



Coding social interactions in naturalistic settings: The taxonomy of dyadic conversation

Kenya W. Mulwa^{1,2} · Sarah C. Kucker^{1,3}

Accepted: 20 November 2022 / Published online: 20 December 2022
© The Psychonomic Society, Inc. 2022

Abstract

Social interaction and conversation is an essential aspect of human behavior, yet existing methods for coding conversations are outdated, and often can only be used in contrived research settings. The Taxonomy of Dyadic Conversation (TDC) is a coding system designed to code dyadic interactions in natural settings by labeling the utterances and turns taken within an interaction using speech categories. The TDC was used to code child–caregiver and adult–adult conversations in a children’s museum and during a public forum, respectively. Results supported hypotheses that predicted adult–adult interactions would contain more Declarative Statement and Acknowledgment utterances than child–caregiver interactions, while child–caregiver interactions contained fewer Conversational Turns, as well as more Command and Encouragement utterances. Results also indicated high levels of inter-rater reliability. The potential for additions and modifications to be applied to the standard TDC is discussed.

Keywords Conversation · Dyadic interaction · Conversational turn · Naturalistic observation · Coding system

The dynamics of conversation is an extensively researched topic across a plethora of academic disciplines. Dyadic interactions are an essential, seemingly simple component of human social life that are critical for a myriad human behaviors, cognitions, and developmental processes. Yet existing research has shown just how complex, interdisciplinary, and multicausal interpersonal interactions tend to be. Capturing such interactions is difficult, but it is essential that we are able to accurately assess these components of human nature. Recent work on similar taxonomies has been sparse, and many of their uses are heavily context-dependent or require extensive off-line analysis of video/audio tapes. The goal of the current study is to refine and expand such observational, dyadic interaction coding to flexibly apply across multiple naturalistic settings and variables.

A host of classic studies have begun to explore the vast implications of the complexity of dyadic interactions for measuring and understanding human behavior. For instance, Freed (1994) identifies 16 types of questions used in informal conversation, and discusses how each type of question functions within an interaction. Similarly, conversations in formal settings can be intricately dictated by subtle factors such as social status or perceptions of power (Keating & Egbert, 2004). Other research describes conversations as a collaborative process, in which interlocutors must work together in order to understand each other and achieve successful conversation (Clark & Wilkes-Gibbs, 1986). Conversation is even one of the primary ways children learn vocabulary and build an understanding of the functions of new objects (Hirsh-Pasek et al., 2015; Kemler-Nelson et al., 2004; Zimmerman et al., 2009).

Across all dyadic interactions, the ability to measure both the quality and quantity of the interactions is vital. Quality is typically captured based on the content of the speech (Freed, 1994; Julien et al., 1989; Pasch et al., 2004); however, it can also be captured by nuanced verbal exchanges. Namely, the production of rich conversation has been described as a joint effort between two individuals that involves complex coordination of understanding and turn-taking (Shockley et al., 2009). Conversational turn-taking has also been associated with vocabulary growth and socio-cognitive development in

✉ Sarah C. Kucker
skucker@smu.edu

¹ Department of Psychology, Oklahoma State University, Stillwater, OK, USA

² Present Address: University of Wisconsin Oshkosh, Oshkosh, WI, USA

³ Present Address: Southern Methodist University, P.O. Box 750442, Dallas, TX 75275-0442, USA

children (Donnelly & Kidd, 2021). More specifically, didactic question-answer conversations have the potential to enhance overall conversation quality because of the multiple functions that questions are used for, including inquiring, suggesting, and even expressing politeness (Steensig & Drew, 2008). Moreover, quantity (i.e., the raw amount) of language input in conversations is associated with similar socio-communicative outcomes and quality (Hirsh-Pasek et al., 2015; Hoff & Naigles, 2002). In sum, the phenomena, nuances, and implications within dyadic conversations are veritable in their ubiquity as well as their complexity. As such, the ability to accurately capture these intricacies can be extremely advantageous.

Current approaches

Dyadic interactions are highly frequent in daily life and social science research, yet the methods for capturing them are less common. Historically, the most common approach for studying interactions has been through experimental observations using highly specific schemas in which observers rate the frequency of various behaviors or thematic verbalizations (e.g., Stiles, 1978; Robinson & Eyberg, 1981; Bales, 1950; Julien et al., 1989; Pasch et al., 2004). With the dawn of technology in research realms, electronic recording systems have become the go-to for their reliability. For instance, LENA[®] is a popular recording system that has been used in over 250 research publications and major conference presentations on language development since 2008 (LENA[®], 2022). However, such systems are not legally plausible in some situations and locations, nor are they always feasible for researchers due to cost or access. Thus, one aim of the current study is to offer an alternative option that utilizes a naturalistic approach.

There have been several variations of methods used to codify myriad components of social interactions, and many have succeeded in advancing the efficacy and accessibility of interaction taxonomies. For instance, Duran et al. (2019) also sought a method that returned to naturalistic observations, and created an open-source Python package in order to improve the systematic assessment of interactions concerned with linguistic alignment.

Similar advancements in methodological rigor have come to fruition in behavioral and affective observations of interactions such as Coan and Gottman's (2007) *Specific Affect Coding System* (SPAFF), through the utilization of latent psychological constructs. Moreover, access to cost-effective alternatives to expensive motion-tracking systems have been explored and have become more readily available (Romero et al., 2017). Indeed, methods such as these are excellent options for researchers interested in exploring variations of field/topic-specific coding systems.

However, the proposed taxonomy attempts to “cast a wider net,” allowing a diverse cohort of researchers across an array of interests, disciplines, and experience levels to be able to

explore the nuances of interpersonal interactions. Therefore, in order to provide a broadly accessible and generally applicable taxonomy while taking full advantage of the benefits unobtrusive observations afford researchers, here, we focus specifically on methodologies targeting the measurement of observable, verbal components in naturalistic dyadic interactions. For instance, Robinson and Eyberg's (1981) *Dyadic Parent-Child Interaction Coding System*, Stiles' (1978) *Verbal Response Modes*, and Bales' (1950) *Interaction Process Analysis*.

While foundational for research in the field, these methods and others similar in nature are limited by their specificity within and across speech categories as well as their modes of production. The *Dyadic Parent-Child Interaction Coding System* consists of more than 20 categories of verbalization, and its application is confined to clinical assessments and parent-child dyads alone (see Table 1) (Robinson & Eyberg, 1981). *Verbal Response Modes* (VRMs, Stiles, 1978) also have limited applicability, as their categories of verbalization are designed specifically to evaluate dimensions of interpersonal roles (Stiles, 1978) (see Table 2). Moreover, Stiles (1978) describes VRMs as “... analytic rather than empirical” and are “... based on a theory of the verbal communication of experience” (p. 694). As for Bales' (1950) *Interaction Process Analysis*, its goal is to capture the essence or commonalities between interactions instead of being focused on testing specific hypotheses from observations. Each of these methods are highly regarded for their specific application and population. However, the variability between existing methods with regard to what exactly they measure, how they measure it, and what implications can be drawn from the measurement limits their ability to be used across dyadic and environmental contexts.

