EnGaze: Designing Behavior Visualizations with and for Behavioral Scientists

Ha Kyung Kong[1*], John Lee[1*], Jie Ding[2], Karrie Karahalios[1]
[1]Department of Computer Science
[2]School of Art and Design
University of Illinois at Urbana Champaign
{ hkong6, lee98, jieding2, kkarahal } @ illinois.edu

ABSTRACT
Joint attention is widely recognized as an important developmental milestone for children, and experts consider a lack of joint attention a defining characteristic of autism spectrum disorders (ASDs). While clinicians and researchers agree on the importance of joint attention, their definitions and methods for assessing joint attention vary. In this paper, we present the design process and the evaluation of EnGaze, a visualization-based Web tool for dyadic communicative behavior that highlights commonly discussed features of joint attention. While such visualization styles are not yet the norm in the clinical practices of behavioral and developmental psychology, we argue they should be and find that the introduction of these visual artifacts helped clinicians and researchers conceptualize their personal joint attention rules. Researchers envisioned a number of uses for EnGaze in their personal workflow, including identifying atypical communication patterns and providing a visual record for tracking behavior. The contributions of this paper are 1) an interactive visualization for exploring joint attention, 2) the documentation of an iterative design process for a clinical visualization tool, 3) illustrations of EnGaze use cases with active practitioners in the behavioral sciences, and (4) the discussions surrounding the implications of introducing such visualizations in behavioral science communities.

Author Keywords
Dyadic Interaction; Behavioral Visualization; Joint Attention; Behavioral Imaging; Autism; Webtool Design

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

*Both authors contributed equally to this work.

DIS 2016, June 04 - 08, 2016, Brisbane, QLD, Australia

ACM 978-1-4503-4031-1/16/06...$15.00
DOI: http://dx.doi.org/10.1145/2901790.2901870

INTRODUCTION
Joint attention is widely recognized as an important developmental milestone for children. Substantial prior research frequently cites the lack of joint attention as a defining characteristic of ASD [25, 28]. The study of joint attention is particularly important in young children, as it can contribute to early detection of ASD [3, 38]. Furthermore, research shows that early intervention is key for minimizing the impact of developmental delays in children with ASD [11, 19].

In this paper, we introduce EnGaze, a Web tool for visualizing coordinated, communicative behavior focused on exploring and identifying joint attention signals between a child and an examiner. Using behavior data from 200 Rapid Attention Back and Forth Communication Test (Rapid ABC) [33] child-clinician sessions, we iteratively designed EnGaze with feedback from a team of ten ASD clinicians and researchers. Over the period of a year, Engaze evolved from a point-line visualization into an interactive timeline visualization.

To evaluate EnGaze, we conducted a user study with five external ASD clinicians and researchers to assess its efficacy for identifying children in need of further assessment and to see if EnGaze fits into their research or clinical practice. When presented with EnGaze, the clinicians and researchers formed their own strategies for exploring the visualizations and independently identified the same children in need of further evaluation. The study participants discovered that the visualizations reveal unique dynamics of a child's behavior that cannot be easily observed by traditional methods. They further acknowledged that such visualizations were largely unexplored in their disciplines and might be met with some resistance due to tradition, yet they embraced the approach and expressed excitement at the possibilities. Clinicians envisioned a number of important uses for EnGaze, including identifying atypical communication patterns, serving as a common talking point with parents, and providing a longitudinal visual record for children's behavior and development.

The contributions of this paper are 1) an interactive visualization for exploring joint attention, 2) the documentation of an iterative design process for a clinical visualization tool, 3) illustrations of EnGaze use cases with active practitioners in the behavioral sciences, and (4) the

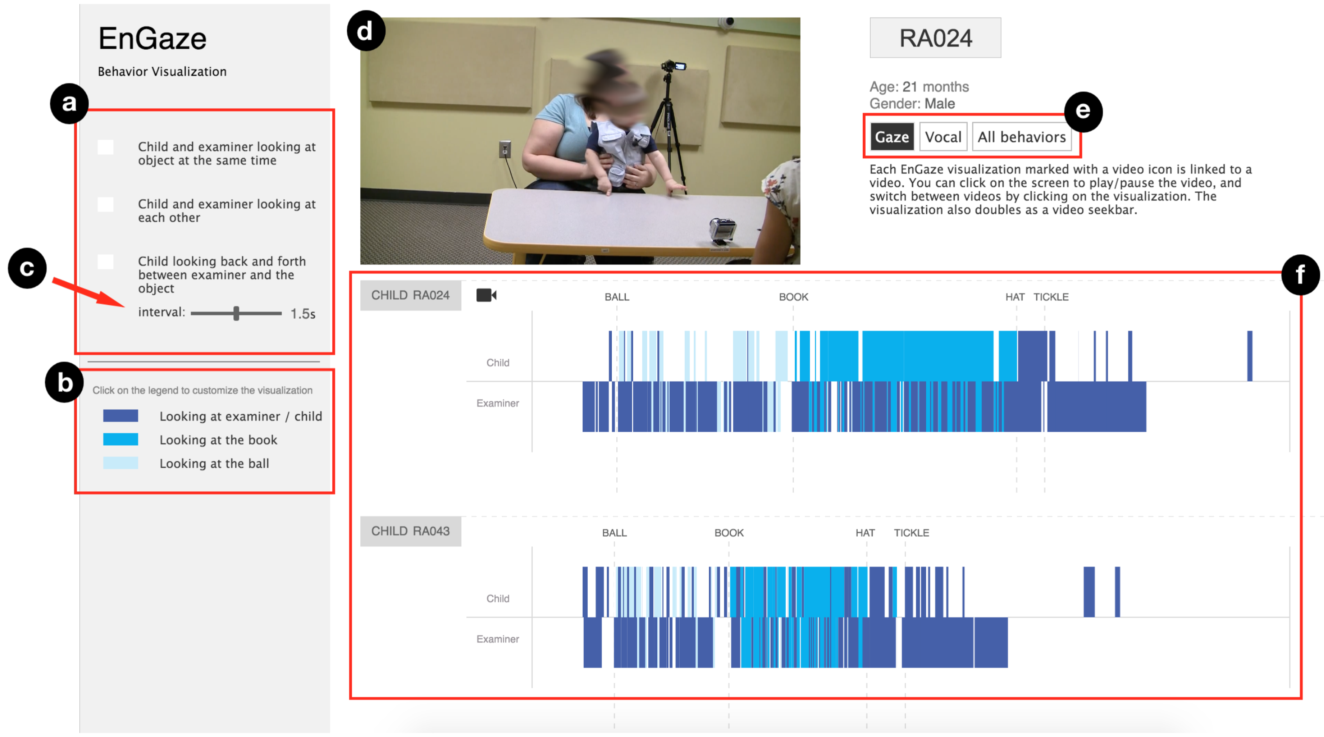

Figure 1. The main view of EnGaze. This image highlights several features of the tool: (a) a control panel where the user can toggle the authoring options offered for each view, (b) the interactive legend where the user can toggle the visibility of specific behaviors in the visualization, (c) a slider to control the delay tolerance of a child's back-and-forth glances, (d) the video, (e) the buttons to select between the three modes, and (f) the visualizations. Note that the top visualization in (f) where the child does not look at the examiner for long periods of time, was identified by all our study participants as belonging to a child that needed further assessment.

discussions surrounding the implications of introducing such visualizations in behavioral science communities.

