraut, R., Galegher, J., Fish, R., & Chalfonte, B. (1992). Task requirements and media choice in collaborative writing. Human-Computer Interaction, 7(4), 375-407.

Krishnan, J., Cusimano, A., Wang, D., & Yim, S. (2018). Writing together: Online synchronous collaboration in middle school. Journal of Adolescent & Adult Literacy, 62(2), 163-173.

Kukreja, U., Stevenson, W. E., & Ritter, F. E. (2006). RUI: Recording user input from interfaces under Windows and Mac OS X. *Behavior Research Methods, 38*, 656–659.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). *lmerTest* package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 1–26.

Lea, M., & Spears, R. (1992). Paralanguage and social perception in computer-mediated communication. Journal of Organizational Computing, 2, 321–341.

Lea, M., & Spears, R. (1995). Love at first byte? Building personal relationships over computer networks.

Lee, A. S. 1994. "Electronic Mail as a Medium for Rich Communication: An Empirical Investigation Using Hermeneutic Interpretation," MIS Quarterly (18:2), pp. 143-157.

Levelt, W. J. M. (1989). ACL.

Libby, R., Trotman, K. T., & Zimmer, I. (1987). Member variation, recognition of expertise, and group-performance. Journal of Applied Psychology, 72, 81-87. doi:10.1037/0021-9010.72.1.81

Littlepage, G. E., Hollingshead, A. B., Drake, L. R., & Littlepage, A. M. (2008). Transactive memory and performance in work groups: Specificity, communication, ability differences, and work allocation. Group Dynamics, 12, 223-241.

Littlepage, G., Robison, W., & Reddington, K. (1997). Effects of task experience and group experience on group performance, member ability, and recognition of expertise. Organizational Behavior & Human Decision Processes, 69 (2), 133-147.

Lowry, P. B., Curtis, A., & Lowry, M. R. (2004). Building a taxonomy and nomenclature of collaborative writing to improve interdisciplinary research and practice. The Journal of Business Communication (1973), 41(1), 66-99.

Lowry, P. B., & Nunamaker, J. F. (2003). Using Internet-based, distributed collaborative writing tools to improve coordination and group awareness in writing teams. IEEE Transactions on Professional Communication, 46(4), 277-297.

Loy, J. E., Rohde, H., & Corley, M. (2017). Effects of disfluency in online interpretation of deception. *Cognitive Science*, *41*, 1434-1456.

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*, 1494-1502.

Lysander, K., & Horton, W. S. (2012). Conversational grounding in younger and older adults: The effect of partner visibility and referent abstractness in task-oriented dialogue. *Discourse Processes*, *49*, 29-60.

MacMillan, J., Entin, E. E., & Serfaty, D. (2004). Communication overhead: The hidden cost of team cognition. Team cognition: Process and performance at the interand intra-individual level. American Psychological Association, Washington, DC.

Malone, T. W., & Crowston, K. (1994). The interdisciplinary study of coordination. ACM Computing Surveys (CSUR), 26(1), 87-119.

March, J. & Simon, H. (1958). Organizations: John Wiley and Sons.

Maricchiolo, F., Gnisci, A., Bonaiuto, M., & Ficca, G. (2009). Effects of different types of hand gestures in persuasive speech on receivers' evaluations. Language and cognitive processes, 24(2), 239-266.

Marlow, J., Dabbish, L., & Herbsleb, J. (2013, February). Impression formation in online peer production: activity traces and personal profiles in github. In Proceedings of the 2013 conference on Computer supported cooperative work (pp. 117-128). ACM.

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. Journal of applied psychology, 85(2), 273.

McGrath, J. E. (1990). Time matters in groups. In J. Galegher, R. E. Kraut, & C. Egido (Eds.), Intellectual teamwork (pp. 23-61). Hillsdale, NJ: Lawrence Erlbaum.

McNeill, D. (1992). Hand and mind: What gestures reveal about thought. University of Chicago press.

Mendoza-Chapa, S., Romero-Salcedo, M., & Oktaba, H. (2000, October). Group awareness support in collaborative writing systems. In Proceedings Sixth International Workshop on Groupware. CRIWG 2000 (pp. 112-118). IEEE.

Mills, G. J. (2014). Dialogue in joint activity: Complementarity, convergence and conventionalization. New ideas in psychology, 32, 158-173.

Mitchell, A., Posner, I., & Baecker, R. (1995, May). Learning to write together using groupware. In CHI (Vol. 95, pp. 288-295).

Neuwirth, C., Chandhok, R., Kaufer, D., Erion, P., Morris, J. and Miller, D. (1992). Flexible Diff-ing In A Collaborative Writing System. Proceedings of the ACM '92 (Toronto ON, 1992) ACM Press, 147 - 154.

