A Network Analytic Approach to Gaze Coordination during a Collaborative Task

A critical component of collaborative learning is the establishment of intersubjectivity, or the construction of mutual understanding. Collaborators coordinate their understanding with one another across various modes of communication, including speech, gesture, posture, and gaze. Given the dynamic, interdependent, and complex nature of coordination, this study sought to develop and test a method for constructing detailed and nuanced models of coordinated referential gaze patterns. In the study, 13 dyads participated in a simple collaborative task. We used dual mobile eye tracking to record each participant's gaze behavior, and we used epistemic network analysis (ENA) to model the gazes of both conversational participants synchronously. In the model, the nodes in the network represent gaze targets for each participant, and the connections between nodes indicate the likelihood of gaze coordination. Our analyses indicate: (a) properties and patterns of how gaze coordination unfolds throughout an interaction sequence; and (b) differences in gaze coordination patterns for interaction sequences that lead to breakdowns and repairs. In addition to contributing to the growing body of knowledge on the coordination of gaze behaviors in collaborative activities, this work suggests that ENA enables more effective modeling of gaze coordination.

Keywords: Epistemic network analysis; collaborative learning; conversational repair; eye tracking; gaze behavior

Introduction

A critical component of successful collaborative learning is the establishment of intersubjectivity, or the construction of mutual understanding. Key to successful development of mutual understanding during collaborative learning activities is *coordination*: participants coordinate whose turn it is to speak (Sacks et al., 1974), and they use actions such as pointing, placing, gesturing, and gazing to coordinate attention (Clark, 2003; Clark and Krych, 2004). Coordinating turns of talk and attention thus ensures that joint activities flow easily and intelligibly (Clark, 1996; Garrod and Pickering, 2004).

Of particular importance to successful learning interactions is the coordination of gaze and attention across a shared visual reference space (Brown-Schmidt, Campana, & Tanenhaus, 2005; Clark, 1996; Clark and Brennan, 1991; Schober, 1993). Gaze coordination describes the coupling of gaze patterns (Richardson, Dale, & Tomlinson, 2009), which arises not from an explicit attempt to synchronize gaze movements but from the gradual alignment of gaze patterns over time due to the nature of interaction in a joint activity.

Mechanisms of gaze coordination, including mutual gaze and joint attention, have been revealed to be a primary instrument of prelinguistic learning between infants and caregivers (Baldwin, 1995), and they play a crucial role in coordinating conversations and collaborative interaction more generally (Bavelas, Coates, & Johnson, 2002).

In this paper, we explore the application of a new analytic approach to the challenge of modeling the creation of intersubjectivity through gaze coordination. We argue that coordination in collaborative activity is constructed, in part, through *reference-action sequences*: short interactions between collaborators in which one person indicates an object in the collaborative workspace that another person is supposed to manipulate in some way. In some collaborative learning tasks, the objects and desired manipulations are abstract, such as a request to solve some algebraic problem. But in many tasks, the objects are explicitly present in the shared environment, either as actual objects or as notations on paper or in digital form.

The phenomenon we focus on here is the extent to which in face-to-face collaboration, these reference-action sequences are accompanied by the development of gaze coordination—that is, the extent to which as collaborators communicate about parts of a task to attend to, they use gaze coordination to facilitate their communication.

The premise of our investigation is that gaze is a form of communication through which participants indicate, clarify, or amplify meaning to one another. In other words, we can think of gaze patterns as similar to discourse, or a sequence of speech acts in a conversation, such that each participant's gaze pattern is best understood as resulting from interactions with the gaze patterns of other participants. To model gaze coordination, then, we use a technique from learning analytics called *epistemic network analysis* (ENA), which measures, visualizes, and enables both quantitative and qualitative comparison of complex, collaborative interactions (Shaffer, 2017; Shaffer et al., 2009; Shaffer, Collier, & Ruis, 2016; Shaffer & Ruis, 2017).

Developing effective models of gaze coordination is of particular importance for the design of cognitive tutors, virtual characters, and other pedagogical agents, including social robots (Karaman & Sezgin, 2018). While researchers have developed intelligent tutors that use gaze tracking to identify inattentiveness or disengagement in a variety of contexts (e.g., D'Mello, Olney, Williams, & Hays, 2012; Hutt, Mills, White, Donnelly, & D'Mello, 2016; Jaques, Conati, Harley, & Azevedo, 2014; Szafir & Mutlu, 2012), these applications are strictly reactionary. Improvements in intelligent tutoring will depend not on reacting to student behavior but on the extent to which the system can proactively encourage behaviors conducive to learning, including collaboration (Hayashi, 2016; Huang & Mutlu, 2016; Olsen, Aleven, & Rummel, 2016). To do this effectively, however, requires understanding how interlocutors establish and maintain gaze coordination and developing effective techniques for modeling it.

To examine gaze coordination and explore the affordances of ENA for modeling gaze coordination, we collected data from 13 dyads outfitted with mobile eye-tracking glasses. Each dyad was assigned a sandwich-building task: one participant made verbal references to visible

ingredients they would like added to their sandwich, while the other participant assembled those ingredients into a sandwich. We chose this task to represent collaborative interactions that contain a large number of reference-action sequences, and because it exemplifies a large class of learning situations in which one person asks another to take some action and then has to provide feedback on the results. Thus, while we are not specifically studying a classroom learning task, we believe that the results of the analyses here will generalize to many interactions that involve reference-action sequences of the kind that are commonly found in learning interactions.

To analyze and visualize the gaze targets of both participants as a complex and dynamic network of semiotic relationships, we conducted three separate analyses of the dyadic gaze data using ENA. In the first analysis, we used ENA to characterize different phases of a reference-action sequence, identifying clear differences in gaze behavior at each phase. This analysis also revealed a consistent pattern of gaze behavior that progresses in an orderly and predictable fashion throughout a reference-action sequence. In the second analysis, we explored the progression of gaze alignment between the two interacting participants throughout a reference-action sequence. In general, we identified a common rise and fall in the amount of aligned gaze throughout a sequence as well as a back and forth pattern of which participant's gaze "led" the other's. In the third analysis, we explored the difference in gaze behaviors that arose during sequences with *repairs* (responses to confusion or requests for clarification) versus sequences without repairs. ENA revealed different patterns of gaze behavior for these two types of sequence, even in very early phases of the sequences before any repair occurs.

Our results suggest that a network-based approach to gaze analysis is a useful way to model gaze during collaborative tasks, such as those commonly supported by intelligent tutoring systems. Gaze coordination is an important measure of the extent to which the collaborating individuals, such as a tutor and student, are able to construct a shared understanding. In addition to contributing to the growing body of knowledge on the coordination of gaze behaviors in collaborative activities, we believe this work has implications for the design and implementation of technologies that support collaborative learning, whether supporting gaze coordination among learners or simulating gaze by computer-controlled pedagogical agents, such as automated tutors or virtual characters.

Background

Prior research on gaze has focused primarily on the eye movements of speakers and listeners in isolation, which has shown that people typically attend to (a) what is being discussed (Griffin, 2004; Meyer, van der Meulen, & Brooks, 2004), (b) things specifically referenced (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), and (c) things about to be referenced (Altmann & Kamide, 2004). When collaborating in a shared space, conversational partners use each others' gaze to gauge attention and comprehension (Gergle & Clark, 2011). As a result, collaborators are more likely to exhibit gaze coordination when they discuss specific referents in the space (Arai, Bard, & Hill, 2009). Referencing is usually multimodal: speakers use various actions and other

contextual cues, such as gestures or head nods, to indicate the referent. However, speakers often under-specify their referents, relying on the listener to seek clarification if more information is needed (Campana et al., 2001), and thus they look toward their listeners to monitor comprehension (Nakano, Reinstein, Stocky, & Cassell, 2003). In turn, listeners rely on the speaker's gaze to disambiguate references, often before the reference could be clarified linguistically (Hanna & Brennan, 2007). In other words, gaze minimizes the time needed for interlocutors to coordinate their attention and confirm that the appropriate coordination has occurred. In a collaborative learning context, then, gaze can indicate both (a) the extent to which a speaker is providing appropriate non-verbal cues so that the listener can identify the referent, and (b) the extent to which the listener can use those cues to accurately identify the referent.

In education research, gaze behavior has been studied most often in individual learning scenarios (Sharma et al., 2017). When gaze behavior has been studied in collaborative learning contexts, the collaborations were typically remote, involving two participants looking at two computer screens in separate locations (Schneider, 2017). However, coordination of attention in shared spaces is a critical element of learning, and developing methods for analyzing gaze coordination is a high priority in the learning sciences (Schneider & Pea, 2017). Recent studies of gaze behavior in interactive contexts typically employ *mobile dual eye-tracking*, which allows researchers to develop nuanced and ecologically valid accounts of how interlocutors coordinate their gaze during natural, situated conversations (Clark & Gergle, 2011). Mobile dual eye-tracking methods can reveal how gaze is used as a referential resource—how people specify the person, object, or entity that they are talking about (Clark & Gergle, 2012).