Limitations of prior approaches

The most common limitation across the vast majority of coding systems similar to the Taxonomy of Dyadic Conversation (TDC), including recording devices, is their restriction to contrived observational settings that require participants to know observations are taking place. This ultimately subjects the results to the possibility of being skewed by reactivity effects (Jimenez-Buedo & Guala, 2016). Although contrived research settings and recording devices have a number of benefits and are common practice within psychology research, investigation into something as casual and innate as conversation often warrants the use of naturalistic observations that cannot be made with many extant methods. The benefits of conducting research in naturalistic settings have been organized into five “roles” by Miller (1977), four of which can be applied to the use of unobtrusive observations of dyadic interactions:

The roles include (1) studying nature for its own sake, (2) using nature as an initial starting point from which to develop a subsequent program of laboratory

Table 1 Summary of the *Dyadic Parent-Child Interaction Coding System* (Eyberg & Robinson, 2000)

Parent behaviors	Child deviance	Child responses to commands	Parent and child affect
*Direct Command	*Cry/Whine/Yell	Compliance	Parent Positive Affect
*Indirect Command	Smart talk	Noncompliance	*Child Positive Affect (Verbalization)
*Statement	*Destructive	No Opportunity	Child Positive Affect (Nonverbal)
*Descriptive Comment/Encouragement	Physical Negative-Child		Child Physical Warmth
*Descriptive Question/Encouragement	*Sibling Deviance		
*Questions			
*Acknowledgement			
*Irrelevant Verbalization			
*Unlabeled Praise			
*Labeled Praise			
*Problem Solving			
Physical Positive			
Physical Negative-Parent			
Physical Intrusion			
*Critical Statement			
*Negative Command			
*Grandma's Rule			
*Warning			
*Timeout Warning			
*Critical Statement-Marital			
Parent Ignore			

Asterisk* indicates verbalization category. Table was constructed from descriptions found in *Dyadic Parent-Child Interaction Coding System: A Manual*

Table 2 *Taxonomy of verbal response modes* (Stiles, 1978)

Focus	Frame of reference	Source of experience	
		Speaker	Other
Speaker	Speaker	Disclosure (D)	Question (Q)
	Other	Edification (E)	Acknowledgement (K)
Other	Speaker	Advisement (A)	Interpretation (I)
	Other	Confirmation (C)	Reflection (R)

research, (3) using nature to validate or add substance to previously obtained laboratory findings... (4) using the field as a naturalistic “laboratory” to test some hypothesis or theoretical concept. (p. 211)

Moreover, the majority of prior coding systems are exhaustive to such an extent that their applicability is restricted to specific contexts, which in turn makes it logistically difficult to capture interactions naturalistically. In essence, prior approaches to coding conversations lack a balance between capturing details of a conversation and the feasibility of the method's use.

In addition to the limited context and application of prior work, the tremendous advancements in technology since the

beginning of the twenty-first century have introduced extraneous variables that similar, older methods simply cannot account for. Most notably, the evolution of technology has drastically changed the way people, especially young adults in Western cultures, communicate in general. An abundance of research has shown that devices such as cell phones and personal computers are often used to avoid face-to-face social interactions, especially amongst people with social anxiety and other psychological disorders (King et al., 2013; Leung, 2007; Lu et al., 2011; Sapacz, Rockman, & Clark, 2016). Furthermore, cell phones disrupt caregiver–child interactions to such an extent that it limits both the frequency of conversations and a child's learning (Reed et al., 2017). In sum, it is critical that, when necessary, these coding metrics are able to capture elements of an interaction beyond verbal exchanges such as the use of technology, nonverbal gestures or behavior, and any other variable of interest relevant to a particular research question.

It should be noted that the prior coding methods discussed above were created and used to assess specific aspects of dyadic interactions in unique contextual circumstances. The limitations discussed regarding these coding procedures are not limitations of the taxonomies themselves, but rather the limitations of their applicability across unspecified research settings and dyadic contexts.

Present coding system

The purpose of the TDC is to provide the research community with a foundational coding system that can capture both quality and quantity across a variety of settings for speech in interpersonal interactions, particularly in naturalistic settings. The advantages of naturalistic research settings and observations should be formally considered when analyzing such innate experiences as social interactions. Naturalistic observations are both convenient and generally effective for researchers and participants alike. Here, the advantages of observing and coding interactions unobtrusively are threefold: (1) it reduces the effects of subject reactivity, (2) requires minimal technical or financial resources, and (3) provides an opportunity to include a diverse sample of participants across a wide range of contexts because it is adaptable to multiple settings.

Evidence of reactivity effects in psychology research has been discussed extensively in the existing literature (Bandura, 1991; Bandura et al., 1977; Carver & Scheier, 1981; Scheier & Carver, 1988). Additionally, observing conversations in a public setting does not require technological equipment such as audio or video recording devices, which is especially advantageous for research programs with limited funding, and is ultimately less invasive for participants. Moreover, this approach is thus more accessible for researchers and allows for a wider context in which data can be collected. Indeed, because unobtrusive observations in public settings are often not considered human subjects research, they are often determined to be exempt from institutional review board (IRB) oversight, which can further broaden the use of such methods. The benign nature of data collection fosters increased anonymity of participants, and allows for reliable data collection in all-party consent states or countries where consent for other recording methods is impossible (e.g., places that require consent from every single individual in the public space prior to recording, which is often impossible in busy places such as museums, malls, or parks). Perhaps more importantly, these methods introduce an opportunity to collect data from samples that are more representative of populations, including minoritized groups who are likely to be skeptical of participating in research for a variety of reasons (Corbie-Smith et al., 2002; Farmer et al., 2007; Scharff et al., 2010). Still, research using the TDC in private environments such as academic and professional settings or organized community events would require varying levels of institutional oversight, and more intimate uses would typically require comprehensive institutional review. Nonetheless, the feasibility of conducting unobtrusive observations in public settings creates an opportunity for the TDC to be used to collect data from members of marginalized communities, of which, processes within the conventional research endeavor routinely ignore (Roberts et al., 2020).

In order to establish a firm foundation for this type of coding system, limiting the specificity of criteria for speech categories is vital. Flexible, less specific speech categories create an opportunity for the TDC to be used across a variety of academic disciplines, environmental contexts, and dyadic dynamics. Further, an updated method like this is required so as to achieve an accurate evaluation of the complexities that technology has brought to social interactions in our contemporary world. Finally, the formulation of a systematic, empirical, valid, and reliable coding system for coding naturalistic conversations should be prioritized. This is our goal here.

Method

Overall coding system

The main components of the standard TDC method capture both the quality of interactions through classifying the type of talk and conversational turns as well as the quantity of utterances and turns. The TDC uses ten speech categories commonly found in dyadic communication coding systems (Robinson & Eyberg, 1981; Stiles, 1978) in order to assess the overall quality as well as the frequency of utterances within a dyadic interaction: (1) “Wh-” Questions (i.e., who, what, when, where, why), (2) Yes/No Questions (i.e., polar questions), (3) Declarative Statements, (4) Commands or directive statements, (5) Acknowledgments, (6) Announcements, (7) Evaluative–encouragement, (8) Evaluative–criticism, (9) Irrelevant utterances, and (10) Unknown, which is primarily used when utterances are unintelligible. Conversational turns are used to code responses from one participant to another, and are indicated using an arrow on the coding sheet toward the responder. If any of the aforementioned speech categories are followed by a response from another subject within 2 seconds, it is considered a conversational turn. A back-and-forth conversation will thus have a series of conversational turns, whereas a string of comments without a response will have no or minimal turns. Interruptions are not considered conversational turns. See Table 3 for a description of speech categories and coding symbols. Figure 1 shows a demonstration of the TDC being used in an example interaction. Detailed coding manual and examples are provided via the Supplementary Information and Open Science Framework (OSF) (https://osf.io/82xu5/?view_only=c7f78024499d4f62a19287d21135d9f6).

At the end of the interaction, the valence of the interaction is also recorded, and refers to the overall affect (i.e., positive, negative, or neutral) of the interaction. Positive affect is signaled through smiling, laughing, using a positive tone of voice, employing terms of endearment, demonstrating physical affect, and exhibiting synchronous, responsive interaction.