RELATED WORK

The design of EnGaze builds on related work on data visualization, joint attention, and existing diagnostic instruments for ASD. We present this related work in this section.

Data Collection and Visualizing Behavior

Abaris, created by Kientz et al. facilitated the automated capture of data during therapy sessions for clinicians and therapists [22]. Abaris provided a simple and effective method for collecting behavioral data and presented this data as line graphs. EnGaze extends the visualization repertoire for behavioral data to represent behaviors using accessible, alternative visualization forms that emphasize the dyadic nature of the clinician-child relationship. Existing health-related visualizations focus primarily on visualizing personal medical histories [34, 41] or patient treatment patterns and outcomes for individuals [26]. Commercial products such as Fitbit and Nike's Fuelband explore the the visualization of health-related behavioral data (e.g., steps walked, steps climbed) [1, 2]. These visualization interfaces focus on activity data, whereas Engaze focuses on coordinated *communicative* behaviors.

Work in communicative behavior visualization to date has often been limited to one aspect of behavior, such as vocalization (e.g. [5, 15, 13]) or eye gaze (e.g. [6]). TipoVis, a behavior analysis tool, visualized two communicative modalities or behaviors chosen by the viewer, emphasizing their overlap [16]. Plexlines combined multiple communicative behaviors (gaze, vocalization, and gesture) into a single view [23]. Unlike EnGaze, Plexlines' child-centric visualization emphasizing the childrens' responses to examiner bids did not convey reciprocal actions between the child and the examiner. Like TipoVis, the only examiner behavior displayed was the examiner bid. This asymmetry in child-examiner behavior did not capture the brief dyadic micro-interactions between the child and examiner that are critical in identifying joint attention.

Diagnostic Instruments for ASD

Envisioning the rise of data collection from screenings (e.g., the five-minute Rapid ABC assessment in the pediatrician's office), we built EnGaze—not as a diagnostic tool such as the Autism Diagnostic Observation Schedule (ADOS) or Autism Diagnostic Interview-Revised (ADI-R)—but as a visualization Web tool to facilitate pre-screening for ASD. ADOS and the ADI-R are currently the most widely used diagnostic exams for ASD [24, 36]. For both ADOS and ADI-R, clinicians manually score specific behavior categories defined within the instrument through personal observation and parent interviews, respectively. With increased accessibility to sensors and recording devices, we see ASD screening instruments augmenting behavioral data observations and interviews with physiological and computationally generated data (e.g., eye gaze and waving from video, speech-like utterances from audio) [7, 17, 35].

Joint Attention

The term "joint attention" refers to a *group* of social-communicative behaviors (e.g., making eye contact with another person and pointing to something that develop between 9 and 18 months of age [9, 18, 37]. Researchers agree that joint attention involves the shared engagement of two or more individuals [4, 10, 18, 27, 40]. In the child development literature, the two individuals are typically a child and an adult. The object of the shared attention is often described as an external third entity—a person, an object, a concept, or an event such as seeing a fire truck drive by [8, 9, 27].

Although the importance of joint attention has been confirmed in previous studies [25, 28], the concept of joint attention is quite nuanced. Moore and Dunham assert that in joint attention, the child must understand that "the other participant has a focus of attention to the same entity as the self" [27]. Some are more exclusive in their definition and distinguish between *requesting* and joint attention. For example, Jones and Carr argue that the act of a child looking back and forth between a ball and her father to indicate that she wants the ball is a request and not an initiation of joint attention [18]. The reward in this case—obtaining the desired ball—is non-social. In joint attention, the reward is the social interaction between the two people attending to the same object or entity—e.g., pointing in excitement to the ball, looking at the father, and saying, "Wow! Look at that awesome ball!"

Various studies explore joint attention through different combinations of communicative modalities such as gaze, gesture, and vocalization. Gaze includes the eye contact between the examiner and the child and the gaze alternation of the child (i.e., when the child looks back and forth between the examiner and the object). Pointing at or holding up the object of interest to lead the other person's attention to the object are common ways of using gesture in joint attention. These gestures are often accompanied by a vocalization such as *"Look!"* when initializing joint attention. The modalities chosen for studies vary as the definition of joint attention leaves room for different interpretations. For example, while one study observed *gaze* and *vocalization* as the main features of joint attention [37], another focused on *gaze* and *gestures* [18]. The age of the child also plays a role in gauging the importance of different modalities. For example, gaze is the dominant modality for a six month old child, while multimodal coordination of behaviors is expected and acts as a development milestone as the child grows older [31].

Rather than defining joint attention, our research aims to provide a dynamic visual method of exploring joint attention. More specifically, we incorporated the most commonly mentioned features of joint attention into our tool and studied how clinicians and researchers used the tool to explore their own definition of joint attention.

DATA COLLECTION

The data set used in our visualization was gathered using the Rapid ABC, a five-minute, semi-structured play protocol between a clinician and a child that assesses the development of a child's communicative skills [33]. The Rapid ABC consists of five stages of play: a greeting, rolling a ball, reading a book, wearing the book as a hat, and tickling. In each stage, the examiner initiates and tries to sustain interaction with a set of predefined and consistent *examiner bids*. These examiner bids were designed to initiate a response from the child. Examples include: *"Can you turn the page?"* or *"Look at my hat!"* The examiner evaluates the child's responses to the examiner bids based on the ease of engaging the child according to the Rapid ABC specifications. Specifically, the examiner seeks to elicit social attention, back-and-forth interaction, and social communication from the child [33].

As of the writing of this paper, we have data from over 200 children, aged 9 to 30 months. Our collaborators at the Marcus Center for Autism and Georgia Tech collected the Rapid ABC data as part of a joint NSF Expeditions effort. Three independent coders hand-annotated Rapid ABC video sessions frame by frame from digital video footage captured

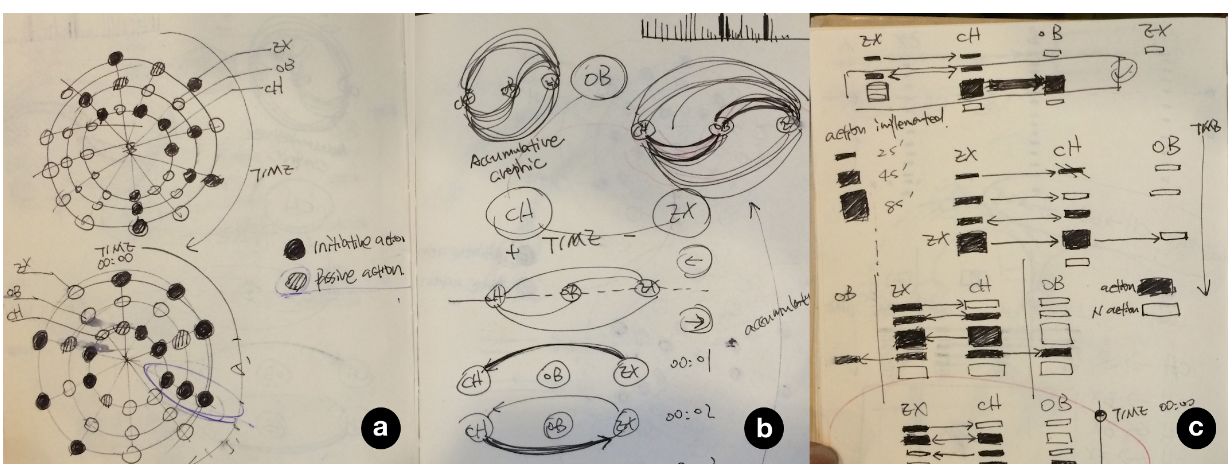

Figure 2. Early design sketches of EnGaze. (a) an early design sketch for visualizing joint attention (this design was not implemented). (b) design sketch of the second iteration. (c) design sketch of the third iteration.