With mobile dual eye-tracking data from participant dyads, *cross-recurrence analysis* (Zbilut, Giuliani, & Webber, 1998) is a commonly used analytic technique, as it permits the visualization and quantification of recurrent patterns of states between two time series—in this case, the gaze patterns of two conversational participants (see Figure 1; for recent examples of cross-recurrence analysis in the learning sciences, see Schneider et al., 2016; Hayashi, 2016). This approach can reveal the temporal dynamics of a mobile dual eye-tracking dataset without a priori assumptions about its statistical properties. The coordinate axes of a cross-recurrence plot indicate the gaze behavior of each of the two partners in interaction. Diagonal slices through the plot (from the lower left to the upper right) correspond to gaze alignment between the participants within some lag time.

Prior research on gaze coordination using cross-recurrence analysis has shown, for example, that a listener's eye movements most closely match a speaker's at a delay of 2 sec (Richardson & Dale, 2005; see Figure 1). Moreover, the more closely a listener's eye movements are aligned with a speaker's, the more likely the listener is to perform well on a test of comprehension administered after the conversation. Of course, gaze is not always well aligned. When speakers' referring expressions do not take into account listeners' needs, for instance, conversational dyads show poor coordination of visual attention (Arai, Bard, & Hill, 2009). Dyads whose members more effectively produce referring expressions coordinate their attention better and in a way linked to the elaboration of the referring expressions.

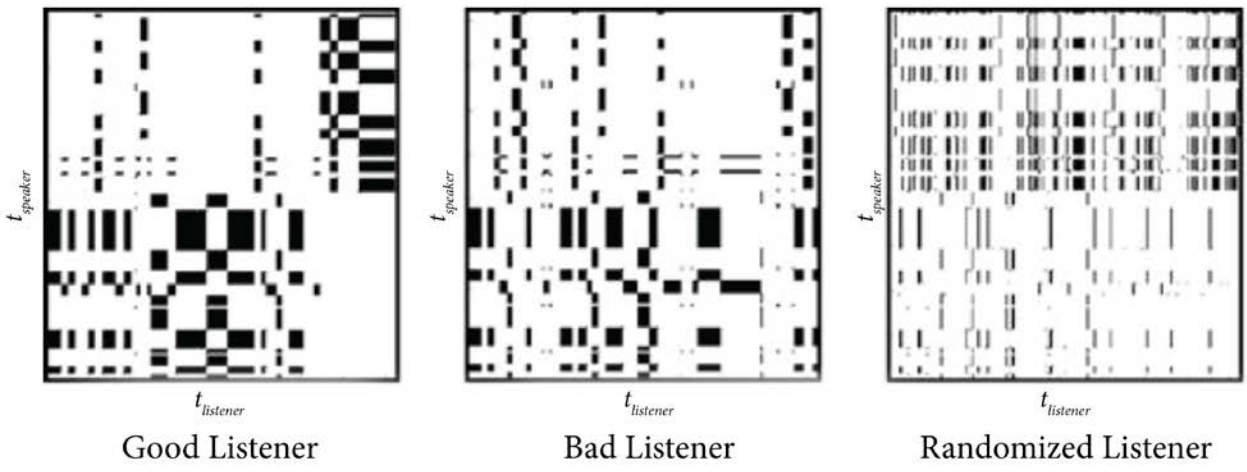

Figure 1: Cross-recurrence plots adapted from work by Richardson and Dale (2005). The vertical and horizontal axes indicate the gaze behavior of a speaker and a listener, respectively. Diagonal slices through the plot (lower-left to upper-right) correspond to an alignment of the participants' gaze within some lag time. A point is plotted on the diagonal whenever the gaze is recurrent. These plots show the difference between a "good" listener (good alignment with the speaker's gaze) and a "bad" listener (poor alignment with the speaker's gaze). The third plot shows the lack of alignment between someone with a random gaze and the speaker's gaze.

Although cross-recurrence analysis has been used successfully to analyze gaze coordination in collaborative learning contexts, it is best suited for processing data from short time windows, and it enables comparison of only one pair at a time. Cross-recurrence plots do not allow researchers to aggregate data from multiple dyads over long time spans in order to minimize the effects of individual differences and discover generalizable patterns of interaction. The plots are also very difficult to interpret visually, and they are not able to represent the complex, dynamic relationships between interlocutors that characterize coordinated gaze in a shared physical workspace. In the next section, we present epistemic network analysis as an alternative analytical tool that is able to measure and visualize dual mobile eye-tracking data such that researchers can explore gaze coordination both within individual dyads and across numerous dyads, supporting discovery of generalizable patterns in gaze coordination in some collaborative context.

Research Questions

The goals of this study were (a) to explore patterns and progressions in the gaze coordination exhibited by dyads engaged in a simple instructional task, and (b) to develop and test a method for constructing detailed and nuanced models of gaze coordination. To do this, we used ENA to address three challenges in modeling gaze coordination for which existing techniques, such as cross-recurrence analysis, are not well suited:

RQ1. Do the gaze behaviors of collaborating dyads during reference-action sequences reflect a particular progression of gaze coordination?
RQ2. How does gaze coordination change in the different phases of a reference-action sequence?

RQ3. Do gaze behavior patterns differ in reference-action sequences that include breakdowns or repairs?

Methods

To explore how gaze coordination develops over reference-action sequences in dyadic collaborations, we collected data in which pairs of participants engaged in a simple, collaborative task in which one person instructed another on how to prepare a sandwich. While this is not the kind of learning interaction one would find in a classroom, it provides a clear case of collaborative dyadic interactions with which to develop a method for modeling gaze coordination, and as such, we believe the results will generalize to a wide range of learning interactions involving reference-action sequences. Such sequences are common, for example, when teachers instruct students in classroom or laboratory settings, or when students work in teams to solve complex problems, such as how to create a design solution for an engineering problem.

Data Collection

Thirteen previously unacquainted dyads of participants were recruited from a large Midwestern university. Participants sat across from each other at a table containing 23 potential sandwich ingredients and two slices of bread (see Figure 2). One participant was assigned the role of *instructor*, and the other was assigned the role of *worker*. The instructor played the role of a customer at a deli and used verbal instructions to indicate to the worker which ingredients to put on the sandwich. The worker carried out these instructions, placing the desired ingredients on the bread.

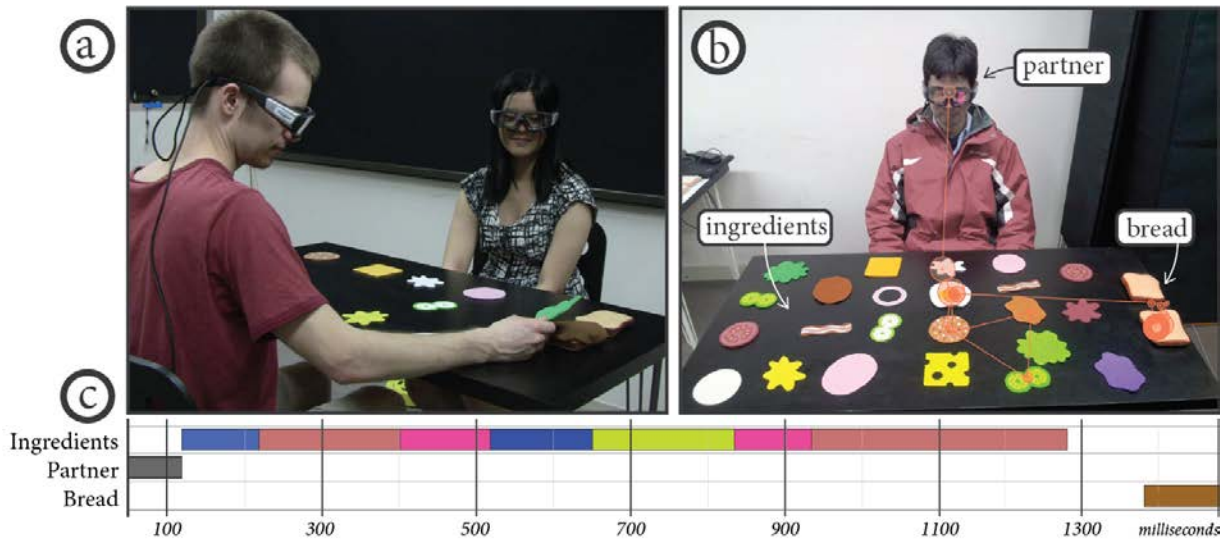

Figure 2: (a) The experimental set up. (b) A view from one participant's eye-tracking glasses. (c) A timeline of one participant's gaze fixation on the ingredients, the partner, and the bread during one reference-action sequence.