Table 3 Taxonomy of dyadic conversation

Speech categories	Coding symbol	Criteria
"Wh-" Questions	Qw	An open-ended question that typically begins (or is implied to begin) with "Who," "What," "Where," "When," "Why," or "How"
Yes/No Questions	Qy	A closed-ended question in which the answer is likely "Yes" or "No"
Declarative Statements	S	Gives an account of the objects or people in the situation or activity
Commands (Directive Statements)	C	Order, demand, directive of something a subject should do or not do. Intended to initiate or stop a child's behavior
Irrelevant	I	Does not pertain to relevant ongoing activity, objects, or individuals
Acknowledgement	A	Brief verbal response to another's verbalization or behavior that contains no descriptive content
Announcements	AA	Brief verbalizations made independently from other conversations or behavior.
Evaluative (Encouragement)	Ee	Statement or phrase expressing approval, appreciating, positive acknowledgment, or praising of effort, attributes, or product
Evaluative (Criticism)	Ec	Statement of phrase expressing disapproval or criticism
Unknown	U	When a talk item is produced by toward a target subject, but is not easily classified in one of the above ways
Conversational turn	←/→	Direction of the arrow indicates a subject's response to an utterance

Negative affect is signaled through frowning at the other person, using a harsh tone of voice, exhibiting closed-off body language, disengaging from the interaction, scolding, or showing angry, frustrated, or anxious behavior. Neutral interactions are neither positive nor negative; they are verbal interactions without significant emotion. Valence is labeled as follows: 1 = extremely positive affect, 2 = positive affect, 3 = neutral, 4 = negative affect, 5 = extremely negative affect.

Training

The TDC can be used in many different contexts within naturalistic settings and across a variety of dyad dynamics. Caregiver–child dyadic interactions can be recorded in public places such as parks, zoos, and museums. Adult–adult interactions can be recorded in formal business, academic, and other professional settings and contexts such as public forums or meetings. Other general dyads can be recorded in virtually any public domain such as restaurants, coffee shops, libraries, public chat rooms or virtual environments, and more. In order to maintain the naturalistic essence of an interaction, it is important that raters remain unobtrusive to subjects, but within earshot of the conversation. Having multiple raters is also preferred to increase reliability of the code. Before beginning the coding process, raters should designate target subjects as subject A and subject B. Then, the raters should record each utterance and conversational turn in conjunction with the speech category and turn criteria.

Before raters code live interactions, it is important that they demonstrate a comprehensive understanding of the coding scheme. After reading the coding manual, raters practiced coding with a set of preselected video interactions

compiled from YouTube (see Supplementary Information or OSF). Before coding live interactions, each coder coded a minimum of eight interactions from the videos, in real time (i.e., without pausing, rewinding, or replaying the video), with at least 90% accuracy. Accuracy was based on a standard key developed by the authors (see Supplementary Materials or OSF). Evidence suggests that coders were able to pick up the coding system with ease and at a high level of reliability without extensive retraining. Indeed, across five new coders who completed at least eight training videos, average agreement to the standard was considered excellent (see Table 4). These coders also represented diverse backgrounds—Coder 1 identified as a non-White female, Coder 2 as a Black male, and the other three as white females. All coders were young adults enrolled in college courses. As noted further below, coders went on to code live sessions at an equally high reliability rate.

Potential research-specific code modifications

Finally, and importantly, there are optional additional notes or codes for study-specific elements, such as location or contextual features, number of individuals in a group, duration or length of interactions, use of technology, and others. For instance, if a researcher is interested in assessing the frequency of cell phone distraction in an interaction, the Irrelevant speech category could be omitted and a "Distraction" category could be added. Alternatively, if a study aims to capture emotional reactions, Acknowledgements can be modified to include laughs, smiles, child babbles, or other nonverbal responses.

Example Interaction

Person A: Hey, Tonya!
Person B: Hey, did you get the email I sent you?
Person A: No, I haven't been at my computer all day. What was it about?
Person B: Nothing urgent, I just wanted to see if you've had a chance to work on the presentation yet. I haven't had a chance to look at it with all the work I've had to do this week. I'm really sorry I haven't gotten the chance to—
Person A: —That's actually what I came to talk to you about. I have some ideas I want to run by you.
Person B: Great!
Person B: Have a seat.
Person A: Thanks!
Person A: Before we get to the Q&A portion, I was thinking we could present the answers to questions we expect to be asked.
Person B: That's a great idea!

Annotated Code

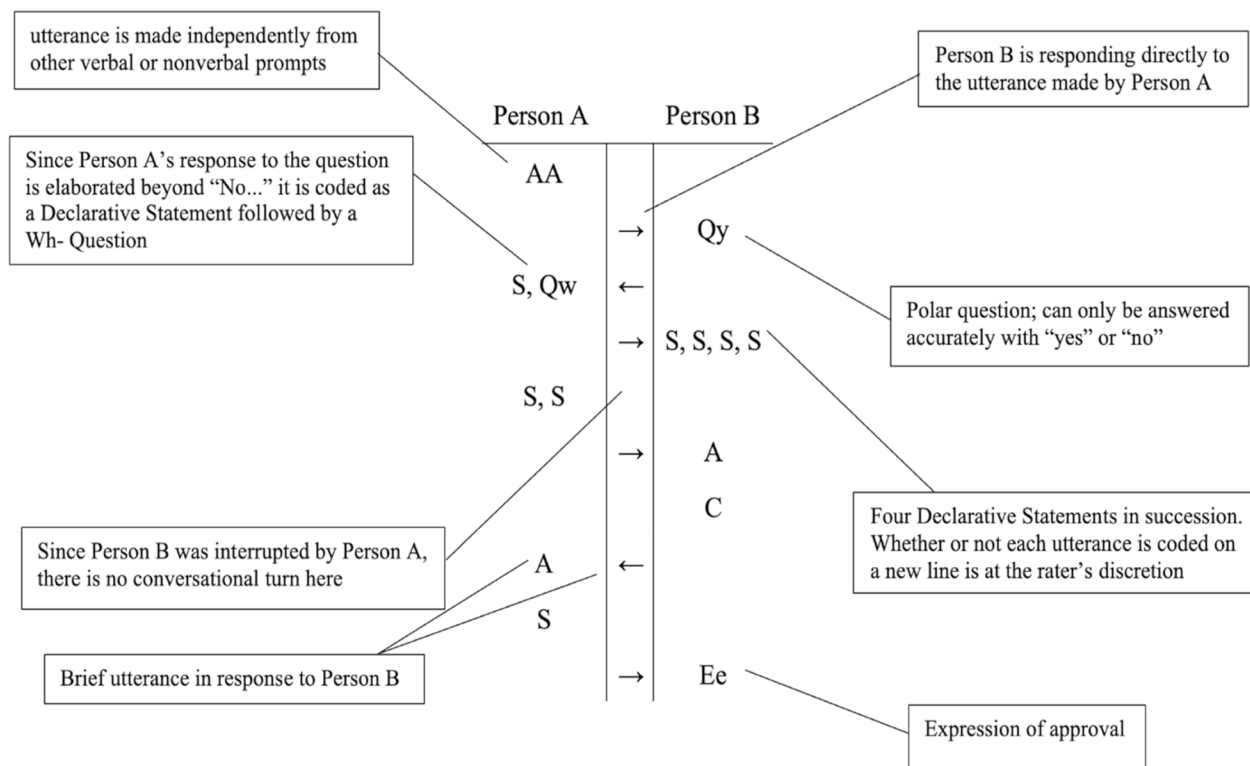


Fig. 1 Coding example

An applied example is well illustrated by the “Announcements” speech category. This particular category was not included in the original coding manual for a study involving

interactions between caregivers and their children in a museum, which will be discussed in detail below. We noticed that many conversations included utterances that

Table 4 Average agreement between coders on training videos

Coding item	Coder 1	Coder 2	Coder 3	Coder 4	Coder 5	Overall
Speech categories	.86 (.07)	.79 (.18)	.82 (.11)	.90 (.10)	.83 (.18)	.84 (.14)
Conversational turns	.89 (.21)	.60 (.41)	.84 (.24)	.80 (.45)	.67 (.45)	.76 (.37)
Valence	+	.90 (.32)	.88 (.34)	.61 (.21)	1.00 (0)	.80 (.31)
Demographics	+	.91 (.10)	.96 (.06)	.97 (.06)	1.00 (0)	.96 (.07)

Average percent agreement shown for each coding item across coders with SD in parentheses. Overall agreement shown in bold

+Coder 1 did not have valence or demographic data for the training videos

were used to get the attention of a child or their caregiver, and initiate a dialogue without being prompted by another utterance (e.g., “Hey” or “Wow”). These types of utterances were initially coded as Acknowledgements; however, a “Hey!” used to get someone’s attention seems to serve a different purpose in an exploration-based STEM dialogue from what is described as an Acknowledgement. In fact, this type of attention-orienting utterance did not fall under any of the speech categories we initially identified. Therefore, we solved this problem by simply adding a new speech category that captures this type of utterance. Additionally, modified versions can be juxtaposed alongside the presented standard TDC to assess the efficacy of modified components, such as we did with the “Announcements” category. Below are illustrative cases that demonstrate both the feasibility and reliability of the TDC as well as the ease by which two such examples—the inclusion of contextual information and speech category modification—can be seamlessly integrated into the coding.