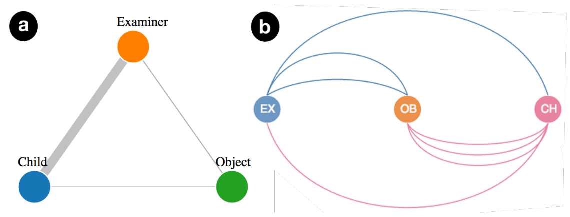

Figure 3. Two early iterations of the visualization. (a) highlights the triadic relationship between the child, examiner, and object. The thickness of the lines signifies the duration of the gaze. (b) highlights the shared attention between the child and the examiner. A line appears between either the examiner, child, or object circle when a gaze behavior connects them (e.g., child looks at the object).

at thirty frames per second. The coder training required over 90% agreement in the amount of overlap, annotation by annotation (in time), from selected sets of ten Rapid ABC sessions. The data set currently contains categorizations and sub-categorizations with timestamps of gaze, gesture, and vocalization behaviors for each child.

THE DESIGN OF ENGAZE

We were motivated by two main goals as we began the design process: (1) to create a visualization that accommodated the varying definitions of joint attention and (2) to develop a platform where clinicians and researchers could explore, analyze and share behavioral data after an assessment. In this section, we will first present the structure of EnGaze and then explain how each component stemmed from our two motivating goals.

EnGaze consists of three modes — *Gaze*, *Vocal*, and *All Behaviors*.[1] Users can switch between the modes using the buttons shown in Figure 1e. Each mode has three features: a behavior visualization, a video component, and an authoring panel. The authoring panel consists of a legend and options for highlighting specific moments during a session (Figure 1a, b, and c).

We started with initial visualization sketches that captured our two goals and lessons from the related literature. We iterated on these sketches (see examples in Figure 2) and implemented approaches recommended by clinicians and researchers in our bi-weekly online research group meetings that focused on behavior imaging. Approximately ten researchers and clinicians, including child developmental psychologists and autism researchers, participated in each of these meetings. We received feedback on all of the interactive Web tool visualizations presented in this paper over a span of one year. We also presented our designs and received feedback in a joint meeting involving researchers from ten different universities and an autism center studying computational behavior science. We describe our previous iterations and design decisions in this section.

[1]We wanted to include each key modality as a mode (gaze, vocal, and gesture). However, the examiner's gestures were not visible from the camera views that captured the Rapid ABC sessions. We therefore could only code the children's gestures. Due to the lack of symmetry, we did not include a Gesture mode.

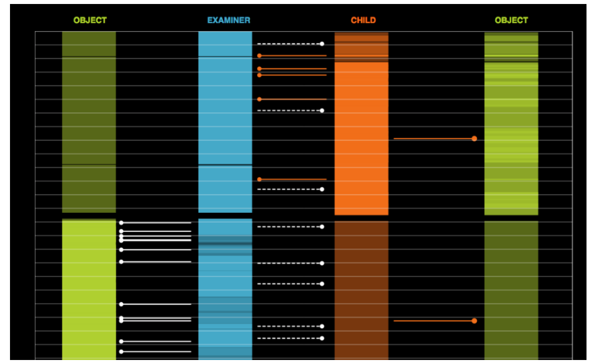

Figure 4. The third iteration of the visualization displaying the examiner's gaze and the child's gaze using rounded arrows for each second of the sessions. The color and the direction of the arrows indicate the direction of gaze. The orange arrows represent the child's gaze, and the white arrows represent the examiner's gaze. For example, the topmost white arrow indicates that the examiner was looking at the child in the first second of the session.

Visualizing Joint Attention

Our iterative design process followed three progressive stages: the visualization of gaze, the archival of temporal changes, and a persistent timeline view. We describe these stages below.

Visualizing Gaze

Previous work in joint attention revealed that gaze, especially gaze alternation, is an essential trait of joint attention [10, 12, 18, 40, 32], and thus we started by focusing on gaze and the three entities required for joint attention—the child, the examiner, and the objects (in this case, the ball and the book from the Rapid ABC). The initial design portrayed the triadic relationship between the entities using a triangle (Figure 3a). This animated visualization connected the entities with lines that varied in thickness depending on the duration of the gaze. Continuous gaze on an entity resulted in a thicker line, which thinned when the gaze stopped. The line between the child and examiner became red if they looked at each other. However, this visualization did not capture the direction of gaze when only one person gazed at the other. While most viewers found the visualization intuitive, it was difficult to observe temporal patterns in this initial triangle design as none of the temporal changes were archived in the animation.

Archiving Temporal Changes

In the next iteration, we archived the temporal patterns between the child and the examiner through cumulative renderings in the visualization; arcs were drawn between the examiner's and the child's circles when gaze behaviors occurred, and the arc renderings persisted until the animation ended (Figure 3b). The inner arcs represented early gazes, and the outer arcs represented later gazes. This provided an overall temporal summary of the child-examiner session. The visualization, however, lacked the details the clinicians needed such as the exact moment and duration of eye contact. And similar to the previous triangular visualization, the resulting summary image did not provide meaningful information about the session as a whole. From these two visualizations, we discovered that precise timing of

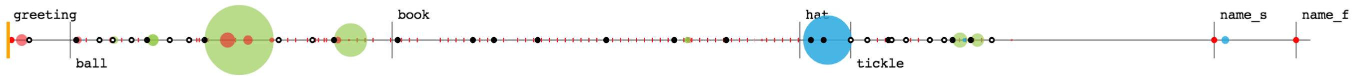

Figure 5. A Plexline visualization. The diameter of the circles corresponds to the duration of the child's behavior. Actions with longer durations are clearly visible on the timeline. Short interactions more easily visible in EnGaze are not as apparent in this visualization. The red, green, and blue circles correspond to vocalization, gesture, and gaze, respectively, similar to the EnGaze visualization.

the child's and the examiner's behavior was critical in our clinicians' and researchers' practice.

A Persistent and Comprehensive Timeline View

We then began exploring behavior across different persistent representations of time. In the following iteration, the vertical axis represented time (Figure 4). In this visualization, each column represented an entity (with the object entity duplicated in the leftmost and rightmost columns), and the arrows indicated the direction of the child's or the examiner's gaze for each depicted time interval of the session. The orange arrows represented the child's gaze, and the white arrows represented the examiner's gaze. Each session started at the top of the visualization and ended at the bottom. This facilitated the viewing and the interpretation of interactions and patterns over time. However, this visualization suffered from visual complexity and our audience struggled with interpretations. This was because the same single object was represented in two different columns. Furthermore, the visualization required considerable vertical space, making it difficult to present multiple children's behavior visualizations in one view, a feature that proved effective in the Plexlines visualization (Figure 5).