Each of the 13 dyads carried out the sandwich-making task twice, allowing the participants to switch roles. This resulted in 26 dyadic interactions. Instructors were told to request any 15 ingredients for their sandwich from among 23 ingredients laid out on the table. Instructors chose their own ingredients with no guidance from the research team. Instructors were told to request only a single ingredient at a time and not to point at or touch the ingredients directly. Upon completion of each sandwich-making task, a member of the research team placed all of the ingredients back in their original locations on the table, and the participants switched roles for the second sandwich-making task.

During each sandwich-making task, both participants wore mobile eye-tracking glasses developed by SMI (http://www.smivision.com/en/gaze-and-eye-tracking-systems). The glasses perform binocular dark-pupil tracking with a sampling rate of 30 Hz and gaze position accuracy of 0.5 degrees. A built-in high-definition camera (forward facing) was used to record audio and video (24 fps). The glasses worn by a dyad were time-synchronized so that the gaze data from both participants could be correlated.

We used BeGaze software (SMI) to automatically segment the data into *gaze fixations*—time periods when the eyes were at rest on a single target—and *saccades*—time periods when the eyes were engaged in rapid movement. This segmentation minimizes the complexity of eye-tracking data while preserving information about cognitive and visual processing behavior (Salvucci & Goldberg, 2000). BeGaze uses a dispersion-based (spatial) algorithm to compute fixations, emphasizing the spread distance of fixation points under the assumption that fixation points generally occur in proximity. Eye fixations and saccades are computed relative to a forward-facing camera located in the bridge of the eye-tracking glasses worn by the user. In other words, fixations and saccades are defined within the coordinate frame of the user's head, and user head movements do not interfere with the detection of eye movements.

Gaze fixations are characterized by their duration and coordinates in the forward-facing camera view. *Area-of-interest (AOI) analysis*, which maps fixations to labeled target areas (AOIs), is commonly used to annotate raw gaze fixation data (Salvucci & Goldberg, 2000). In this study, all fixations were manually labeled. The labeled AOIs serve as the input data for epistemic network analyses, rather than the raw gaze fixation data. Target AOIs included each of the sandwich ingredients, the slices of bread, and the conversational partner's face and body. Roughly 80% of gaze fixations were mapped to these AOIs (79.47% for instructors, 81.65% for workers), and the remaining gaze fixations were directed elsewhere (e.g., to a part of the table on which there were no sandwich ingredients). The speech of each participant was also transcribed. Instructor requests for specific objects were tagged with the referenced object, and worker speech was tagged when it was either confirming a request or asking for clarification.

To successfully reference something in the sandwich-making space, the speaker needs feedback from the addressee. Despite the best efforts of speakers to clearly communicate their thoughts, there are inevitably instances of breakdowns in communication—e.g., misunderstandings—that can hamper progress or lead to further breakdowns in subsequent exchanges. To correct

breakdowns when they occur, people engage in *repair*, a process that allows speakers to correct misunderstandings or clarify understanding of the relayed information (Hirst, McRoy, Heeman, Edmonds, & Horton, 1994; Zahn, 1984). In this study, if an instructor provided clarification for an initially inadequate reference—for example, in response to the worker's request for more information—that sequence was marked as containing a *repair*.

Each interaction between an instructor and a worker was divided into a set of reference-action sequences, such as a request for ham followed by the action of placing the ham on a slice of bread. Each reference-action sequence was then further divided into five discrete phases: (1) *pre-reference*, the time prior to an instructor making a reference, a verbal request for a particular sandwich ingredient; (2) *reference*, the time during which the instructor asks the worker to add a specific ingredient to the sandwich; (3) *post-reference*, the time immediately after the instructor's reference and up to the initiation of the worker's action in response to the reference; (4) *action*, the time during which the worker selects the requested ingredient, or the referent, and places it on the slice of bread; and (5) *post-action*, the time immediately following the completion of an action.

These phases are defined based on verbal indicators and physical actions, not gaze behaviors. The pre-reference phase (average length = 1.90 s) ends at the onset of the instructor's verbal reference. The reference phase (average length = 1.32 s) ends with the completion of the instructor's verbal reference. The post-reference phase (average length = 0.78 s) begins when the worker first touches the referent, initiating an action. The action phase (average length = 1.68 s) ends when the worker releases the ingredient after placing it on the bread, thus completing the action and initiating the post-action phase. The post-action phase (average length = 0.81 s) includes any feedback provided by the instructor and ends with the beginning of a preparatory utterance for the next reference, e.g., "so, um, next I want ...."

Epistemic Network Analysis

The dynamic and interdependent nature of dual mobile eye-tracking data makes analysis of the temporal patterns of gaze coordination particulary challenging. In this study, we adapted techniques from a learning analytic method, epistemic network analysis (ENA) (Shaffer, 2017; Shaffer et al., 2009; Shaffer, Collier, & Ruis, 2016; Shaffer & Ruis, 2017), to study interactive gaze behaviors across multiple dyads collaborating on the same task.

ENA was originally developed to model complex, collaborative thinking based on *discourse*, or the actions and interactions of people engaged in some cognitive task. While ENA has been widely used to study learning when students are engaged in complex problem-solving (see, e.g., Arastoopour, Shaffer, Swiecki, Ruis, & Chesler, 2016; Hatfield, 2015; Quardokus Fisher, Hirshfield, Siebert-Evenstone, Arastoopour, & Koretsky, 2016; Svarovsky, 2011), it is well suited to model patterns of association in any system characterized by complex, dynamic relationships among a relatively small, fixed set of elements.

The basic method with which ENA creates such models is described in detail elsewhere (Shaffer,

Collier, & Ruis, 2016; Shaffer & Ruis, 2017), but in brief, ENA creates adjacency matrices that quantify the co-occurrence of coded elements within some temporal context. The resulting adjacency matrices are normalized and embedded in a high-dimensional space. A dimensional reduction is performed using singular value decomposition (SVD), and the nodes of the networks—the coded elements—are placed in a metric space formed by the reduced dimensions using an optimization algorithm, such that the centroid of each network corresponds to the location of the network in the dimensional reduction. The result is two coordinated representations: (1) the location of each network in a projected metric space, in which all units included in the model are located, and (2) weighted network graphs for each network, which indicate *why* the network is positioned where it is.

ENA thus has several key affordances for modeling dual mobile eye-tracking data. ENA constructs network models that visualize and quantify the extent to which various dyads are looking at the same thing at the same time—or not. Because ENA produces network graphs that are coordinated with the underlying statistical properties of the model, ENA makes it possible to use the network graphs to interpret any statistically significant differences. For example, network centroids that appear farther to the right side of the ENA space will have more or stronger connections on the right side of the network graph, and centroids that appear farther to the left will have more or stronger connections on the left. Statistical tests can be used to determine if differences between different groups of centroids are significant on a given dimension, and if there are significant differences, the network graphs indicate which connection(s)—which co-occurring gaze behaviors—are most responsible for a given difference.

To model dual mobile eye-tracking data using ENA, data were annotated to indicate the *gaze targets* of each participant during the interaction, and these became the nodes in the network model. To do this, each interaction was segmented into reference-action sequences, and each sequence was further semented into 50 ms intervals. Each interval was then annotated, using a binary coding process, to designate which gaze targets were active during that sequence and which were not. Gaze coordination is defined as the co-occurrence of gaze targets within the same 50 ms interval. For any two gaze targets, then, the frequency of their association in the network of a given unit of analysis is computed based on the frequency of their co-occurrence in the data for that unit; frequency of association is represented by thicker, more saturated edges in the network graph. In this study, the units are defined by dyads, interactions, and phases: each dyad is represented as ten distinct units because each dyad had two interactions (switching roles after the first scenario), and each interaction involved reference-action sequences in five distinct phases.

Data in this study were analyzed and visualized using the ENA web tool version 3.0 (http://www.epistemicnetwork.org/).

Results

Before conducting epistemic network analyses, we computed descriptive statistics for the gaze

data. Not surprisingly, there was virtually no *mutual gaze*—direct eye contact between participants—during the reference-action sequences (0.92%), but there was a fair amount of simultaneous shared gaze, where both participants were looking at the same target (31.16%). Instructors made verbal reference utterances on average 1.31 s after first fixating on the referent, but they also made an average of 1.93 fixations on the referent before verbalizing their reference. Workers fixated on the referent an average of 1.65 s after the instructor made a verbal reference. Referential gaze in speech typically precedes the verbal reference by approximately 800–1000 ms, and listeners typically fixate on the referent about 2000 ms after the reference (Griffin & Bock, 2000; Meyer, Sleiderink, & Levelt, 1998). The data collected in this study largely reflect these general trends reported in the literature, and the slightly longer lag between gaze fixation and verbal reference among the instructors is likely due to the fact that instructors at times had to search for a sandwich ingredient rather than having one already in mind.