Illustrative cases

Hypotheses

The conversations adults have with other adults differ broadly from conversations adults have with children. With a more sophisticated understanding and extended experience with social interactions, utterances from an individual in an adult-adult conversation tend to be complex, elaborate, and narrative (Ochs, 2004). Moreover, adults often use utterances referred to by Schegloff (1982) as “continuers” (e.g., “uh-huh” or “mhm”) in order to communicate a variety of signals such as attention or understanding, and they can also serve as a prompt for conversational turns. Therefore, we predicted that conversations within a formal adult–adult context would include significantly more utterances within the Declarative Statements and Acknowledgments speech categories, as well as more conversational turns than the caregiver–child dyads.

Research also suggests that conversational turn-taking is a skill that children continue to develop as they age (Maroni, Gnisci, & Pontecorvo, 2008). In addition, caregivers use high rates of directives (i.e., commands; Girolametto et al., 2000) and increase the use of encouragements with children in learning settings (Willard et al., 2019). This led us to predict that, compared with adult–adult dyads, caregiver–child dyads would contain fewer conversational turns, more Commands, and more Encouraging utterances than interactions between adults in the formal setting. We also predicted that adult–adult interactions would tend to be neutrally valenced given their more formal, professional setting, while the valence of caregiver–child interactions would be generally positive due to the context of an interactive museum and play exploration.

Illustrative Case 1: Caregiver–child dyad in public setting

Overview and adaptation of coding system In the first illustrative case, observations were conducted at a local children’s museum by Coders 1–3 (Table 4) and two other coders who both identified as white females. In order to reduce reactivity effects, raters positioned themselves near exhibits in such a way that they did not disrupt or engage with any of the museum guests, but were still able to clearly hear conversations between caregivers and their children. Here, Person A was the target child being observed and Person B was their collaborative partner (typically perceived as the caregiver). The exhibits varied in the presence of signage (no signage, some signs, or many signs). In addition to the standard TDC coding, a portion of the sample was tested under the modification adding Announcements (AA). Moreover, additional elements were added including the start and stop times of conversations, and notes about the presence of signage were recorded by the rater.

Participants A total of 324 caregiver–child dyads were observed as they interacted with exhibits within the Tulsa Children’s Museum and Discovery Lab. Of these, 240 were observed using the standard TDC coding and 104 were observed under the addition of Announcement. All

Table 5 Perceived participant demographics based on self-presentation

	Person A (child, host)				Person B (caregiver, colleague)			
	Average	SD	Min	Max	Average	SD	Min	Max
Illustrative Case 1: Caregiver-child dyad in a public setting								
Age	6.06	2.09	1	12	34.01	9.40	5	65
Gender	Female <i>n</i> = 167 Male <i>n</i> = 139 Ambiguous <i>n</i> = 3				Female <i>n</i> = 210 Male <i>n</i> = 97 Ambiguous <i>n</i> = 2			
Race/ Ethnicity	White <i>n</i> = 253 Black <i>n</i> = 18 Asian <i>n</i> = 3 Native <i>n</i> = 1 Multiracial <i>n</i> = 1 Hispanic <i>n</i> = 31 Ambiguous <i>n</i> = 2				White <i>n</i> = 257 Black <i>n</i> = 15 Asian <i>n</i> = 1 Native <i>n</i> = 1 Multiracial <i>n</i> = 0 Hispanic <i>n</i> = 32 Ambiguous <i>n</i> = 2			
Illustrative Case 2: Adult-adult dyad in formal public discussion								
Age	54.5	8.23	40	70	46.5	11.60	30	65
Gender	Female <i>n</i> = 4 Male <i>n</i> = 16				Female <i>n</i> = 18 Male <i>n</i> = 2			
Race/ Ethnicity	White <i>n</i> = 19 Black <i>n</i> = 0 Asian <i>n</i> = 1 Native <i>n</i> = 0 Multiracial <i>n</i> = 0 Hispanic <i>n</i> = 0 Ambiguous <i>n</i> = 0				White <i>n</i> = 20 Black <i>n</i> = 0 Asian <i>n</i> = 0 Native <i>n</i> = 0 Multiracial <i>n</i> = 0 Hispanic <i>n</i> = 0 Ambiguous <i>n</i> = 0			

Demographic information was based on observation by the coders and not self-report. Thus, the demographics reported here are perceived based on the participants' self-presentation and do not include any ambiguous or non-observable identities

individuals had timing and exhibit data noted. Demographic information as noted by the coders is listed in Table 5.

Analysis The total counts for each speech category observed in each dyad were recorded, as was the number of conversation turns, the total amount of time the conversation lasted (dwell time), and the overall valence. To compare sets of observations (comparing illustrative cases, modifications, subgroups, signage), generalized linear mixed models were run predicting total amount of talk, turns, and valence per dyad. Initial models included fixed factors of speech category and group (illustrative case setting, modification pre vs. post, subgroups, or signage), their interaction, and a random intercept of coder. All fixed factors were sum-coded. The reference group in speech category was Declarative Statements, and references for the other groups were, respectively, Illustrative Case 1, Pre-modification, Subgroup A, and No signage. When the hypothesis included differences in speaker, it was also included as a fixed factor (sum-coded with Person A as the reference) and allowed to interact with the others. In some cases, speech category was highly collinear (VIFs > 10; using the performance package and check_collinearity function of R), and in these cases, individual linear regressions were run for each speech category independently. All regression models were fitted

using RStudio 1.4.1717 (R Core Team, 2022) with the lme4 (Bates et al., 2015) and lmerTest (Kuznetsova et al., 2017) packages. We used the maximum likelihood approximation for lmer and Satterthwaite approximation for degrees of freedom in lmerTest for the models. Post hoc tests were run using lsmeans (Lenth, 2016) and Tukey adjustment. See OSF page for R scripts: https://osf.io/82xu5/?view_only=c7f78024499d4f62a19287d21135d9f6.

Reliability for the coding scheme was also assessed for both illustrative cases. Inter-rater reliability was assessed using a mean, two-way, random absolute effects intraclass correlation (ICC) model and percent agreement. As per guidelines by Cichetti (Cicchetti, 1994; Cicchetti et al., 2006; see also Fleiss, 1986), an ICC above .90 and percent agreement at 90% or higher was considered excellent.

Data and results

Overall findings Descriptive data on average amount of each speech category, dwell time, and valence are reported in Table 6. Importantly, nearly all categories of talk were identified in dyadic conversations to some extent, with Declarative Statements being the most common utterance

Table 6 Descriptives of Illustrative Case 1 vs. Illustrative Case 2

	Illustrative Case 1 (child–caregiver; public setting)			Illustrative Case 2 (adult–adult; formal discussion)		
	Person A	Person B	Total	Person A	Person B	Total
Speech category						
Qwh	.22 (.48)	.49 (.82)	.71 (.99)	.95 (1.05)	.05 (.22)	1.0 (1.17)
Qyn	.53 (.70)	1.01 (1.04)	1.54 (1.26)	.70 (1.05)	.15 (.37)	.85 (1.04)
S	2.67 (2.43)	2.80 (2.70)	5.48 (4.38)	3.90 (2.51)	6.45 (4.89)	10.35 (5.13)
C	.31 (.77)	1.68 (1.74)	1.99 (1.87)	.35 (.75)	0 (0)	.35 (.75)
I	.11 (.37)	.15 (.90)	.27 (.99)	0 (0)	0 (0)	0 (0)
A	.36 (.70)	.60 (.78)	.96 (1.09)	.40 (.59)	1.40 (1.05)	1.80 (1.11)
AA*	.29 (.73)	.16 (.50)	.85 (1.09)	0 (0)	0 (0)	0 (0)
Ee	.02 (.15)	.53 (.97)	.55 (.99)	0 (0)	0 (0)	0 (0)
Ec	.06 (.33)	.02 (.15)	.08 (.36)	0 (0)	0 (0)	0 (0)
U	.07 (.32)	.01 (.08)	.07 (.37)	0 (0)	0 (0)	0 (0)
Turns	1.42 (1.17)	1.26 (1.29)	2.68 (2.05)	1.45 (1.05)	3.10 (1.45)	4.55 (2.11)
Dwell time (seconds)			100.71 (72.70)			52.15 (17.96)
Valence			2.22 (.71) [^]			2.05 (.51) [^]