In developing EnGaze, we strived to create a timeline visualization that cogently depicted dyadic concurrent communicative behaviors, specifically micro-behaviors. We drew inspiration from a presentation of results in [43] and Plexlines. While Plexlines functions as an intuitive visualization for child responses, it primarily depicted macro-behaviors, that is, behaviors of long duration. Behaviors that occurred for very short periods of time were eclipsed by the longer behaviors in this visualization. We visualized gaze before incorporating the other behaviors. In this EnGaze visualization, the child's and the examiner's gazes are represented in parallel horizontal timeline visualizations, with the child gazes located above the examiner gazes (Figure 6). Unlike Plexlines that focused on the child's gaze patterns, EnGaze creates a reciprocal visualization that displays behaviors of both the child and the examiner. The adjacency of the child and examiner behaviors on the timeline allows for quick interpretation of the *dance* that occurs in their communication. As seen in the legend in Figure 6, the darker blue represents moments when the child is looking at the examiner or when the examiner is looking at the child. Blue and light blue are used to represent moments when the examiner or the child is looking at the book or the ball, respectively.

Since different colors are used to depict different behaviors, a connected vertical rectangular block will form across the child's strip and the examiner's strip when the child and the examiner concurrently look at the same object or each other (Figure 6c), indicating a possible joint attention moment. The visualization depicts sequential micro-behaviors through thin bars of alternating colors. For example, short back-and-forth gazes are highlighted through alternating blue and dark blue bars, as shown in Figure 6b. Such short interactions went unnoticed in visualizations such as Plexlines (Figure 5).

The EnGaze Design: Designing for Behavioral Scientists

We designed EnGaze with and for clinicians and researchers. As non-traditional visualizations are not commonly used in the areas of behavioral science that we explored, we developed EnGaze based on the feedback received throughout the three previous iterations and from feedback from Plexlines. Three common themes emerged in the feedback from the earlier iterations: (1) the need to visualize three key modalities of communicative behavior (gaze, gesture, and vocalization), (2) the ability to compare children's behaviors, and (3) the need for interaction-based features, such as filtering or customizing. In the following section, we present three visual elements that address the first two themes and two interactive components of EnGaze that address the third theme.

The Three Visual Elements

Multimodal Element: The first element is a multimodal visualization of behavior that included gaze, gesture, and vocalization. Our first iteration included only the gaze of the child and examiner since gaze appeared as the main common modality across the joint attention literature [18, 37]. In a joint project meeting, two child specialists emphasized that concurrent behaviors across modalities (e.g., vocalization coupled with gesture) are as important as gaze. We then added gesture and vocalization modalities to EnGaze to accommodate concurrent behaviors in the visualization.

Dyadic Element: Joint attention depends on both the child's and the clinician's behavior. Yet, current assessment instruments record the actions of the child, often summarized by a numeric value, without noting the explicit actions of the examiner [24, 36, 33]. While such assessments are the norm, they do not hold explanatory power and do not capture the temporal complexities of the interaction between the child and the clinician. A trained third party looking at an ADOS assessment could imagine possible behaviors in the observation session, but without scanning a video of a session from beginning to end, would not see the nuances in the behavior. We designed EnGaze to include both the child and the examiner behaviors to capture the dynamics of joint attention, e.g., see the gaze interplay between the child and the examiner in the ball and book phases in the bottom visualization in Figure 6. EnGaze can further be used for repeat observation by the examiner or by other clinicians.

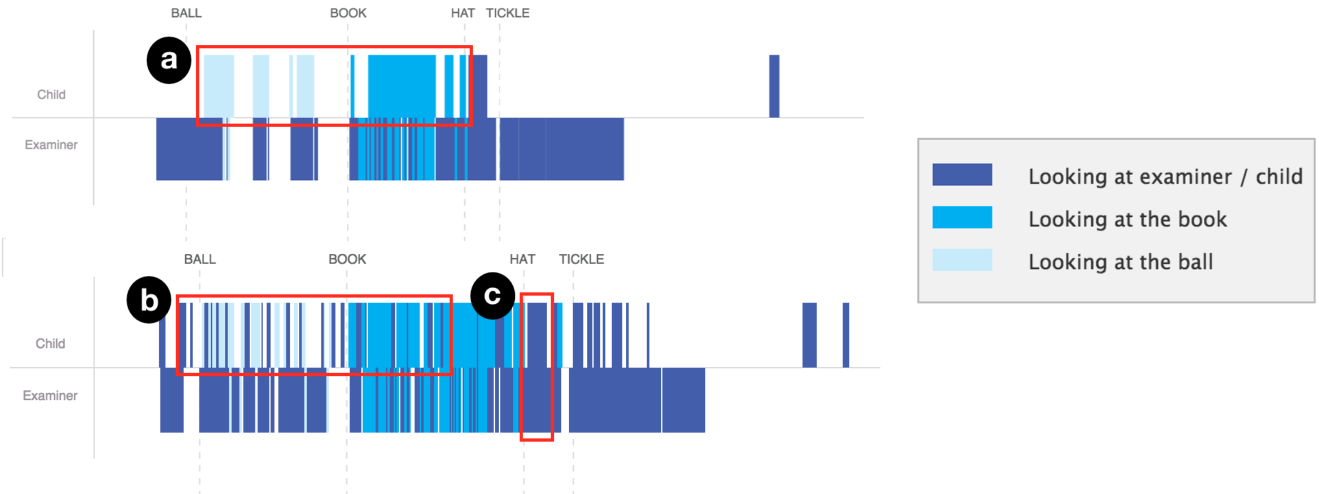

Figure 6. Two EnGaze visualizations of gaze between a child and an examiner. (a) a child fixated only on the object creates large rectangular blocks of the same color. (b) back-and-forth gaze switching between the examiner/child and the object creates a pattern that resembles a barcode, which is often a positive sign of joint attention. (c) a connected vertical rectangular block forms when the child and examiner make eye contact.

Multisession Element: The third visual element— the ability to see and compare multiple sessions—is closely related to research in human behavior analysis that extensively employ single-subject research methodologies [20, 21], in which the subjects often serve as their own control. Researchers and clinicians commented on how visualizing progress over the entire treatment period helps to measure the effectiveness of an intervention. The current Web tool is vertically compact and displays approximately four visualizations of different children on a typical desktop screen. The same layout could similarly display visualizations of multiple sessions for the same child in a sequence on one page. With this progressive view of a child, each visualization can act as a graphical record of a session, and multiple sessions can be interpreted at a glance. Clinicians, and as study participants pointed out, possibly parents, could see longitudinal assessments over time.

The Two Interactive Components

In addition to the visual elements mentioned above, we integrated two interactive components—a video component and an authoring panel—into the Web tool based on the positive feedback from earlier studies. Both features allow users to explore the sessions in more depth and according to their own definitions of joint attention as we explain below.