Nguyen, L. S., & Gatica-Perez, D. (2015, November). I would hire you in a minute: Thin slices of nonverbal behavior in job interviews. In Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (pp. 51-58). ACM.

Olson, G. M., Olson, J. S., Carter, M. R., & Storrøsten, M. (1992). Small group design meetings: An analysis of collaboration. Human-Computer Interaction, 7(4), 347-374.

Orasanu, J. (1990). Shared mental models and crew decision making (CSL Report No. 46). Princeton, NJ: Princeton University, Cognitive Science Laboratory.

Oser, R. L., Prince, C., Morgan Jr, B. B., & Simpson, S. S. (1991). An analysis of aircrew communication patterns and content (No. NTSC-TR-90-09). NAVAL TRAINING SYSTEMS CENTER ORLANDO FL.

Paul, S., He, F., & Dennis, A. R. (2018). Group Atmosphere, Shared Understanding, and Team Conflict in Short Duration Virtual Teams.

Pietrandrea, P. (2002). Iconicity and arbitrariness in Italian sign language. Sign Language Studies, 2(3), 296-321.

Poole, M. S., Holmes, M., & DeSanctis, G. (1991). Conflict management in a computer-supported meeting environment. Management Science, 37(8), 926-953.

Postma, A., & Kolk, H. (1993). The covert repair hypothesis: Prearticulatory repair processes in normal and stuttered disfluencies. Journal of Speech, Language, and Hearing Research, 36(3), 472-487.

Postmes, T., Spears, R., & Lea, M. (1998). Breaching or building social boundaries? SIDE-effects of computer-mediated communication. Communication research, 25(6), 689-715.

Rice, R. E. (1992). Task analyzability, use of new media, and effectiveness: A multi-site exploration of media richness. Organization science, 3(4), 475-500.

Rico, R., Sanchez-Manzanares, M., Gil, F., & Gibson, C. (2008). Team implicit coordination processes: A team knowledge-based approach. Academy of Management Journal, 33, 163–184. doi:10.5465/ AMR.2008.27751276

Riordan, M. A. (2017). Emojis as tools for emotion work: Communicating affect in text messages. *Journal of Language and Social Psychology*, *36*, 549-567.

Roseano, P., Gonzàlez, M., Borràs-Comes, J., & Prieto, P. (2016). Communicating epistemic stance: How speech and gesture patterns reflect epistemicity and evidentiality. *Discourse Processes*, *53*, 135-174.

Rutter, D. R., & Stephenson, G. M. (1979). The role of visual communication in social interaction. Current Anthropology, 20(1), 124-125.

Sauter, D. A., Eisner, F., Calder, A. J., & Scott, S. K. (2010). Perceptual cues in nonverbal vocal expressions of emotion. The quarterly journal of experimental psychology, 63(11), 2251-2272.

Salas, E., Fowlkes, J., Stout, R., Milanovich, D., & Prince, C. (1999). Does CRM training improve teamwork skills in the cockpit?: Two evaluation studies. Human Factors, 41, 326–343.

Serfaty, D., Entin, E., & Deckert, J. (1993). Team adaptation to stress in decision making and coordination with implications for CIC team training (Report No. TR-564, Vol. 1/2). Burlington, MA: ALPHATECH.

Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. Journal of Cross-cultural psychology, 32(1), 76-92.

Schiffer, S. R. (1972). *Meaning*. Oxford: Oxford University Press.

Schnadt, M. J., & Corley, M. (2006). The influence of lexical, conceptual and planning based factors on disfluency production. In Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 28, No. 28).

Shah, J., & Breazeal, C. (2010). An empirical analysis of team coordination behaviors and action planning with application to human–robot teaming. Human factors, 52(2), 234-245.

Sheldon, O. J., Thomas-Hunt, M. C., & Proell, C. A. (2006). When timeliness matters: The effect of status on reactions to perceived time delay within distributed collaboration. Journal of Applied Psychology, 91(6), 1385.

Singmann, H., & Kellen, D. (2019). An introduction to mixed models for experimental psychology. In D. H. Spieler & E. Schumacher (Eds.), *New Methods in Cognitive Psychology* (pp. 4-31). Psychology Press.

Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. Journal of memory and language, 32(1), 25-38.

Spears, R., & Lea, M. (1994). Panacea or panopticon? The hidden power in computer-mediated communication. Communication Research, 21(4), 427-459.

Stout, R., Cannon-Bowers, J., & Salas, E. (1996). The role of shared mental models in developing team situational aware- ness: Implications for training. Training Research Journal, 2, 86–116.