Mean number reported for each speech category; time interacting for dwell time reported in seconds. SD reported in parentheses

*The first 225 child observations did not include AA

[^]Valence for both cases was significantly higher than neutral (3); Case 1: $t(307) = -19.26, p < .001$; Case 2: $t(19) = -8.32, p < .001$

in both illustrative cases, and caregivers (Person B) doing the majority of the talking in child–caregiver dyads. We predicted that child–caregiver dyads would produce more Commands and more Encouragements, but fewer conversational turns than in adult–adult conversations. Each of these hypotheses were supported, see Table 7. We also predicted that caregiver–child dyads would have a relatively positive valence; indeed, overall valence was significantly different from neutral, $t(307) = -19.26, p < .001, d = 1.10$. Importantly, the majority of the variables included in the coding system were identified, suggesting that the TDC does indeed capture critical elements of dyadic conversation. Moreover, the majority of the specific hypotheses proposed were also confirmed, again giving weight to the validity of the TDC to capture variables of interest.

Inter-rater reliability The standard TDC coding system was used with the addition of a few modifications, as described below. On three¹ of these observed dyads, two trained coders made independent observations for reliability purposes. Each coder's responses were then compared using both a mean, two-way, random absolute effects ICC model and percent agreement. The ICC was good to excellent for the quantity of utterances and conversational turns that occurred

¹ The small number of dual-coded dyads was due to limited personnel, partially related to the COVID-19 pandemic and limitations on staff. Because of the high number of dual-coded dyads in the second illustrative case, we did not add further personnel here.

(Cicchetti, 1994; Cicchetti et al., 2006; Fleiss, 1986), see Table 8. This suggests that the TDC coding system has high inter-rater reliability for quantity of talk. However, because the TDC coding also captures the quality of utterances as defined by speech categories, conversational turn, speaker, and orders of utterances, a total percentage of agreement between coders was also calculated. In this case, agreement was counted when the utterance type, speaker, and order all matched. There was 89% agreement (SD = 13%) between the coders on speech categories, 70% agreement (SD = 26%) on conversational turns, and 100% agreement on valence and demographics. Broken down by observed dyad, both coders agreed 100% on the first dyad, and on the second dyad one coder missed one conversational turn and one Statement. The final dyad had two missing conversational turns, two speech categories that were substituted for a different type (C for a Q, C for S), and one missing S. Taken together, these results show that the TDC has excellent inter-rater reliability for quantity and very good reliability for quality.

Modifications While the main advantage of TDC is its context-independence, modifications to the standard TDC can be made in order to complement context-specific research questions. There were two main modifications to the standard TDC: the addition of a new coding type (Announcements; AA) and the notation of exhibit signage. Both are elaborated below.

Adding Announcement speech category The first modification to the standard TDC was the additional category of

Table 7 Regressions comparing Illustrative Case 1 vs. Illustrative Case 2

Variable Model	β	SE	<i>t</i>	<i>p</i>
Qwh Model				
Illustrative Case	−.03	.07	−.40	.691
Speaker	.16	.05	2.88	.004
Illustrative Case*Speaker	.29	.05	5.38	<.001
Qyn Model				
Illustrative Case	−.05	.09	−.58	.562
Speaker	.02	.07	.28	.779
Illustrative Case*Speaker	.26	.07	3.71	<.001
S Model				
Illustrative Case	1.52	.27	5.65	<.001
Speaker	−.67	.21	−3.21	.001
Illustrative Case*Speaker	−.61	.21	−2.90	.004
C Model				
Illustrative Case	−.53	.13	−3.97	<.001
Speaker	−.25	.11	−2.42	.02
Illustrative Case*Speaker	.43	.11	4.08	<.001
I Model				
Illustrative Case	−.32	.07	−4.65	<.001
Speaker	−.01	.05	−.18	.855
Illustrative Case*Speaker	.01	.05	.18	.855
A Model				
Illustrative Case	.20	.08	2.62	.009
Speaker	−.31	.06	−5.22	<.001
Illustrative Case*Speaker	−.19	.06	−3.18	.002
AA Model				
Illustrative Case	−.04	.07	−.64	.524
Speaker	.05	.06	.85	.397
Illustrative Case*Speaker	−.05	.06	−.85	.397
Ee Model				
Illustrative Case	−.13	.06	−2.06	.044
Speaker	−.13	.05	−2.35	.019
Illustrative Case*Speaker	.13	.05	2.35	.019
Ec Model				
Illustrative Case	−.15	.02	−5.99	<.001
Speaker	.01	.02	.47	.642
Illustrative Case*Speaker	−.01	.02	−.47	.642
U Model				
Illustrative Case	−.01	.02	−.72	.480
Speaker	.02	.02	.83	.405
Illustrative Case*Speaker	−.02	.02	−.83	.405
Turns				
Illustrative Case	.62	.12	4.99	<.001
Speaker	−.37	.10	−.38	<.001
Illustrative Case*Speaker	−.45	.10	−4.66	<.001
Dwell Time				
Illustrative Case	−19.43	9.88	−1.97	.050
Valence				
Illustrative Case	−.64	.09	−6.96	<.001

Individual regressions were run for each speech category. Coder was included as a random effect. Significant findings shown in bold

Announcements (AA) in order to capture attention or orient an individual, which was previously imperfectly captured as either an Acknowledgements (A) or Declarative Statement (S). After Announcements were added, it was seen an average of .85 times per interaction by both the child and adult, representing approximately 4% of the total talk. After adding Announcements, we would predict relative decreases in the quantity of Statements and Acknowledgments, but fewer differences in the amount of the other talk categories. This hypothesis was partially supported as there were no differences in any speech category pre- vs. post-modification. See Supplementary Materials, Table S.

Thus, the modification of adding Announcement as a category does indeed appear to capture a present and unique element of talk without changing the presence of other core conversational elements.

One reason we may not have seen decreases in Statements after the modification is that there was a shift in the overall amount of talk that occurred before the modification of adding Announcement and post-Announcement; more talk occurred in dyads observed after the modification, $t(93.35) = -3.20$, $p = .002$. One possibility for this difference is a historical threat—pre-Announcement data was collected before the COVID-19 global pandemic began (June 2019), whereas the post-Announcement data was collected while COVID restrictions were in effect (November 2020–October 2021). During the height of the COVID pandemic, the museum implemented restrictions on who and how many people could attend; museum members (who are more likely to be of higher socioeconomic status (SES) and talk with their children more; Ridge et al., 2015) had priority for time slots, and fewer dyads were allowed inside at a time, leading to a quieter, less busy environment. This could have changed the dynamic of the interaction and increased the amount of talk. For reasons like this it remains important for researchers modifying the TDC to provide their validity metrics for their modifications.

As a final step to confirm that the TDC remains reliable even with the addition of Announcement and changes in exhibit interactions during COVID, two subsets of post-modification data were compared to offer another metric of reliability—a set of 24 child–caregiver dyads observed shortly after the implementation of Announcements (in the Spring of 2021) and a set of 60 similar child–caregiver dyads observed six months later by a different coder but with similar exhibits and COVID circumstances. The quality and quantity of talk between these subsets was nearly identical (see Supplementary Materials, Table S2), again offering support for the reliability of the TDC.