Video Component: With the video feature in EnGaze, the visualization acts not only as a visual representation of a child-examiner session, but also as a video seek bar. When a user clicks on a specific point of interest on a visualization (e.g., when the child responds to an examiner bid), the video syncs to show that moment (Figure 1). This feature provides additional contextual information when needed.

Authoring Component: The authoring panel allows users to select a subset of all possible behaviors for a more detailed analysis. Users can highlight moments of interest by selecting different authoring options and clicking on the colors on the interactive legend (Figure 1a and b).

Each mode has a different set of authoring options. The *Gaze* mode panel options include: "Child and examiner looking at object at the same time", "Child and examiner looking at each other" (*eye contact*), and "Child looking back and forth between examiner and the object" (*gaze alternation*) (Figure 1a). Joint attention is not instantaneous. A child can look at the examiner 1.5 seconds after gazing at an object. To capture such cases of joint attention, the *gaze alternation* authoring panel option has a custom delay tolerance that the user can set using the slider shown in Figure 1c. For example,

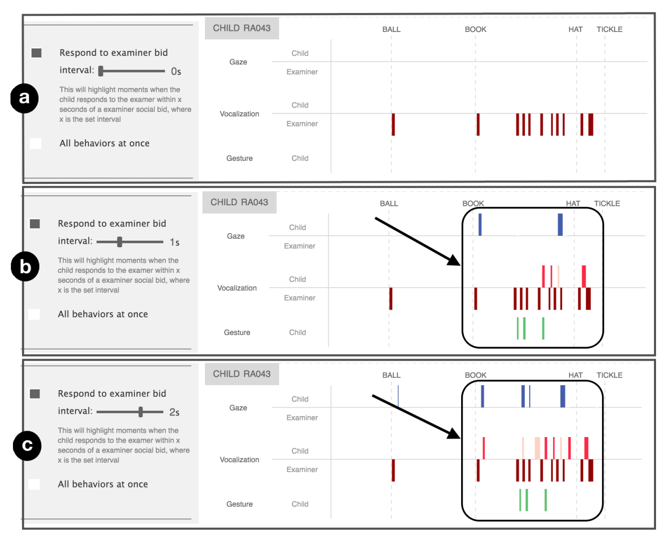

Figure 7. Three EnGaze visualizations of the same child in *All Behaviors* mode. The delay for "Respond to examiner bid" is set to 0, 1, and 2 seconds for (a), (b), and (c), respectively, to show the changes in the visualization with the varying delay options. When the delay is set to X seconds, the visualization will highlight all child actions that happen within X seconds of the examiner's bid. (a) The delay is set at 0 seconds. The only visible actions are the examiner bids (in dark red). (b) The delay is set at 1 second, and the child's gaze, gesture, and vocalizations appear in the visualization (see box). (c) The delay is set at 2 seconds, and more child behaviors are visible in the visualization (see box).

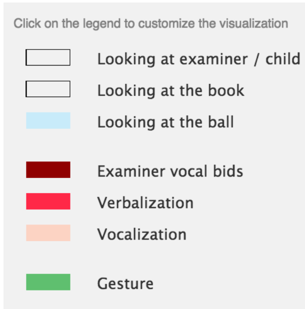

Figure 8. The legend for the *All Behaviors* mode doubles as a part of the authoring panel for EnGaze. "Looking at the examiner / child" and "Looking at the book" are currently toggled off, as indicated by the blank boxes next to the description. Clicking on each of the rectangles toggles the corresponding feature, making it visible again in the visualization .

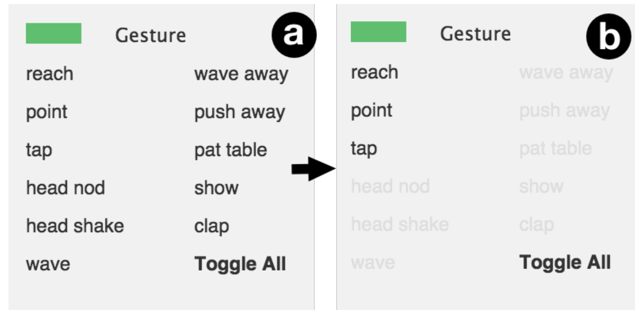

Figure 9. The legend for the gestures in the *All Behaviors* mode. All gestures are visible in (a). In (b), only "reach", "point", and "tap" are toggled on, and only these three gestures would appear in the resulting visualization.

if a user sets the delay time to 1.5 seconds, every instance where the child looks at the examiner within 1.5 seconds of looking at the object and then looks back at the object within 1.5 seconds will be highlighted (an "object-examiner-object" pattern). Moments exhibiting an "examiner-object-examiner" pattern are highlighted as well.

The *All Behaviors* mode has two options. The first option is unique to the *All Behaviors* mode; it allows users to highlight durations when the child exhibits all behaviors at once (gaze, gesture, and vocalization). This option was implemented to address the clinicians' requests to explore concurrent multimodal behaviors of a child. The second option allows users to highlight all the child's responses to examiner bids (*response to bid*) (Figure 7). This option has a custom delay slider, similar to the one for the *gaze alternation* option, and is available in the *Vocal* mode as well, where only the vocal responses are shown in the visualization.

The interactive legend allows users to display or hide specific behaviors by clicking on the corresponding rectangles (Figure 8). A colored rectangle indicates that the relevant behavior is currently visualized. A transparent rectangle indicates that the behavior is hidden. If users to hide the "Looking at the examiner / child" and "Looking at the book" entries, they can click on the top two rectangles on the legend to remove them from the *Gaze* view. Figure 8 shows the resulting legend. The moments when the examiner or the child looked at the ball will remain in the visualization, as well as other vocal and gesture behaviors from the child.

Clicking on the green *Gesture* rectangle reveals a list of annotated gestures (e.g., reach, point, and tap). We built a separate gesture list as there are eleven different types of annotated gestures (Figure 9a) in the data set. All of the available gestures are represented with a green rectangle in the visualization. The list of gestures allows users to show or hide specific gestures. If a user is only interested in "reach", "point", and "tap" gestures, the user can hide all other gestures by clicking on the unwanted gestures one by one. The labels of the hidden gestures appear in a lighter font color (Figure 9b). Selecting different options using the authoring panel and the legend allowed users to customize the visualization based on their interpretation of joint attention and their research or clinical needs.

EnGaze was implemented as a web application built using HTML, JavaScript, and CSS. The browser-based environment allows for easy user access and does not require additional software. Note that such a browser model would require an authentication interface to abide by the Health Insurance Portability and Accountability Act (HIPAA) to move beyond this lab study.

EVALUATION

To evaluate the usability and the efficacy of EnGaze, we conducted a user study with five participants, clinicians and researchers who had a background in child development or behavioral analysis and prior experience working with children with ASD. We recruited participants by placing flyers in academic buildings at the University of Illinois in Urbana-Champaign and local libraries. We also used snowball sampling to recruit child development researchers and special education professionals with clinical experience. Each participant received a $12 Amazon gift card at the end of the session. Of our five participants, two were male and three were female. Three participants had backgrounds in behavioral psychology and two in developmental psychology. Each had a minimum of four years of experience working directly with children with developmental delays. Four of the participants spent approximately one hour participating in the study. One participant spent one hour on the day of the study, and asked to return and experiment with the interface for another hour the following week.