Analysis 1: General Gaze Behavior Patterns in Reference-Action Sequences

For our first analysis, all collected data were analyzed using ENA (see Figure 3). The unit of analysis was defined by dyad (n = 13), interaction (n = 2: each participant played the role of instructor in one interaction and worker in the other), and phase in the reference-action sequences (n = 5: pre-reference, reference, post-reference, action, or post-action). Thus each plotted point in the ENA space (center-top plot in Figure 3) represents the centroid of a network for one dyad, in one interaction, with data from one of the five phases collapsed across all reference-action sequences that occurred in the interaction. The colored squares represent the mean locations of the networks for all dyadic interactions in each of the five phases. The boxes surrounding the means are the 95% confidence intervals on the first (x) and second (y) dimensions. There is a clear separation between each of the five phases in the ENA metric space, indicating that the patterns of gaze behavior are significantly different in each of the five phases. The locations of the phases also correspond with their temporal sequence, as indicated by the arrows, suggesting a cyclical pattern (see Figure 3, center-top graph).

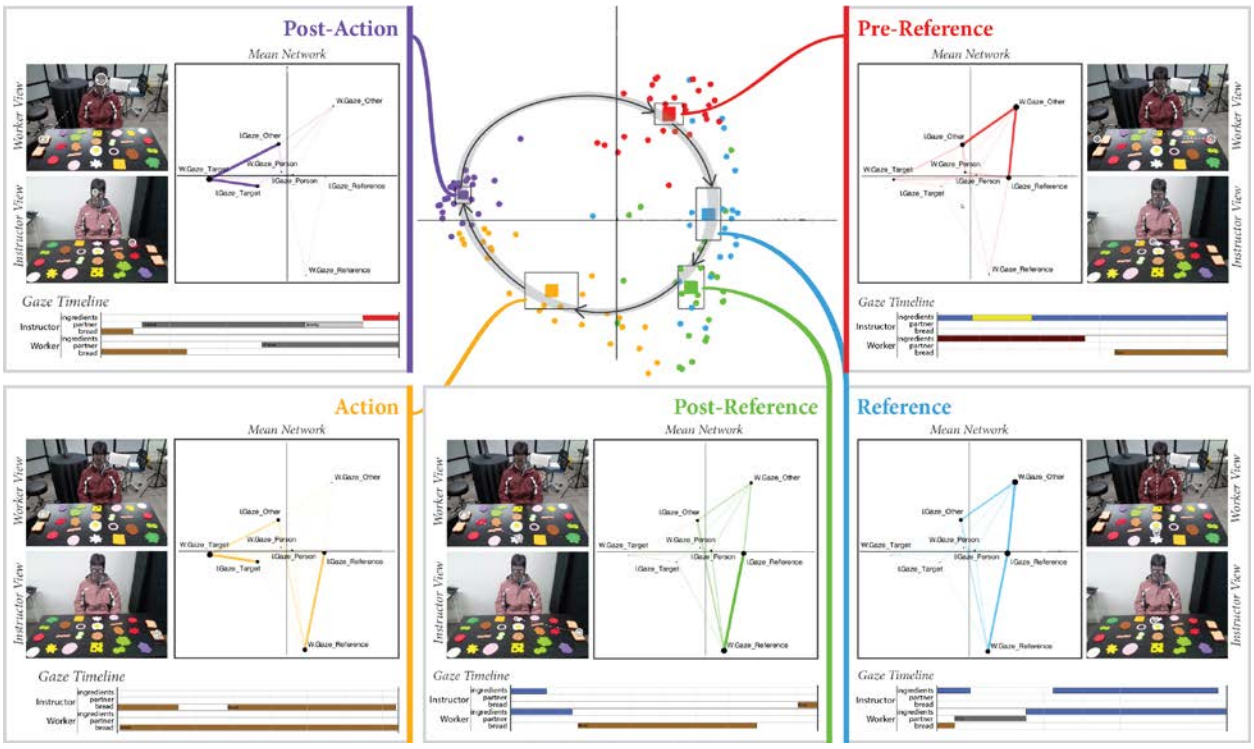

Figure 3: **Center Top:** The network centroids of each unit of analysis, color-coded by phase (points), with the corresponding means (colored squares) and 95% confidence intervals (boxes around the means). Note that the network centroids for each of the five phases appear in different parts of the ENA space. **Peripheral Graphs:** The remaining five graphs show the mean network for each of the five phases. Note that the strongest connections in the mean networks correspond with the locations of those means in the center-top graph, explaining the positions of the centroids in the metric space. An example gaze timeline and scan path is included with each of the mean networks to illustrate common gaze behaviors during that phase.

The mean networks for each of the five phases are also shown in Figure 3. The nodes in the networks are gaze targets, and the edge weights (thickness and saturation of the lines connecting the nodes) are determined by the relative amount of recurrent gaze on those targets. Each participant has four possible gaze targets: (1) the referent, (2) the other participant in the dyad, (3) the action target (i.e., the bread slices), and (4) something in the space other than the referent, other participant, or action target. Note that each gaze target is represented by two nodes, one for the instructor (I.*) and one for the worker (W.*). In this network model, connections can only occur between instructor and worker gaze target nodes, as one individual cannot simultaneously gaze at more than one target. Table 1 provides definitions for the annotations used in each of the three analyses.

ENA Network Node Names and Meanings

Analysis 1, 2, 3	*I.Gaze_Reference*	Instructor gazing at reference ingredient
	I.Gaze_Other	Instructor gazing at non-reference ingredient
	I.Gaze_Target	Instructor gazing at target bread
	I.Gaze_Person	Instructor gazing at the worker
Analysis 1, 3	*W.Gaze_Reference*	Worker gazing at reference ingredient
	W.Gaze_Other	Worker gazing at non-reference ingredient
	W.Gaze_Target	Worker gazing at target bread
	W.Gaze_Person	Worker gazing at the instructor
Analysis 2	*W.Same*	Worker gazing at same object as instructor
	W.Different	Worker gazing at different object than instructor

Table 1. Names and definitions of the gaze annotations used in each of the three analyses.

The node positions in the ENA space shown in Figure 3 help define the gaze behavior patterns that each dimension is capturing. Networks with centroids located high on the *y*-axis (second dimension), for example, are characterized by strong connections to *W.Gaze_Other*. That is, these networks indicate times when the worker is not looking at the referents, action target, or instructor. Networks with centroids located low on the *y*-axis, in contrast, are characterized by strong connections to *W.Gaze_Reference*, or times when the worker's gaze is oriented toward the referents (i.e., the sandwich ingredients). Similarly, moving from left to right along the *x*-axis seems to indicates a shift from the worker's gaze on the action target (i.e., the bread) toward the worker's gaze on the referents.

Figure 3 thus shows a different pattern of gaze behavior in each of the five phases of a reference-action sequence:

In the pre-reference phase, the strongest connections are between *W.Gaze_Other* and *I.Gaze_Other*, and between *W.Gaze_Other* and *I.Gaze_Reference*. As a result, the centroids appear in the upper right part of the ENA space. The pre-reference phase is thus characterized by the worker looking elsewhere while the instructor scans the sandwich ingredients, including the ingredient that they will verbally indicate as the referent in the next phase of the sequence.

In the reference phase, the strongest connection is, unsurprisingly, between W.*Gaze_Reference* and *I.Gaze_Reference*, and there is also a strong connection between *I.Gaze_Reference* and *W.Gaze_Other*, reflecting the lag between the instructor verbally indicating a referrent and the worker fixating on it. As a result, the network centroids have lower *y*-values that in the pre-reference phase.

In the post-reference phase, the strongest connection remains between W.*Gaze_Reference* and *I.Gaze_Reference*, but connections to *W.Gaze_Other* became much weaker, pulling the

network centroids even lower on the *y*-axis. This makes sense, as in this phase the worker is attempting to locate the sandwich ingredient that the instructor requested.

In the action phase, the strongest connection is between *W.Gaze_Target* and *I.Gaze_Target*, indicating that both participants were looking at bread. The network centroids, as a result, appear in the lower left part of the ENA space.

In the post-action phase, the strong connection between *W.Gaze_Target* and *I.Gaze_Target* remains, but a new strong connection appears between *W.Gaze_Target* and *I.Gaze_Other*, indicating that the instructor has started to scan other ingredients in anticipation of the next reference-action sequence while the worker is still looking at the sandwich. The network centroids thus have a low *x*-value, but they have a higher *y*-value than the centroids in the action phase.