Adding variable for exhibit signage A second modification to the standard TDC was the addition of a variable marking signage. In this particular case, exhibits within the children’s

Table 8 Intraclass correlations between coders

Item	Illustrative Case 1 (caregiver–child)				Illustrative Case 2 (adult–adult)			
			F test with true value 0				F test with true value 0	
	ICC	95% CI	Value (df 1, df 2)	P value	ICC	95% CI	Value (df 1, df 2)	P value
Overall # of utterances	.996	[.835, 1.0]	170.33 (2, 2)	.006	.987	[.965, .995]	69.25 (19, 19)	<.001
# of Conv. turns	.889	[−.801, .997]	9.0 (2, 2)	.10	.937	[.84, .98]	15.38 (19, 19)	<.001

Calculated using SPSS, based on mean-rating, absolute agreement two-way random effects model

museum varied in the amount of directive signage that was present; some exhibits had no signage and were purely exploration-based, some had minimal signs, and others were heavy on signs and explanations. Prior work suggests that signs significantly increase the quantity of talk that occurs by providing prompts for caregivers (Ridge et al., 2015) as well as increasing the amount of time spent engaging and the quality of talk (Callanan et al., 2020), which is often defined as conversational turns. Thus, these predictions were tested here. Differences in valence across exhibits were not expected to differ and are offered as a control prediction. All hypotheses were confirmed (see Supplemental Materials, Table S3), most notably that overall amount of talk and conversation turns increase when there is some level of signage present. Once again, this suggests that the TDC coding reliably captures critical elements of conversation, paralleling expected patterns even with the addition of extra variables.

Illustrative Case 2: Adult-adult dyad in formal public discussion

Overview and adaption of coding system The TDC was used to code the interactions between a host and guest in a university-sponsored webinar series in order to use the coding in a formal context between two adults. The webinar series involved a discussion between the host and guest—a local researcher—about their recent publications, presentations, and ongoing research. The host was the same across all webinar episodes. In this setting, Person A was the host of the meeting and Person B was the guest speaker. Interactions after the formal presentation (the discussion session) were coded for the illustrative case. In addition to coding utterances and turns in real time, raters recorded presenting demographic information (age, sex, race) to the best of their ability, as well as the duration and any relevant notes about the interaction.

Participants A total of 20 adult–adult dyads were observed as they conversed in a formal setting. See Table 5 for demographics.

Data and results

Overall findings Descriptive results are reported in Table 6. Again, the most common utterance was a Declarative Statement, followed by Acknowledgements and Questions. Irrelevant, Announcements, and Evaluative utterances were nonexistent in these formal adult dyadic conversations. As noted above, we hypothesized that adult–adult dyads would include more Declarative Statements and Acknowledgements than caregiver–child talk, which was confirmed here. It was also predicted that valence in these formal settings would be more neutral, which was not the case, $t(19) = -8.32$, $p < .001$, $d = 1.86$. This is further discussed below. As in the first illustrative case, the majority of the variables of interest were identified during the conversation and the majority of the specific hypotheses proposed were also confirmed. Together with Illustrative Case 1, this again suggests that the TDC may be a valid yet flexible approach for coding naturalistic dyadic interactions.

Inter-rater reliability Because Illustrative Case 2 utilized a recorded webinar, all 20 observations were able to be coded by two trained research assistants. The mean, two-way, random absolute effects intraclass correlation model for the quantity of utterances and conversational turns that occurred was good to excellent, based on standards in the field (Cicchetti, 1994; Cicchetti et al., 2006; Fleiss, 1986); see Table 7. This again suggests that the TDC has high inter-rater reliability for quantity of talk. As before, in order to capture the reliability of conversational quality, a total percentage of agreement (matching in utterance type, speaker, and order) between coders was also calculated. There was 95% (SD = 7%) agreement between the coders on speech categories, 93% (SD = 13%) agreement on conversational turns, and 100% agreement on valence and demographics. Overall, the TDC has excellent inter-rater reliability for both quantity and quality of talk.

Illustrative cases summary

Most importantly, both illustrative cases demonstrate that the TDC captures the quantity and quality of speech during social interaction. Across 300 dyads and conversations, every utterance was observed and coded with less than .05% of utterances coded as Unknown and excellent reliability between raters. Moreover, with the exception of hypotheses regarding valence, the majority of predictions were supported. The ability to accurately capture speech combined with the expected differences between contexts being observed suggests that the TDC is a viable option for observing the quality and quantity of talk within dyadic interactions.

Discussion

Dyadic conversations are complex yet ubiquitous social interactions. Children speak to their caregivers about the things they experience in the world. Adolescents build relationships and communicate hardships as well as their good fortunes, and colleagues discuss different ways to solve some of the world's most imminent problems. As commonplace and mundane as conversations seem, a large body of research suggests that there is much more to these social interactions than meets the eye. However, there is a lack of methods able to systematically evaluate conversations where they occur most genuinely and most often: natural settings. Although there are plenty of published coding systems capable of analyzing a dyadic interaction, many of them are unable to do so without sacrificing reactivity effects. Moreover, many coding mechanisms can be expensive, time-consuming, context-dependent, and even exclusive.

The TDC approach proposed here fills this gap. It is simple in its design and function. An interaction can be coded with a pen and paper by a rater with little training, and from that interaction, the TDC allows us to assess how rich a conversation is by the number of utterances spoken and the length of the conversation. It also allows us to see the depth of a particular interaction through the use of questions being asked, commands given, turns taken, and more.

Results from Illustrative Cases 1 and 2 demonstrate that the TDC's speech categories are inclusive of virtually all types of utterances, and that these speech categories are detectable across unparalleled contexts such as informal child–caregiver interactions and formal interactions between adults. Further, we can see that the TDC captures how interlocutors within an interaction can change the dynamic of a conversation with regard to types of utterances used, their frequency, and the valence of the interaction. Excellent interrater reliability reinforces the notion that the TDC is very easy to learn and accurate in live, natural settings. Thus,

the TDC provides a crucial, structured foundation for future research questions involving dyadic interactions.

Limitations and future directions

Due to its novelty, many of the factors limiting the TDC have an opportunity to be resolved over time. Since the TDC is primarily used in natural settings, data collection was slightly stunted by COVID-19 regulations. This contributed to a general lack of social interactions in public spaces, which in turn reduced opportunities to observe interactions and access raters. These restrictions resulted in fewer dual coders for Illustrative Case 1 and the use of video interactions for data collection in Illustrative Case 2. A bonus, however, is that this suggests that TDC can be easily adapted to different modalities and capture interactions even if virtual. Nonetheless, because of their methods and the flexibility of the TDC to adapt to various outcomes, naturalistic observational approaches such as this provide some initial validity data. However, consistent with standards for assessment validity in the field, validity and reliability ought to be confirmed for each unique study (see also Chmielewski et al., 2015). While the TDC offers a robust foundation, future studies should still include their own metrics for validity such as a priori hypotheses, criterion metrics for the outcome of interest, and/or assessments of convergent and/or divergent validity depending on the target construct.

Additionally, while valence differed across the two contexts, it did so in the opposite direction as predicted, such that adult–adult conversations were rated as more positively valenced than child–caregiver interactions. While the contexts were similar in their goals—both were public settings with potential audiences, and both were teaching/learning situations in which participants had autonomy to participate in the discussion and did so under their own volition—they differed in their participant demographics and activities. There is evidence suggesting that agentic motivation can lead to increases in positive affect (Gherghel et al., 2020). One can argue that in child–adult conversation, agentic motivation would be present to a lesser extent given children are often being directed by their caregivers to engage in specific activities—indeed, there were higher rates of Commands in child–adult dyads than in adult–adult conversations. Thus, while the average valence for both contexts would be neutral–high as was seen here, it could be slightly higher in the adult–adult dyad. Nonetheless, it is also feasible that the metric of valence ought to be more nuanced in order to capture subtle differences—indeed, there were very few scores of 1 or 5 given across the entirety of observations. Some scholars have also suggested that positive affect and negative affect are independent constructs and thus should be measured on separate metrics (Cacioppo & Berntson, 1994; Kaplan, 1972). Similar work has also elucidated the difficulty in capturing ambivalent attitudes on bipolar attitude scales (Cacioppo et al., 1997). As such, the coding manual for

the TDC was updated to address this, and future users of the TDC may want to include two unipolar measures of positive and negative valence.