The user study consisted of a preliminary interview, an introductory video, an exploration phase, a semi-structured open-ended interview, and a final survey. In the preliminary interview, we asked the participants about their expertise in ASD and their experience working with children with developmental delays. We then asked them to define joint attention in their own words. We specifically asked them for their interpretation of joint attention before we introduced EnGaze, as its features might bias their responses. Then, we asked if there were any specific modalities, such as gaze, gesture, or vocalization, that they considered important in joint attention. All five participants mentioned a combination of one to three modalities out of the three modalities mentioned above.

Participants then watched a five minute introductory video that briefly described the Rapid ABC protocol and the essential features of EnGaze (e.g., switching between the three available modes, highlighting joint attention features, and using the legend as a filter). Following the video, participants spent approximately thirty minutes exploring the Web tool. In this exploration phase, we provided the following tasks to guide the users and to test the various features offered in EnGaze:

- Find a child who is especially talkative.
- Find a child who does not look at the examiner often.
- Find a child who reaches a lot.
- Find a child who looks back and forth often.
- Find a child who does not make a lot of eye contact with the examiner.
- Find a child who responds to a vocal bid with both gesture and vocalization.

These tasks were provided as a starting point to prompt the participants to explore different features in the tool. However, the main goal of this exploration stage was to go beyond the given tasks and observe how the participants explored their own definitions of joint attention through the tool, and how they used it for assessment. We asked participants to talk aloud as they used EnGaze as we observed and took notes on their use patterns and any comments they shared throughout the study. After the exploration phase, we conducted a semi-structured open-ended interview. We asked if and how they would incorporate EnGaze into their work flow, and what features relating to joint attention they would add to EnGaze. We also asked if there were any children for whom they would recommend a follow-up based on the visualization. The study session ended with a written survey that included nine five-point Likert scale questions to evaluate the participants' expertise and the interpretability, usability, and satisfiability of EnGaze, followed by four open-ended questions about the strengths and the weaknesses of EnGaze.

Overall, EnGaze was well received by the study participants. On a five-point Likert scale ranging from strongly disagree (1) to strongly agree (5), the study participants were satisfied with the EnGaze visualization ($\mu = 4.6$, $\sigma^2 = 0.54$), and found the visualization to be interpretable ($\mu = 4.2$, $\sigma^2 = 0.45$). The participants were also satisfied with the EnGaze features ($\mu = 4.2$, $\sigma^2 = 0.44$) and found them easy to use ($\mu = 4.4$, $\sigma^2 = 0.54$).

Responses

During the exploration stage, the study participants could complete any of the given tasks or explore EnGaze freely. For a given task, each participant had a unique style of browsing through the interface, utilizing certain features more than others. However, there were two authoring options that received the most positive feedback from the participants: the *response to bid* and the *eye contact* options.

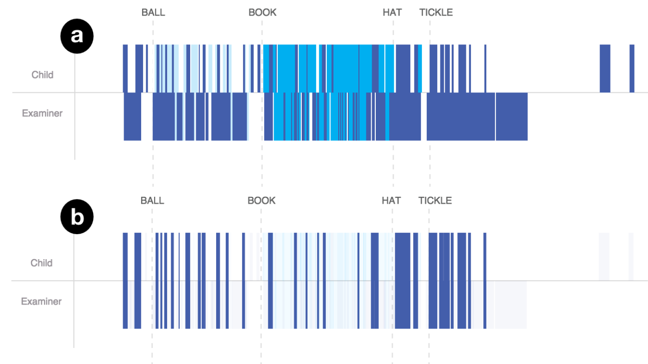

Figure 10. Two EnGaze visualizations. (a) the default EnGaze visualization. (b) the visualization after the user has selected to highlight the eye contact between the child and the examiner in the interactive tool. Only the moments when the child and examiner make eye contact are highlighted in the visualization.

Popular Options when Searching for Joint Attention

Participants were very interested in the *response to bid* option in the authoring panel. This was cited by four of the five study participants as one of the most useful authoring options. One participant specifically stated that it was their favorite feature when asked about the usefulness of the features offered in the Web tool (P4). When asked how they would integrate EnGaze into their workflow when examining a child, another participant stated *"I would [first] look at all [behaviors], and see if the child is responding to examiner bids"* (P1). The participants especially expressed concerns when children showed no response to multiple examiner bids, and often checked the video to see how the child behaved immediately after the examiner bids. The *eye contact* feature was another popular feature that all five participants frequently used during the study (Figure 10). Children with ASD often have a hard time making and maintaining eye contact with others; this feature allowed participants to easily detect eye contact patterns [29]. While working with the *eye contact* feature, a participant stated that this feature made detecting and analyzing eye contact very easy: *"seeing when the child looks, when the examiner looks, you can see the frequency and the amount they see from the visualization."* (P3). This feature also brought certain children to the participants' attention. One participant stated, *"children who did not have much gaze [stood out]"* (P1). The lack of eye contact prompted the participant to take a closer look at the other communicative behaviors of the child.

We also received positive feedback regarding the dyadic and multimodal representation of the sessions. Three participants complimented the equal emphasis on the child and the examiner in EnGaze. As previously stated, joint attention is not a skill that can be observed in one individual. One participant commented, *"I found it fascinating to visualize the variables examined and their relationship!"* (P2). Another participant commented on the multimodal aspect of the visualization. As a speech pathologist working with children with disabilities, the participant primarily dealt with speech and said, *"I appreciate [that] you are looking at multimodal. We focus so much on speech that once kids become verbal we* *don't study gaze, gesture as much, but I think it is influential"* (P5).

Meeting Clinicians' and and Researchers' Needs

A participant described his process of coding an intervention as follows: *"[We] film interactions between the child and the parent, [and then] observers code the video tapes. [After defining] joint attention (not eye-gaze specifically) manually going through video tapes, [we] count the number of specific interactions in the video"* (P3). In this participant's current workflow, if the examiner wants to revisit a hand-coded point of interest, they must watch the video again to identify the exact moments of interaction. P5 similarly explained, *"whenever you want to go back and see this episode, you need to go back to the video,"* and added that *"it would be nice to see turn-taking and go right to the section and watch it."* The participants appreciated the ease with which one could immediately jump to points of interest in the video while referring to the visualization and using it as a seek bar. P2 specifically mentioned that the video in this video-visualization coupling is critical for distinguishing between what he considered joint attention and *requesting*.

While all the participants found EnGaze interpretable and easy to use, they offered suggestions for improvements. P3 and P4 specifically suggested we include zooming features for the visualization so they could look at interactions more closely. The same two participants suggested an aggregated view of a child's multimodal behaviors rather than the separate views as they appear in the current *All Behavior* view. Of note is that during our formative EnGaze design sessions, we experimented with aggregate views but did not find a design solution that aggregated multimodal behaviors clearly.