Stout, R., Cannon-Bowers, J., Salas, E., & Milanovich, D. (1999). Planning, shared mental models, and coordinated performance: An empirical link is established. Human Factors, 41, 61–71.

Swerts, M., & Krahmer, E. (2005). Audiovisual prosody and feeling of knowing. Journal of Memory and Language, 53(1), 81-94.

Tam, J., & Greenberg, S. (2004, September). A framework for asynchronous change awareness in collaboratively-constructed documents. In International Conference on Collaboration and Technology (pp. 67-83). Springer, Berlin, Heidelberg.

Tam, J., & Greenberg, S. (2006). A framework for asynchronous change awareness in collaborative documents and workspaces. International Journal of Human-Computer Studies, 64(7), 583-598.

Tidwell, L. C., & Walther, J. B. (2002). Computer-mediated communication effects on disclosure, impressions, and interpersonal evaluations: Getting to know one another a bit at a time. Human communication research, 28(3), 317-348.

Toma, C. L., & Hancock, J. T. (2010). Looks and lies: The role of physical attractiveness in online dating self-presentation and deception. Communication Research, 37(3), 335-351.

Tomlinson Jr, J. M., & Tree, J. E. F. (2011). Listeners' comprehension of uptalk in spontaneous speech. Cognition, 119(1), 58-69.

Treem, J. W. (2013). Technology use as a status cue: The influences of mundane and novel technologies on knowledge assessments in organizations. Journal of Communication, 63(6), 1032-1053.

Trichas, S., & Schyns, B. (2012). The face of leadership: Perceiving leaders from facial expression. The Leadership Quarterly, 23(3), 545-566.

Trichas, S., Schyns, B., Lord, R., & Hall, R. (2017). “Facing” leaders: Facial expression and leadership perception. The Leadership Quarterly, 28(2), 317-333.

Trigg, R. H., & Suchman, L. A. (1989). Collaborative writing in NoteCards. Hypertext: theory into practice, 45-61.

Utz, S. (2000). Social information processing in MUDs: The development of friendships in virtual worlds. *Journal of Online Behavior*, 1.

Valacich, J. S., Mennecke, B. E., Wachter, R. M., & Wheeler, B. C. (1994). Extensions to media richness theory: A test of the task-media fit hypothesis. In Proceedings of the 27th Hawaii International Conference on System Sciences (HICSS-27). Part 4 (of 5) (pp. 11-20). Publ by IEEE.

Vertegaal, R. (1997, March). Conversational awareness in multiparty VMC. In CHI'97 Extended Abstracts on Human Factors in Computing Systems (pp. 6-7). ACM.

Vizer, L. M., Zhou, L., & Sears, A. (2009). Automated stress detection using keystroke and linguistic features: An exploratory study. International Journal of Human-Computer Studies, 67(10), 870-886.

Vrij, A., Edward, K., Roberts, K. P., & Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. Journal of Nonverbal behavior, 24(4), 239-263.

Walther, J. B. (1992). Interpersonal effects in computer-mediated interaction: A relational perspective. Communication research, 19(1), 52-90.

Walther, J. B., & Parks, M. R. (2002). Cues filtered out, cues filtered in. Handbook of interpersonal communication, 3, 529-563.

Wang, D., Tan, H., & Lu, T. (2017). Why users do not want to write together when they are writing together: Users' rationales for today's collaborative writing practices. Proceedings of the ACM on Human-Computer Interaction, 1(CSCW), 107.

Weick, K. (1990). "The Vulnerable System: An Analysis of the Tenerife Air Disaster". Journal of Management, 16(3): 571-593.

Weick, K. (1993). "The Collapse of Sensemaking in Organizations: The Mann Gulch Disaster". Administrative Science Quarterly, 38(4): 628-652.

Weick, K. & Roberts, K. (1993). "Collective Mind in Organizations: Heedful Interrelating on Flight Decks". Administrative Science Quarterly, 38(3): 357-381.

Wengelin, A. (2006). Examining pauses in writing: Theory, methods and empirical data. In K.P.H. Sullivan & E. Lundgren (Eds.), *Computer Keystroke Logging and Writing: Methods and applications* (pp. 107-130). Oxford: Elsevier.

Wittenbaum, G. M., Stasser, G., & Merry, C. J. (1996). Tacit coordination in anticipation of small group task completion. Journal of Experimental Social Psychology, 32(2), 129-152.

Yeh, H. C. (2014). Exploring how collaborative dialogues facilitate synchronous collaborative writing. Language Learning & Technology, 18(1), 23-37.

Yus, F. (2005). Attitudes and emotions through written text: The case of textual deformation in Internet chat rooms. *Pragmalingüística*, *13*, 147-174.