This analysis suggests an overall progression of gaze behavior patterns in reference-action sequences during dyadic collaboration. Gaze coordination was distinctly different in each of the five phases in the reference-action sequence, and the cyclical progression through the ENA space reflects the sequence of phases. Importantly, although the phases themselves are defined temporally based on the speech and actions of the participants, the ENA model was constructed using only the annotated gaze data. That is, ENA was agnostic to the sequence of phases. Thus, gaze behavior patterns are uniquely different across different phases of a reference-action sequence. Moreover, these patterns progress in an orderly way through the abstract metric space constructed by ENA.

This analysis suggests that given a segment of annotated gaze data with no information about its phase, ENA could compute the network for the segment of data and predict the phase of the reference-action sequence from which it came. To validate this claim, we computed an ENA model as described above, but we omitted data from one randomly selected dyad, which resulted in an ENA space very similar to the one shown in Figure 3. Using the dyad data omitted from the model, we randomly selected excerpts of 200 ms and 1000 ms. We then computed ENA models for each of the excerpts and projected the networks into the space created for the model of the other 12 dyads. The predicted phase for each of the centroids representing excerpted data was determined based on the location of the centroid in the ENA space. Table 2 gives the results of this analysis as a confusion matrix. The rows indicate the actual phase from which each segment of data came, and the columns indicate the predicted phase. As Table 2 shows, the predictions are fairly accurate except for some confusion in the relatively short phases of *reference* and *action*. Prediction accuracy could easily be improved, however, by employing more sophisticated methods than we used here for demonstration purposes. For example, dynamically updating the phase predictions as segments of gaze data are collected or assigning confidence weights to the predictions based on their distance from phase centroids would increase predictive accuracy.

Predicting phase from segments of gaze data

Actual phase	Predicted phase (200 *ms* segments)					Predicted phase (1000 *ms* segments)				
	Pre-Reference	*Reference*	*Post-Reference*	*Action*	*Post-Action*	*Pre-Reference*	*Reference*	*Post-Reference*	*Action*	*Post-Action*
Pre-Reference	117	3	10	60	16	31	3	1	2	4
Reference	50	5	76	38	3	10	2	18	4	0
Post-Reference	6	0	31	6	0	0	2	7	0	0
Action	7	1	46	52	54	2	0	10	7	13
Post-Action	7	0	0	33	61	0	0	0	2	18

Table 2. Predicting the phase from which gaze data were excerpted based on the location of the network centroids in ENA space. The rows indicate the actual phase from which each segment of data came, and columns indicate the predicted phase. Cells with darker shading indicate higher proportions of excerpts from a given phase (row) predicted to be in the phase indicated by the column.

Analysis 2: Gaze Coordination in Lag-Adjusted Reference-Action Sequences

The goal of our second analysis was to determine which participant's gaze leads the other's, and by how much time, in each phase of the reference-action sequences. For this analysis, two additional annotations were used: *same* if the worker and instructor exhibited the same gaze behavior (person, referent, target, or other), and *different* if they did not. For each phase of the reference-action sequence, and for all dyads and interactions, we shifted the instructor's gaze from –2000 ms to 2000 ms in 50 ms increments, then computed the values for each of the new annotations (*same* and *different*). To find the optimal overlap, we divided the sum of the *same* values by the total number of increments to construct a measure of synchronization at each time lag.

The peak of the line for each phase shown in Figure 4 indicates the optimal time lag. These lags and the corresponding extent to which the participants' gazes were aligned, are given in Table 3. Positive lags indicate that the instructor was leading the gaze behavior of the worker, and negative lags indicate that the worker was leading them. In the pre-reference phase, neither participant led (t = 0 ms) and there is not much gaze coordination (22.5%). In the reference phase, the instructor starts to lead (t = 700 ms), and the gaze alignment increases (27.6%). In the post-reference phase, the worker begins leading (t = –300 ms), and the dyad reaches peak gaze coordination (36.1%). In the action phase, the worker leads by a small amount (t = –50 ms) and there is a slight drop in gaze coordination (34.6%). In the post-action phase, the instructor resumes leading (t = 300 ms), and gaze coordination declines further (27%).

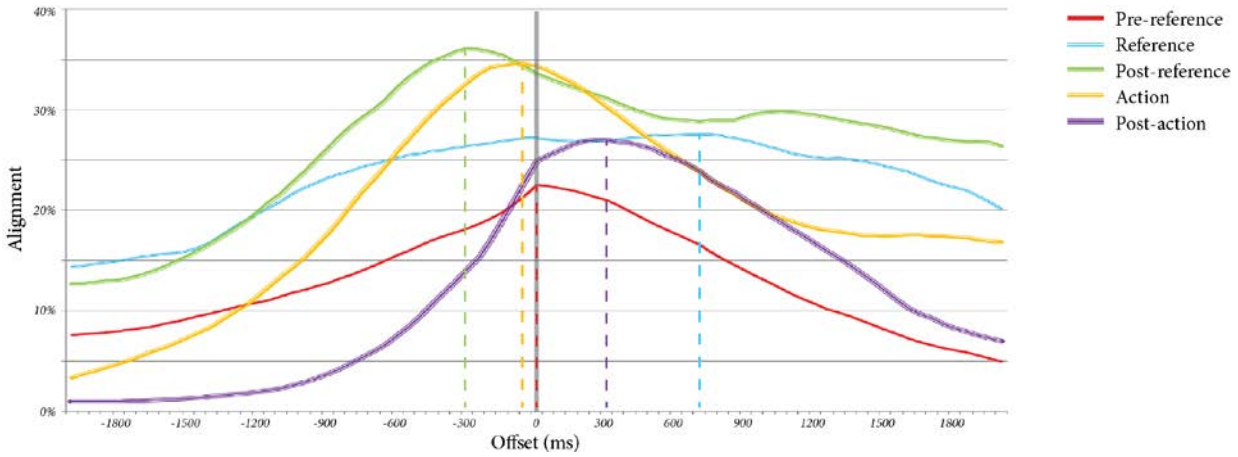

Figure 4: Percentage of gaze alignment between the instructor and worker in each of the five phases, plotted at offset lags from –2 s to 2 s in 50 ms intervals. Positive lags indicate that the instructor led the gaze alignment, while negative lags indicate that the worker led.

Based on the data shown in Table 3, we shifted the gaze data in each phase of the reference-action sequences by the optimal time lag for that phase. Then, we develop an ENA model of the adjusted data from the perspective of the instructors (see Figure 5). Four nodes represent the possible gaze targets for the instructor, as in Analysis 1, but there are only two nodes included for the worker: *W.same*, which indicates that the worker was looking at the same thing as the instructor, and *W.different*, which indicates that the worker was looking at something different than the instructor.

Optimal Lag & Alignment Percentage

	Pre-Reference	Reference	Post-Reference	Action	Post-Action
Optimal Lag (ms)	0	700	–300	–50	300
Alignment (%)	22.5	27.6	36.1	34.6	27.0

Table 3. Optimal time lags and the percentage of alignment at the optimal lag for each phase.

As Figure 5 shows, the mean location of each phase in the ENA metric space (center-top graph) is distinct, and the locations of the phases correspond with their temporal sequence, as indicated by the arrows (the pattern is similar to that found in Analysis 1). In this case, though, the networks indicate the extent to which the worker's gaze behavior is coordinated with the instructor's. Connections with *W.same* get stronger as the phases move from pre-reference to reference to post-reference, and then they get weaker through the action and post-action phases. The ENA model shows a similar pattern as the data on the level of gaze coordination for the optimal lag in each phase given in Table 3.