Ethical considerations in community sampling

Although the samples from each illustrative case were diverse with regard to sex and age across vastly different settings, they had similarly little racial diversity as many other research samples. As mentioned previously, the use of the TDC for unobtrusive/noninvasive observational studies permits access to populations of color and people of low socioeconomic status whose voices and experiences are often excluded from participation in research, and their inclusion is vital. However, even when access to people of color and their communities is possible, it does not necessarily mean access is granted. In order to take advantage of this benefit the TDC offers, the inclusion of people from racially, ethnically, and culturally diverse backgrounds must be intentional. Still, future studies using the TDC for unobtrusive observations may potentially receive exemption from comprehensive review from their respective institutions. In these cases, ethical considerations, especially those pertaining to marginalized communities, should be maintained throughout the research process in its entirety. Research must be rooted in the community, and the flexibility of the TDC should not be used to conduct research disengaged from the communities, cultures, and experiences it investigates. Even in the first illustrative case here, despite being deemed exempt from extended institutional review, the objectives, pragmatics, and procedures associated with observations were discussed, created, and implemented in collaboration with program leaders and museum staff. It is recommended that researchers interested in using the TDC to investigate constructs within a given community follow concordant ethical procedures, engaging with communities of interest with empathic and holistic diligence. Moreover, it is recommended that researchers explore best practices for engaging with marginalized communities in literature that details meaningful considerations when engaging with particularly vulnerable social groups, such as those discussed by Fassinger and Morrow (2013), Potnis and Gala (2020), and Woodley and Lockard (2016).

Although more robust coding systems exist for more detailed components of social interactions, the ability to systematically identify, define, and organize every verbal utterance and turn of a dyadic conversation with minimal experience and/or resources is uniquely useful. Future research could examine how power, identity, class, context, and many other factors shape and change the way we interact with others. This is only possible because of the flexibility of the TDC. Any verbal or nonverbal utterance that has not already been identified can be identified, and any new additions or

modifications can be tailored specifically to fit a researcher's needs. It is even feasible to use the TDC for nonverbal text exchanges or, as demonstrated in Illustrative Case 2, virtual calls. Moreover, the TDC allows for conversations to be measured where conversations happen, when conversations happen, and how conversations happen. Without the noise that labs, research assistants, tape recorders, and video cameras bring to data from conversations in contrived settings, unobtrusive observations of natural conversations in natural settings grants researchers access to the true nature of dyadic interactions.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.3758/s13428-022-02033-w>.

Acknowledgements We wish to thank the Language, Learning, and Development Lab of Oklahoma State University as well as Tulsa's Discovery Lab and Children's Museum. We would also like to thank Lauren Ezell, Michael Chmielewski, Jordan Newell, and especially Lynsey Childress for their meaningful contributions.

Code availability Not applicable.

Data availability All coding manuals, training videos, and data are available via OSF https://osf.io/82xu5/?view_only=c7f78024499d4f62a19287d21135d9f6. The technique and methods were not preregistered.

Declarations

Conflicts of Interest Both authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Ethics approval The current study was reviewed by the Oklahoma State University Research Ethics Committee and deemed as exempt from IRB oversight.

Consent to participate Not applicable.

Consent for publication Not applicable.

References

- Bales, R. F. (1950). *Interaction process analysis: A method for the study of small groups*. Addison-Wesley.
- Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes*, 50(2), 248–287. [https://doi.org/10.1016/0749-5978\(91\)90022-L](https://doi.org/10.1016/0749-5978(91)90022-L)
- Bandura, A., Adams, N. E., & Beyer, J. (1977). Cognitive processes mediating behavioral change. *Journal of Personality and Social Psychology*, 35(3), 125–39. <https://doi.org/10.1037//0022-3514.35.3.125>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin*, 115(3), 401–423.

- Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1997). Beyond bipolar conceptualizations and measures: The case of attitudes and evaluative space. *Personality and Social Psychology Review*, 1(1), 3–25.
- Callanan, M. A., Legare, C. H., Sobel, D. M., Jaeger, G. J., Letourneau, S., McHugh, S. R., & Watson, J. (2020). Exploration, explanation, and parent–child interaction in museums. *Monographs of the Society for Research in Child Development*, 85(1), 7–137. <https://doi.org/10.1111/mono.12412>
- Carver, C. S., & Scheier, M. F. (1981). Self-consciousness and reactance. *Journal of Research in Personality*, 15(1), 16–29. [https://doi.org/10.1016/0092-6566\(81\)90003-9](https://doi.org/10.1016/0092-6566(81)90003-9)
- Chmielewski, M., Clark, L. A., Bagby, R. M., & Watson, D. (2015). Method matters: Understanding diagnostic reliability in DSM-IV and DSM-5. *Journal of Abnormal Psychology*, 124(3), 764.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284–290. <https://doi.org/10.1037/1040-3590.6.4.284>
- Cicchetti, D., Bronen, R., Spencer, S., Haut, S., Berg, A., Oliver, P., & Tyrer, P. (2006). Rating scales, scales of measurement, issues of reliability: Resolving some critical issues for clinicians and researchers. *The Journal of Nervous and Mental Disease*, 194(8), 557–564. <https://doi.org/10.1097/01.nmd.0000230392.83607.c5>
- Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22(1), 1–39. [https://doi.org/10.1016/0010-0277\(86\)90010-7](https://doi.org/10.1016/0010-0277(86)90010-7)
- Coan, J. A., & Gottman, J. M. (2007). The specific affect coding system (SPAFF). In J. A. Coan & J. M. Gottman (Eds.), *Handbook of emotion elicitation and assessment* (pp. 267–285). Oxford University Press Inc.
- Corbie-Smith, G., Thomas, S. B., & George, D. M. M. S. (2002). Distrust, race, and research. *Archives of Internal Medicine*, 162(21), 2458–2463. <https://doi.org/10.1001/archinte.162.21.2458>
- Donnelly, S., & Kidd, E. (2021). The longitudinal relationship between conversational turn-taking and vocabulary growth in early language development. *Child Development*, 92(2), 609–625. <https://doi.org/10.1111/cdev.13511>
- Duran, N. D., Paxton, A., & Fusaroli, R. (2019). ALIGN: Analyzing linguistic interactions with generalizable techniques - A python library. *Psychological Methods*, 24(4), 419–438. <https://doi.org/10.1037/met0000206>
- Eyberg, M. S., & Robinson, A. E. (2000). *Dyadic parent-child interaction coding system: A manual*. [Unpublished manuscript]. Department of Family and Child Nursing, University of Washington.
- Farmer, D. F., Jackson, S. A., Camacho, F., & Hall, M. A. (2007). Attitudes of African American and low socioeconomic status white women toward medical research. *Journal of Health Care for the Poor and Underserved*, 18(1), 85–99. <https://doi.org/10.1353/hpu.2007.0008>
- Fassinger, R., & Morrow, S. L. (2013). Toward best practices in quantitative, qualitative, and mixed-method research: A social justice perspective. *Journal for Social Action in Counseling and Psychology*, 5(2), 69–83. <https://doi.org/10.33043/JSACP.5.2.69-83>
- Fleiss, J. (1986). *The Design and Analysis of Clinical Experiments*. Wiley.
- Freed, A. F. (1994). The form and function of questions in informal dyadic conversation. *Journal of Pragmatics*, 21(6), 621–644. [https://doi.org/10.1016/0378-2166\(94\)90101-5](https://doi.org/10.1016/0378-2166(94)90101-5)
- Gherghel, C., Nastas, D., Hashimoto, T., Takai, J., & Cargile, A. C. (2020). Culture, morality, and the effect of prosocial behavior motivation on positive affect. *Ethics & Behavior*, 30(2), 126–149. <https://doi.org/10.1080/10508422.2019.1651651>
- Girolametto, L., Weitzman, E., Lieshout, R. V., & Duff, D. (2000). Directiveness in teachers' language input to toddlers and preschoolers in daycare. *Journal of Speech, Language, and Hearing Research*, 43(5), 1101–1114. <https://doi.org/10.1044/jslhr.4305.1101>
- Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M., Golinkoff, R. M., Pace, A., Yust, P. K. S., & Suma, K. (2015). The Contribution of Early Communication Quality to Low-Income Children's Language Success. *Psychol Sci*, 26(7), 1071–1083. <https://doi.org/10.1177/0956797615581493>
- Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Development*, 73(2), 418–433. <https://doi.org/10.1111/1467-8624.00415>
- Jimenez-Buedo, M., & Guala, F. (2016). Artificiality, reactivity, and demand effects in experimental economics. *Philosophy of the Social Sciences*, 46(1), 3–23. <https://doi.org/10.1177/0048393115618015>
- Julien, D., Markman, H. J., & Lindahl, K. M. (1989). A comparison of a global and a microanalytic coding system: Implications for future trends in studying interactions. *Behavioral Assessment*, 11(1), 81–100.
- Kaplan, K. J. (1972). On the ambivalence-indifference problem in attitude theory and measurement: A suggested modification of the semantic differential technique. *Psychological Bulletin*, 77(5), 361–372.
- Keating, E., & Egbert, M. (2004). Conversation as a cultural activity. In A. Duranti (Ed.), *A companion to linguistic anthropology* (pp. 169–196). Blackwell.
- Kemler-Nelson, D. G., Egan, L. E., & Holt, M. B. (2004). When children ask, “what is it?” What do they want to know about artifacts? *Psychological Science*, 15(6), 384–389. <https://doi.org/10.1111/j.0956-7976.2004.00689.x>
- King, A. L. S., Valenca, A. M., Silva, A. C. O., Baczynski, T., Carvalho, M. R., & Nardi, A. E. (2013). Nomophobia: Dependency on virtual environments or social phobia? *Computers in Human Behavior*, 29(1), 140–144. <https://doi.org/10.1016/j.chb.2012.07.025>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- LENA. (2022). *Journal bibliography-research publications*. Lena.org. <https://www.lena.org/research/>
- Lenth, R. V. (2016). Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software*, 69(1), 1–33. <https://doi.org/10.18637/jss.v069.i01>
- Leung, L. (2007). Unwillingness-to-communicate and college students' motives in SMS mobile messaging. *Telematics and Informatics*, 24(2), 115–129. <https://doi.org/10.1016/j.tele.2006.01.002>
- Lu, X., Watanabe, J., Liu, Q., Uji, M., Shono, M., & Kitamura, T. (2011). Internet and mobile phone text-messaging dependency: Factor structure and correlation with dysphoric mood among Japanese adults. *Computers in Human Behavior*, 27(5), 1702–1709. <https://doi.org/10.1016/j.chb.2011.02.009>
- Maroni, B., Gnisci, A., & Pontecorvo, C. (2008). Turn-taking in classroom interactions: Overlapping, interruptions and pauses in primary school. *Eur J Psychol Educ*, 23(1), 59–76. <https://doi.org/10.1007/BF03173140>
- Miller, D. B. (1977). Roles of naturalistic observation in comparative psychology. *American Psychologist*, 32(3), 211. <https://doi.org/10.1037/0003-066X.32.3.211>
- Ochs, E. (2004). Narrative lessons. In A. Duranti (Ed.), *A companion to linguistic anthropology* (pp. 269–289). Blackwell.
- Pasch, L. A., Harris, K. W., Sullivan, K. T., & Bradbury, T. N. (2004). The social support interaction coding system. In *Couple observational coding systems* (pp. 333–348). Routledge.
- Potnis, D., & Gala, B. (2020). Best practices for conducting fieldwork with marginalized communities. *Information Processing & Management*, 57(3), 1021–1044.
- R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>