Tradition in the clinicians' and researchers' respective fields played a role in their expectations. An audience member at a medical school presentation of EnGaze commented on the lack of quantitative data in the visualization. Two of our participants, P3 and P5 also desired a quantitative summary of the session. P3 commented, *"This is good that you can see where it happens, but can you tally the frequency of data? I care about both the frequency and duration, we care about both, frequency will matter (related to the examiner frequency)—this visualization allows you to see this."* Such quantitative summaries are widely used in screening methods such as ADOS, at the end of which a count and levels of engagement are provided. While it would be trivial to add quantitative measures to the EnGaze visualization, we had originally omitted explicit numbers from our tool intentionally so that our participants would focus on the interactions themselves as represented in the visualization. EnGaze was designed to provide not only summative power but also explanatory power. Instead of explaining behavior from a series of numbers, our goal was to provide additional context through video and interaction dynamics. We believe there is value in combining established and validated existing methods and norms from our intended communities, and in future work, plan to incorporate quantitative measures of frequency into EnGaze.

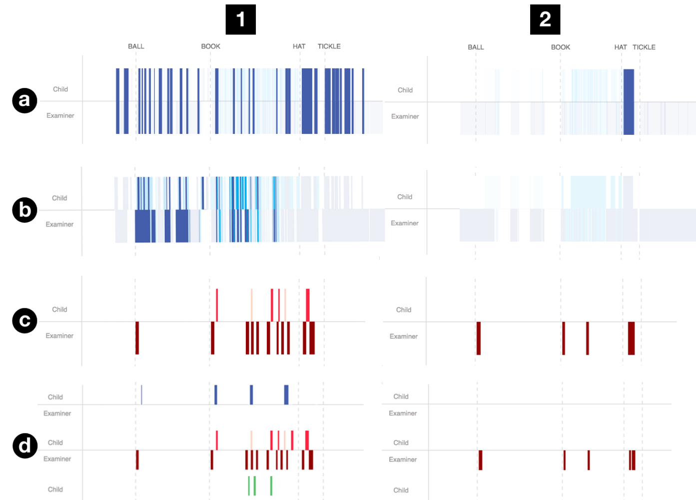

Figure 11. Two different sets of EnGaze visualizations from two children. Child (1) appears to be actively engaged with the examiner while child (2) is disengaged. The following visualizations (a)-(d) are the result of the features mentioned as most useful by our participants. (a) the eye contact feature: highlights moments of eye contact between the child and the examiner. (b) the gaze alternation feature: highlights moments of gaze alternation, when the child has looked back-and-forth between the examiner and the object. (c) the response to bid feature (in the *Vocal* mode): displays the child responding to the examiner's bids vocally. (d) the response to bid feature (in the *All Behaviors* mode): shows all behaviors of the child following a examiner's bid.

ILLUSTRATIONS OF ENGAZE USAGE IN PRACTICE

During the interviews, we asked the participants if and how they would incorporate EnGaze into their field of work. Based on their responses to this question and the main themes from the open-coding results of interviews, we present two main use cases of EnGaze for clinicians and researchers.

Identifying Cases for Further Assessment

EnGaze's compact visualization and color-coded behaviors helped participants identify children in probable need of a follow-up ASD screening session. All five participants independently suggested that the same two children receive further assessment[2]. Throughout the exploration stage, the participants checked for behaviors that related to their definition of joint attention and other behaviors related to autism by using the authoring panel; they often found a child of interest. For example, after browsing for moments where a child is looking at the examiner, P5 exclaimed *"Ah, just looking at the frequency, this kid jumps out!"* When filtering for "Looking at the examiner / child," it is easy to observe when a child does not look at the examiner as often as other children. The visualization often called attention to particular children due to the lack of a certain behavior or combination of behaviors. While looking at different behaviors in the *All Behaviors* view, P3 commented *"this kid is talkative, but does not make much eye contact. I would like to see the video in detail. A follow up would be interesting."*

[2]Families participating in the Rapid ABC sessions were not expected to contact us if they later received an ASD diagnosis.

EnGaze was also used to find a particular session's moments of interest. P5 stated that she would be interested in seeing *"those places where there are lots of engagement or go to the places where there is less engagement if that is available."* The participants also compared the engagement level at different phases of the protocol for a child: *"I would follow up with [this child] especially in the book activity, there is something going on with [the] book [activity]. [There is] vocal[ization] in the ball activity but much less in book and tickling."* (P3). This ability to highlight moments of interest within a session and compare across sessions was highly desirable and beneficial for our participants' pre-screening purposes. Additionally, participants valued the ability to narrow the number and the time frame of videos to explicitly specify what they wished to watch. They believed this could save them time and energy.

EnGaze as a Visual Record

The ability to compare sessions suggested that EnGaze could act as a visual record of children for clinicians and parents. More specifically, P1 suggested that EnGaze could be used to *"'show [the session] to parents, keep track of progress, and use for future instruction."* If EnGaze was used as a visual record for interventions across intervals, the visualization of the first session could act as a solid baseline against which the clinicians can compare the following sessions to measure the effectiveness of an intervention for that child. The norm, or a common response for a protocol, could also be visualized through EnGaze if the data for multiple children with typical development were available. P2 remarked, *"norms would help enormously! [EnGaze could be used to] follow up and compare with past data."*

We also found that for EnGaze to be an effective visual record, more contextual information was necessary. The contextual information that the participants desired was the exact instruction given for the vocal bids. A participant indicated that if the coding was automated, she would use EnGaze in a natural class setting where she had to attend to several students and could not observe a child's behavior fully (P1). She further stated, *"If I were to use it with students, I would like to see how they react to my instructions."* This focus on seeing the children's response to instructions matches the previously mentioned interest in using the *response to bid* feature. Displaying the exact vocal bid is necessary for EnGaze to show the relationship between the instruction and the child's reaction. Functionally, a *tool tip* could reveal the exact vocal bid when a participant hovers over a behavior block in EnGaze.

LIMITATIONS

Our approach to visualizing behavior on EnGaze relies heavily on the availability and the accuracy of the behavior annotations. The details are also affected by the rigidity of the handcrafted annotation schedule. Until we can reliably annotate data automatically, we are reliant on time-intensive hand-coded annotations for our visualization. Although this is certainly a valid concern, the goal of this paper is to convey a proof of concept for EnGaze while we and some of our collaborators are pursuing the automation of the Rapid ABC annotation [14, 35, 42]. We envision that five-minute protocols supported by similar technology will become standard procedures in doctors' offices in the years to come.

We also realize that our current annotation is imperfect and not comprehensive as it only includes gaze, gesture, and vocalization behaviors. Smiles and affect are examples of additional annotations that could be included in the future. Annotations for other red flags of autism, such as echolalia, unusual prosody, and stereotypical, repetitive behaviors [30, 39] could also be explored.

While we received positive feedback from the participants, the study was conducted with only five experts in ASD. In addition, we do not have parental reports of which children were later diagnosed with ASD for validation. An extended study with a larger number of participants comparing the performance of EnGaze to that of traditional tools and with additional data of children diagnosed with ASD will help validate our current findings.

CONCLUSION

In this paper, we presented EnGaze, a visualization Web tool for child-clinician communicative behavior that highlights commonly discussed features of joint attention. We described an iterative design process of an interactive visualization for clinical and research use and a user study with clinicians and researchers to identify how they could use EnGaze to explore different features of joint attention and incorporate them into their practice. While the clinicians and researchers focused on different behaviors based on their past experience, they were able to capture their interpretations of joint attention using the features provided in EnGaze. Study participants commented on how EnGaze could supplement existing evaluation processes by providing context for the numerical values in traditional assessments. When using the visualization, they expressed concern for children who exhibited a lack of joint attention, and independently suggested the same two children for further assessment. Participants also suggested the use of EnGaze visualizations as a visual record for tracking the communicative behavior of children over time. Overall, EnGaze shows potential for providing clinicians and researchers with snapshots of communicative behavior development, aiding their analysis of trends and benchmarks in children, and establishing a common graphical language through which to communicate with parents.