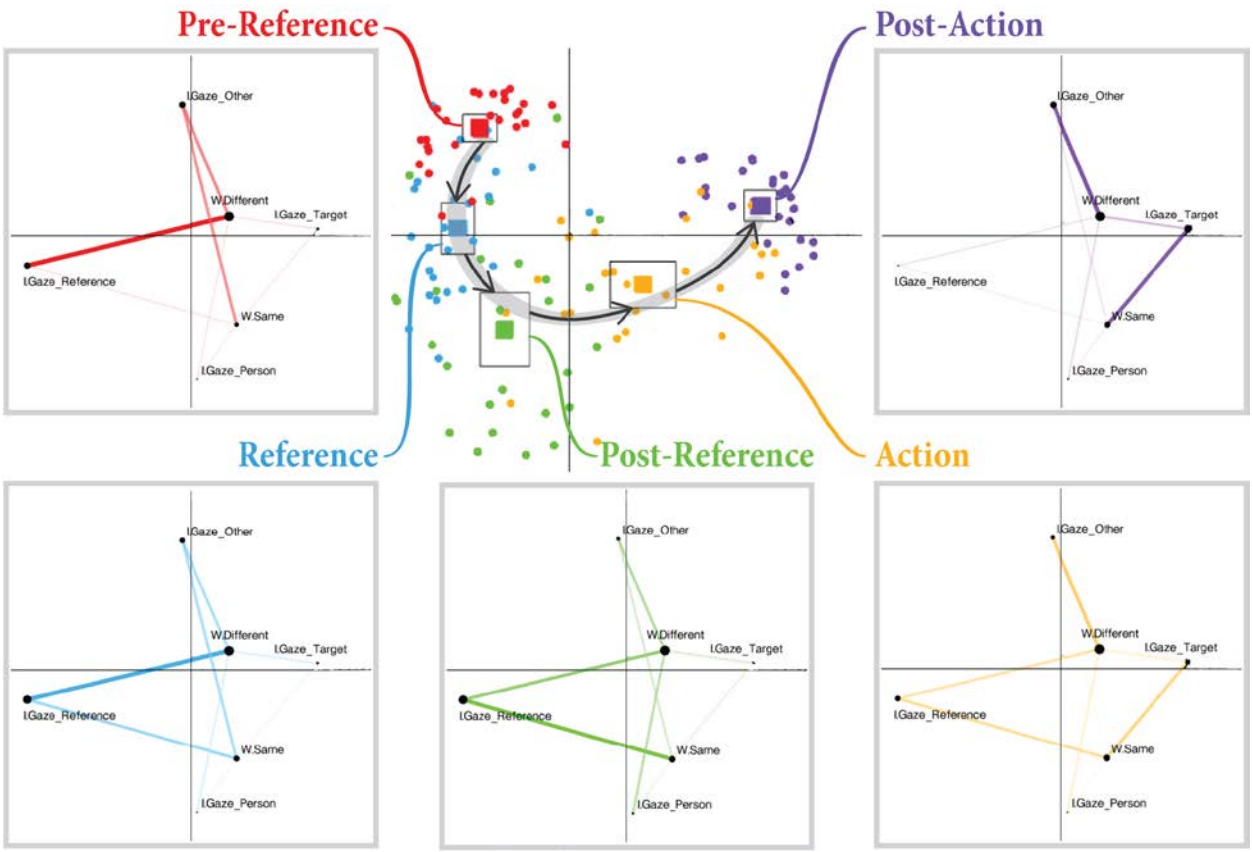

Figure 5: Network centroids and mean networks for the ENA model constructed based on lag-adjusted gaze data. The networks indicate the extent to which the worker's gaze behavior was the same as or different from the instructor's.

Analysis 3: Gaze Behavior Patterns in Reference-Action Sequences That Contain Repairs

In our third analysis, we explore whether patterns of gaze behavior in phases of reference-action sequences that included a repair were significantly different from phases that did not include such repairs. In particular, we examine whether the gaze behavior patterns from early phases (pre-reference, reference, and post-reference) are able to predict breakdowns later in the sequence (action and post-action).

For this analysis, we defined the unit of analysis by dyad, interaction, phase, and *repair*; that is, each reference-action sequence either contains a repair or not. We then analyzed the data using ENA as described above. As Figure 6 shows, the networks for each of the first three phases in the reference-action sequence (pre-reference, reference, and post-reference) are significantly different when the sequence contained a repair and when it did not. The differences are all on the *y*-axis. These phases all occur before or during any possible repair.

For the pre-reference phase, networks with repair are significantly higher on the *y*-axis than

networks without repair: $t = -2.17$, $p = 0.036$, Cohen's $d = 0.25$. This difference is due primarily to the fact that sequences with repairs exhibit a stronger connection between *I.Gaze_Reference* and *W.Gaze_Target*. This connection indicates a situation in which the worker is looking at the bread while the instructor is looking at the referenced ingredient. The worker may have remained engaged in the previous reference-action sequence, still looking towards the bread after having moved the previously requested ingredient there, while the instructor is already preparing to ask for the next ingredient in the current reference-action sequence. Frequent misalignment is thus associated with breakdowns in communication.

In contrast, the networks for the reference and post-reference phases that contain repairs appear lower on the *y*-axis than the networks of the corresponding phases without repairs: for the reference phase, $t = 2.12$, $p = 0.04$, Cohen's $d = 0.37$; for the post-reference phase, $t = 2.79$, $p < 0.01$, Cohen's $d = 0.45$. In both cases, the difference is largely caused by the stronger connections with *W.Gaze_Other* in the sequences with repairs. That is, repairs are more common when the worker's gaze is not fixated on the referent as frequently. In addition, the networks for reference and post-reference phases from sequences without repairs appear to contain stronger connections between *I.Gaze.Reference* and *W.Gaze.Reference*, suggesting that when both the instructor and worker are fixated on the referent, breakdowns in communication are less common.

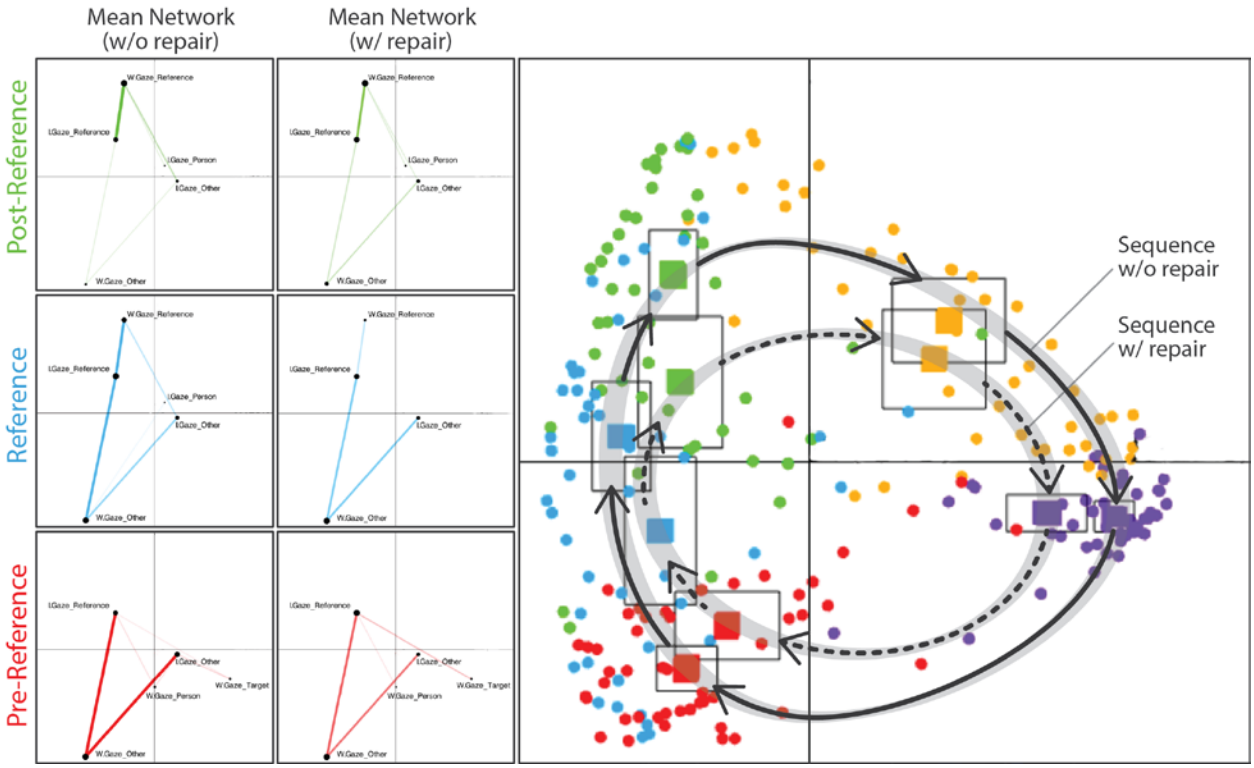

Figure 6: Network centroids and mean networks for the ENA model constructed based on whether the phase came from a reference-action sequence containing a repair or not.

This analysis indicates that the general pattern of gaze behavior identified in Analysis 1 is slightly different when information about repairs is included in the model. In particular, the gaze behavior patterns during the first three phases of a reference-action sequence are significantly different when a repair occurs during the sequence than when it does not. This suggests that the pattern of gaze behaviors early in a reference-action sequence may predict whether a breakdown in communication is likely to occur.