- Reed, J., Hirsh-Pasek, K., & Golinkoff, R. M. (2017). Learning on hold: Cell phones sidetrack parent-child interactions. *Developmental Psychology, 53*(8), 1428. <https://doi.org/10.1037/dev0000292>
- Ridge, K. E., Weisberg, D. S., Ilgaz, H., Hirsh-Pasek, K. A., & Golinkoff, R. M. (2015). Supermarket speak: Increasing talk among low-socioeconomic status families. *Mind, Brain, and Education, 9*(3), 127–135. <https://doi.org/10.1111/mbe.12081>
- Roberts, S. O., Bareket-Shavit, C., Dollins, F. A., Goldie, P. D., & Mortenson, E. (2020). Racial inequality in psychological research: Trends of the past and recommendations for the future. *Perspectives on Psychological Science, 15*(6), 1295–1309. <https://doi.org/10.1177/1745691620927709>
- Robinson, E. A., & Eyberg, S. M. (1981). The dyadic parent-child interaction coding system: Standardization and validation. *Journal of Consulting and Clinical Psychology, 49*(2), 245. <https://doi.org/10.1037/0022-006X.49.2.245>
- Romero, V., Amaral, J., Fitzpatrick, P., Schmidt, R. C., Duncan, A. W., & Richardson, M. J. (2017). Can low-cost motion-tracking systems substitute a Polhemus system when researching social motor coordination in children. *Behavior Research Methods, 49*(2), 588–601. <https://doi.org/10.3758/s13428-016-0733-1>
- Sapacz, M., Rockman, G., & Clark, J. (2016). Are we addicted to our cell phones? *Computers in Human Behavior, 57*, 153–159. <https://doi.org/10.1016/j.chb.2015.12.004>
- Scharff, D. P., Mathews, K. J., Jackson, P., Hoffsuemmer, J., Martin, E., & Edwards, D. (2010). More than Tuskegee: Understanding mistrust about research participation. *Journal of Health Care for the Poor and Underserved, 21*(3), 879–97. <https://doi.org/10.1353/hpu.0.0323>
- Schegloff, E. A. (1982). Discourse as an interactional achievement: Some uses of 'uh huh' and other things that come between sentences. In Tannen, D. (Eds.), *Analyzing discourse: Text and talk* (pp.71–93). Georgetown University Press.
- Scheier, M. F., & Carver, C. S. (1988). A model of behavioral self-regulation: Translating intention into action. *Advances in Experimental Social Psychology, 21*, 303–346. [https://doi.org/10.1016/S0065-2601\(08\)60230-0](https://doi.org/10.1016/S0065-2601(08)60230-0)
- Shockley, K., Richardson, D. C., & Dale, R. (2009). Conversation and coordinative structures. *Topics in Cognitive Science, 1*(2), 305–319. <https://doi.org/10.1111/j.1756-8765.2009.01021.x>
- Steensig, J., & Drew, P. (2008). Introduction: Questioning and affiliation/disaffiliation in interaction. *Discourse Studies, 10*(1), 5–15. <https://doi.org/10.1177/1461445607085581>
- Stiles, W. B. (1978). Verbal response modes and dimensions of interpersonal roles: A method of discourse analysis. *Journal of Personality and Social Psychology, 36*(7), 693. <https://doi.org/10.1037/0022-3514.36.7.693>
- Willard, A. K., Busch, J. T., Cullum, K. A., Letourneau, S. M., Sobel, D. M., Callanan, M., & Legare, C. H. (2019). Explain this, explore that: A study of parent-child interaction in a children's museum. *Child Development, 90*(5), e598–e617. <https://doi.org/ghz5h3>
- Woodley, X. M., & Lockard, M. (2016). Womanism and Snowball Sampling: Engaging Marginalized Populations in Holistic Research. *The Qualitative Report, 21*(2), 321–329. <https://doi.org/jmsh>
- Zimmerman, F. J., Gilkerson, J., Richards, J. A., Christakis, D. A., Xu, D., Gray, S., & Yapanel, U. (2009). Teaching by Listening: The Importance of Adult-Child Conversations to Language Development. *Pediatrics, 124*(1), 342–349. <https://doi.org/10.1542/peds.2008-2267>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.