Discussion

Given the dynamic, interdependent, and complex nature of coordination, this study sought to develop a method for constructing detailed and nuanced models of coordinated referential gaze patterns in dyadic collaborations that take place in a shared physical space. To do this, we used ENA to address three challenges in modeling gaze coordination for which existing techniques, such as cross-recurrence analysis, are not well suited: (a) Do the gaze behaviors of collaborating dyads during reference-action sequences reflect a particular progression of gaze coordination? (b) How does gaze coordination change in the different phases of a reference-action sequence? and (c) Do gaze behavior patterns differ in reference-action sequences that include breakdowns or repairs?

While the specific scenario we used for the purposes of developing this novel analytic approach was fairly simple, it clearly demonstrates the affordances of ENA as a toolkit for analyzing collaboration. This study suggests that ENA can be used to analyze collaborative interactions of the sort that are common in learning contexts, whether it is very young children learning from their parents, schoolchildren learning from a simulated pegagogical agent, or adult learners working in teams to solve complex problems. Such interactions are typically mediated by various referents—toys, books, equations, models, schematics, and so on—and thus reference-action sequences are critical components of effective collaborative learning. Accounting both for what is said and for how the interlocutors reference the objects in their shared space (whether physical or virtual) provides a more complete model of the collaborative interaction.

Each of the three analyses conducted in this study reveal important properties of gaze behavior and patterns of gaze coordination in reference-action sequences. In the first analysis, ENA was able to characterize and distinguish the five phases of a reference-action sequence (pre-reference, reference, post-reference, action, and post-action) based on gaze coordination behaviors. The ENA model shows clear and significant differences in gaze coordination across the five phases, and the model reconstructs the ordered progression of gaze behavior despite having no information about the order of the phases. This analysis suggests that tracking the gaze behaviors of a collaborating dyad could be used to monitor and possibly even predict the dyad's progression through a reference-action sequence.

In the second analysis, we explored the extent to which the interacting participants exhibited gaze coordination throughout a reference-action sequence. By computing the lag between gaze behaviors, we identified a general pattern of increasing and decreasing coordination throughout a reference-action sequence, as well as consistent differences in which participant's gaze led the

other's. In particular, the instructor's gaze leads the worker's at the beginning and end of the reference-action sequence, when the instructor is driving the interaction by making a verbal reference or preparing to make a new one after the worker has responded to the previous one. In contrast, the worker's gaze lead's the instructor's during the middle of the reference-action sequence (i.e., in the post-reference and action phases), when the instructor is monitoring the worker's action in response to the reference (by attempting to locate the referent and move it to the target).

In the third analysis, we examined whether there were differences in gaze behavior patterns between reference-action sequences with repairs and those without. Our ENA models show similar overall patterns of gaze behavior for these two types of sequences, with one important difference: the gaze behavior patterns during the first three phases of a reference-action sequence are significantly different when a repair occurs during the sequence than when it does not. This analysis suggests that gaze coordination behaviors early in a reference-action sequence may predict whether a breakdown is likely to occur, though future research will have to establish whether there is a causal relationship.

These results, and the use of ENA to model gaze behavior patterns more generally, suggest a number of potential learning applications. For example, these kinds of models could be used to better control the gaze behavior of artficial pedagogical agents—such as social robots and virtual characters—to improve gaze coordination in collaborative learning interactions between humans and agents (Admoni & Scassellati, 2017; Andrist, Gleicher, & Mutlu, 2017; Karaman & Sezgin, 2018). In particular, such models will aid in the development of anticipatory rather than reactive control mechanisms for agent behavior, improving coordination in collaborative contexts (Hayashi, 2016; Huang & Mutlu, 2016; Olsen, Aleven, & Rummel, 2016). Of course, to do this will require a shift from developing *descriptive* models, such as those presented in this study, to developing *synthesizing* models that produce gaze behaviors for the purpose of improving gaze coordination, and ultimately learning. By detecting the gaze behavior of a human interlocutor and using that information to inform the the agent's gaze behavior, the agent could facilitate gaze coordination patterns that follow the cycle of human gaze coordination observed in Analysis 1.

The second and third analyses conducted in this study also suggest new directions for modeling and controlling the gaze behaviors of artificial pedagogical agents. Analysis 2 provides new insights on the role of gaze behavior in *mixed initiative* conversations (Novick, Hansen, & Ward, 1996). Specifically, the analysis indicates that an agent may be more effective if it leads the gaze coordination early in reference-action sequence (producing gaze behaviors to which the human interlocutor should respond) and then, later in the sequence, follows the human's gaze (responding to detected gaze behaviors of the human interlocutor). Similarly, Analysis 3 suggests that an agent could be programmed to recognize misunderstandings based on the human's gaze behavior before the human explicitly and verbally requests a clarification, which could produce smoother interactions. Ideally, agents would attempt to avoid gaze behavior patterns that are more likely to engender breakdowns.

This study adds to a growing body of work on gaze behavior and patterns of gaze coordination in collaborative contexts. While the scenario used is a simple one, our analyses suggest several avenues for future work on coordination in learning contexts. For example, research on the extent to which observed differences in gaze coordination affect the outcomes of collaborative learning interactions could inform the development of strategies to promote more effective collaboration. Such analyses will likely need to be multimodal, involving some combination of eye-tracking, video, and discourse data, as well as measures of student learning outcomes. Schneider and Pea (2015), for instance, found that remotely collaborating dyads who could see one another's gaze and who exhibited higher levels of verbal coherence (building on one another's ideas) had better learning outcomes. ENA could be used to create multimodal connectivity models that account for the relationships between gaze behaviors and discourse, providing insight on the ways in which collaborative coordination develops based on both embodied and discursive elements. ENA has been used in other learning contexts to create models that integrate physical and verbal factors (Ruis et al., in press), and such analyses could significantly advance work on coordination in collaborative learning.

Of course, this study has a number of limitations. Most notably, this research was conducted with a small number of participants in a laboratory setting. Because the present study was intended to provide proof of concept for a novel approach to analyzing gaze coordination, we chose a very simple collaborative scenario. Future work will need to replicate this study in collaborative learning contexts, such as one-on-one tutoring or cooperative problem solving. As such studies have largely been conducted on remote learning interactions (Schneider, 2017), it is particularly important to explore the connections between gaze coordination and learning in shared physical spaces (Schneider & Pea, 2017).

Future work should also explore further the temporal aspects of gaze coordination in reference-action sequences. This study divided reference-action sequences into an ordered cycle containing five distinct phases, but the gaze fixations within each phase were aggregated. This means that the ordering of low-level fixations was lost. Scanpath analysis is commonly used to analyze the temporal characteristics of gaze, but such scanpaths represent only the gaze behaviors of individuals. We were able to extract generalizable gaze patterns by aggregating data across multiple dyads and reducing the variability in gaze due to individual differences and changing context. However, ENA could be used in future studies to extend these findings by retaining information on the order of gaze fixations by creating directed network graphs. Such models would show not only when gaze is coordinated but who is leading whom, which would be particularly useful in studies of peer learning.

Conclusions

In this paper, we used ENA to better understand gaze coordination between individuals collaborating in a shared physical space. The context for these analyses was the reference-action sequence in a simple scenario designed to elicit a large number of such scenarios. Our analyses show (a) how gaze coordination progresses through a reference-action sequence, (b) the extent to which gaze coordination occurs in different phases of the sequence, and (c) which gaze

behavior patterns are associated with breakdowns and repairs. A similar analytic approach, we argue, could be applied to any collaborative interaction that involves reference-action sequences, which are commonly found in many learning contexts and require high levels of coordination between interlocutors. While these findings contribute to the growing body of work on gaze coordination during collaborative interactions, they provide a model for analyzing gaze behavior in a wide range of collaborative learning contexts, from classroom instruction and tutoring to cooperative problem-solving and agent-based virtual learning.

This study suggests that ENA can address analytic challenges in research on gaze coordination, providing a robust toolkit that could be used to develop models that account not only for gaze but also for gestures, facial expressions, discourse, and other data used to better understand collaborative learning. Such research could improve understanding of learning in collaborative interactions involving reference-action sequences, such as one-on-one tutoring or cooperative problem solving among students, and it could enable us to design learning technologies that more effectively promote and achieve coordination with human users.

References

Admoni, H., & Scassellati, B. (2017). Social Eye Gaze in Human-Robot Interaction: A Review. *Journal of Human-Robot Interaction, 6*(1), 25–63.

Altmann, G. T., & Kamide, Y. (2004). Now you see it, now you don't: Mediating the mapping between language and the visual world. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 347-386). Psychology Press.

Andrist, S., Gleicher, M., & Mutlu, B. (2017). Looking coordinated: Bidirectional gaze mechanisms for collaborative interaction with virtual characters. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 2571–2582). ACM.

Arai, M., Bard, E. G., & Hill, R. (2009). Referring and gaze alignment: Accessibility is alive and well in situated dialogue. In *Proceedings of the Cognitive Science Society* 31 (pp. 1246–1251).

Arastoopour, G., Shaffer, D. W., Swiecki, Z., Ruis, A. R., & Chesler, N. C. (2016). Teaching and assessing engineering design thinking with virtual internships and epistemic network analysis. *International Journal of Engineering Education, 32*(3B), 1492–1501.

Baldwin, D. A. (1995). Understanding the link between joint attention and language. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 131–158). Psychology Press.

Bavelas, J. B., Coates, L., & Johnson, T. (2002). Listener responses as a collaborative process: The role of gaze. *Journal of Communication, 52*(3), 566–580.

Brown-Schmidt, S., Campana, E., & Tanenhaus, M. K. (2005). Real-time reference resolution by naïve participants during a task-based unscripted conversation. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions* (pp. 153–171). MIT Press.

Campana, E., Baldridge, J., Dowding, J., Hockey, B. A., Remington, R. W., & Stone, L. S. (2001). Using eye movements to determine referents in a spoken dialogue system. In *Proceedings of the 2001 Workshop on Perceptive User Interfaces* (pp. 1–5). ACM.

Clark, A. T., & Gergle, D. (2011). Mobile dual eye-tracking methods: challenges and opportunities. In *Proceedings of the International Workshop on Dual Eye Tracking*.

Clark, A. T., & Gergle, D. (2012). Know what I'm talking about? Dual eyetracking in multimodal reference resolution. In *Proceedings of the CSCW 2012 Workshop on Dual Eye Tracking* (pp. 1–8).

Clark, H. H. (2003). Pointing and placing. In Sotaro Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 243–268). Lawrence Erlbaum.

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

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). American Psychological Association.

Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language, 50*(1), 62–81.

D'Mello, S., Olney, A., Williams, C., & Hays, P. (2012). Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of Human-Computer Studies, 70*(5), 377–398.

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences, 8*(1), 8–11.

Gergle, D., & Clark, A. T. (2011). See what I'm saying? Using dyadic mobile eye tracking to study collaborative reference. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work* (pp. 435–444). ACM.

Griffin, Z. M. (2004). The eyes are right when the mouth is wrong. *Psychological Science, 15*(12), 814–821.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science, 11*(4), 274–279.

Hanna, J. E., & Brennan, S. E. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language, 57*(4), 596–615.

Hatfield, D. L. (2015). The right kind of telling: An analysis of feedback and learning in a journalism epistemic game. *International Journal of Gaming and Computer-Mediated Simulations, 7*(2), 1–23.

Hayashi, Y. (2016). Coordinating knowledge integration with pedagogical agents. In *Intelligent Tutoring Systems* (pp. 254–259). Springer.

Hirst, G., McRoy, S., Heeman, P., Edmonds, P., & Horton, D. (1994). Repairing conversational misunderstandings and non-understandings. *Speech Communication, 15*(3-4), 213–229.

Huang, C. M., & Mutlu, B. (2016). Anticipatory robot control for efficient human-robot collaboration. In *11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 83–90). IEEE.

Hutt, S., Mills, C., White, S., Donnelly, P. J., & D'Mello, S. K. (2016). The eyes have it: Gaze-based detection of mind wandering during learning with an intelligent tutoring system. In *Proceedings of the 9th International Conference on Educational Data Mining* (pp. 86-93).

Jaques, N., Conati, C., Harley, J. M., & Azevedo, R. (2014). Predicting affect from gaze data during interaction with an intelligent tutoring system. In *International Conference on Intelligent Tutoring Systems*, (pp. 29–38). Springer.

Karaman, Ç. Ç., & Sezgin, T. M. (2018). Gaze-based predictive user interfaces: Visualizing user intentions in the presence of uncertainty. *International Journal of Human-Computer Studies, 111*, 78–91.

Meyer, A., van der Meulen, F., & Brooks, A. (2004). Eye movements during speech planning: Talking about present and remembered objects. *Visual Cognition, 11*(5), 553–576.

Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition, 66*(2), B25–B33.

Nakano, Y. I., Reinstein, G., Stocky, T., & Cassell, J. (2003). Towards a model of face-to-face grounding. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics* (Vol. 1, pp. 553-561). Association for Computational Linguistics.

Novick, D. G., Hansen, B., & Ward, K. (1996). Coordinating turn-taking with gaze. In *Proceedings of the Fourth International Conference on Spoken Language* (Vol. 3, pp. 1888–1891). IEEE.

Olsen, J. K., Aleven, V., & Rummel, N. (2016). Enhancing student modeling for collaborative intelligent tutoring systems. In *Intelligent Tutoring Systems* (pp. 485–487). Springer.

Quardokus Fisher, K., Hirshfield, L., Siebert-Evenstone, A. L., Arastoopour, G., & Koretsky, M. (2016). Network analysis of interactions between students and an instructor during design meetings. In *Proceedings of the American Society for Engineering Education* (Paper 17035). ASEE.

Richardson, D. C., & Dale, R. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science, 29*(6), 1045–1060.

Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The art of conversation is coordination: Common ground and the coupling of eye movements during dialogue. *Psychological Science, 18*(5), 407–413.

Richardson, D. C., Dale, R., & Tomlinson, J. M. (2009). Conversation, gaze coordination, and beliefs about visual context. *Cognitive Science, 33*(8), 1468–1482.

Ruis, A. R., Rosser, A. A., Quandt-Walle, C., Nathwani, J. N., Shaffer, D. W., & Pugh, C. M. (In press). The hands and head of a surgeon: Modeling operative competency with multimodal epistemic network analysis. *American Journal of Surgery*.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*(4) 696–735.

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research and Applications* (pp. 71–78). ACM.

Schneider, B. (2017). Dual eye-tracking in co-located spaces. In B. K. Smith, M. Borge, E. Mercier, & K. Y. Lim (Eds.), *Making a difference: Prioritizing equity and access in CSCL: 12th International Conference on Computer-Supported Collaborative Learning* (Vol. I, pp. 729–731). ISLS.

Schneider, B., & Pea, R. (2015). Does seeing one another's gaze affect group dialogue? A computational approach. *Journal of Learning Analytics, 2*(2), 107–133.

Schneider, B., & Pea, R. (2017). Real-time mutual gaze perception enhances collaborative learning and collaboration quality. In M. Orey & R. M. Branch (Eds.), Educational media and technology yearbook (pp. 99–125). Springer.

Schneider, B., Sharma, K., Cuendet, S., Zufferey, G., Dillenbourg, P., & Pea, R. (2016). Detecting collaborative dynamics using mobile eye-trackers. In C.-K. Looi, J. Polman, U. Cress, & P. Reimann (Eds.), *Transforming learning, empowering learners: The International Conference of the Learning Sciences (ICLS) 2016* (Vol. I, pp. 522–529). ISLS.

Schober, M. F. (1993). Spatial perspective-taking in conversation. *Cognition, 47*(1), 1–24.

Shaffer, D. W. (2017). Quantitative ethnography. Cathcart Press.

Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics, 3*(3), 9–45.

Shaffer, D. W., Hatfield, D. L., Svarovsky, G. N., Nash, P., Nulty, A., Bagley, E. A., … Frank, K. (2009). Epistemic network analysis: A prototype for 21st century assessment of learning. *International Journal of Learning and Media, 1*(1), 1–21.

Shaffer, D. W., & Ruis, A. R. (2017). Epistemic network analysis: A worked example of theory-based learning analytics. In C. Lang, G. Siemens, A. F. Wise, & D. Gasevic (Eds.), *Handbook of learning analytics* (pp. 175–187). Society for Learning Analytics Research.

Sharma, K., Jermann, P., Dillenbourg, P., Prieto, L. P., D'Angelo, S., Gergle, D., … Rummel, N. (2017). CSCL and eye-tracking: Experiences, opportunities and challenges. In B. K. Smith, M. Borge, E. Mercier, & K. Y. Lim (Eds.), *Making a difference: Prioritizing equity and access in CSCL: 12th International Conference on Computer-Supported Collaborative Learning* (Vol. I, pp. 727–728). ISLS.

Svarovsky, G. N. (2011). Exploring complex engineering learning over time with epistemic network analysis. *Journal of Pre-College Engineering Education Research, 1*(2), 19–30.

Szafir, D., & Mutlu, B. (2012). Pay attention! Designing adaptive agents that monitor and improve user engagement. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 11–20). New York, NY: ACM.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*(5217), 1632–1634.

Zahn, C. J. (1984). A reexamination of conversational repair. *Communications Monographs, 51*(1), 56–66.

Zbilut, J. P., Giuliani, A., & Webber, C. L. (1998). Detecting deterministic signals in exceptionally noisy environments using cross-recurrence quantification. *Physics Letters A, 246*(1), 122–128.