ACKNOWLEDGEMENTS

We would like to thank the National Science Foundation (Award Number CCF-1029679) for supporting this work, the parents and children for participating in the Rapid ABC studies, our study participants, and the entire NSF Expeditions team for providing data, time and a supportive community to explore behavior imaging.

REFERENCES

1. Fitbit. `https://www.fitbit.com/`. Accessed: 11-30-2015.

2. Nike+Fuel. `http://www.nike.com/us/en_us/c/nikeplus-fuel`. Accessed: 11-30-2015.

3. Simon Baron-Cohen, Jane Allen, and Christopher Gillberg. 1992. Can autism be detected at 18 months? The needle, the haystack, and the CHAT. *The British Journal of Psychiatry* 161, 6 (1992), 839–843.

4. Rachael Bedford, Mayada Elsabbagh, Teodora Gliga, Andrew Pickles, Atsushi Senju, Tony Charman, Mark H Johnson, and others. 2012. Precursors to social and communication difficulties in infants at-risk for autism: gaze following and attentional engagement. *Journal of autism and developmental disorders* 42, 10 (2012), 2208–2218.

5. Tony Bergstrom and Karrie Karahalios. 2007. Seeing more: visualizing audio cues. In *Human-Computer Interaction–INTERACT 2007*. Springer, 29–42.

6. T Blascheck, K Kurzhals, M Raschke, M Burch, D Weiskopf, and T Ertl. 2014. State-of-the-art of visualization for eye tracking data. In *Proceedings of EuroVis*, Vol. 2014.

7. Daniel Bone, Chi-Chun Lee, Alexandros Potamianos, and Shrikanth Narayanan. 2014. An Investigation of Vocal Arousal Dynamics in Child-Psychologist Interactions using Synchrony Measures and a Conversation-based Model. In *Fifteenth Annual Conference of the International Speech Communication Association*.

8. Signe Marie Boucher. 2007. *Joint attention, imitation, and repetitive behaviors as predictors of autism and expressive language ability in early childhood.* ProQuest.

9. Yvonne Bruinsma, Robert L Koegel, and Lynn Kern Koegel. 2004. Joint attention and children with autism: A review of the literature. *Mental retardation and developmental disabilities research reviews* 10, 3 (2004), 169–175.

10. Malinda Carpenter, Katherine Nagell, Michael Tomasello, George Butterworth, and Chris Moore. 1998. Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the society for research in child development* (1998), i–174.

11. Tony Charman. 2003. Why is joint attention a pivotal skill in autism? *Philosophical Transactions of the Royal Society B: Biological Sciences* 358, 1430 (2003), 315–324.

12. Community of Practice for Autism. 2007. Joint Attention and Social Referencing. (2007). `http://www.infantva.org/documents/CoPA-Nov-JointAttentionSocialRefer.pdf` Valley Region CoPA (Community of Practice for Autism) Meeting.

13. Harriet J Fell, Joel MacAuslan, Linda J Ferrier, and Karen Chenausky. 1999. Automatic babble recognition for early detection of speech related disorders. *Behaviour & Information Technology* 18, 1 (1999), 56–63.

14. Nuno Goncalves, José L Rodrigues, Sandra Costa, and Filomena Soares. 2012. Automatic detection of stereotyped hand flapping movements: two different approaches. In *RO-MAN, 2012 IEEE*. IEEE, 392–397.

15. Joshua Hailpern, Karrie Karahalios, Laura DeThorne, and Jim Halle. 2010. Vocsyl: Visualizing syllable production for children with ASD and speech delays. In *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility*. ACM, 297–298.

16. Yi Han, Agata Rozga, Nevena Dimitrova, Gregory D Abowd, and John Stasko. 2015. Visual Analysis of Proximal Temporal Relationships of Social and Communicative Behaviors. In *Computer Graphics Forum*, Vol. 34. Wiley Online Library, 51–60.

17. Jordan Hashemi, Thiago Vallin Spina, Mariano Tepper, Amy Esler, Vassilios Morellas, Nikolaos Papanikolopoulos, and Guillermo Sapiro. 2012. A computer vision approach for the assessment of autism-related behavioral markers. In *Development and Learning and Epigenetic Robotics (ICDL), 2012 IEEE International Conference on*. IEEE, 1–7.

18. Emily A Jones and Edward G Carr. 2004. Joint Attention in Children With Autism: Theory and Intervention. *Focus on autism and other developmental disabilities* 19, 1 (2004), 13–26.

19. Connie Kasari, Marian Sigman, Peter Mundy, and Nurit Yirmiya. 1990. Affective sharing in the context of joint attention interactions of normal, autistic, and mentally retarded children. *Journal of autism and developmental disorders* 20, 1 (1990), 87–100.

20. Alan E Kazdin. 2003. Research design in clinical psychology. (2003).

21. Alan E Kazdin. 2011. *Single-case research designs: Methods for clinical and applied settings* . Oxford University Press.

22. Julie A Kientz, Sebastian Boring, Gregory D Abowd, and Gillian R Hayes. 2005. Abaris: Evaluating automated capture applied to structured autism interventions. In *UbiComp 2005: Ubiquitous Computing*. Springer, 323–339.

23. John Lee, Sanny Lin, and Karrie Karahalios. 2013. Visualizing patterns of social and communicative behavior in children using plexlines. In *Proceedings of the 2013 Workshop on Visual Analytics in Healthcare (Nov. 2013)*. 28–32.

24. Catherine Lord, Susan Risi, Linda Lambrecht, Edwin H Cook Jr, Bennett L Leventhal, Pamela C DiLavore, Andrew Pickles, and Michael Rutter. 2000. The Autism Diagnostic Observation Schedule-Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of autism and developmental disorders* 30, 3 (2000), 205–223.

25. Katherine A Loveland and Susan H Landry. 1986. Joint attention and language in autism and developmental language delay. *Journal of autism and developmental disorders* 16, 3 (1986), 335–349.

26. Megan Monroe, Rongjian Lan, Hanseung Lee, Catherine Plaisant, and Ben Shneiderman. 2013. Temporal event sequence simplification. *Visualization and Computer Graphics, IEEE Transactions on* 19, 12 (2013), 2227–2236.

27. Chris Moore and Phil Dunham. 2014. *Joint attention: Its origins and role in development*. Psychology Press.

28. Peter Mundy, Marian Sigman, and Connie Kasari. 1990. A longitudinal study of joint attention and language development in autistic children. *Journal of Autism and developmental Disorders* 20, 1 (1990), 115–128.

29. Peter Mundy, Marian Sigman, and Connie Kasari. 1994. Joint attention, developmental level, and symptom presentation in autism. *Development and Psychopathology* 6, 03 (1994), 389–401.

30. Peter Mundy, Marian Sigman, Judy Ungerer, and Tracy Sherman. 1986. Defining the social deficits of autism: The contribution of non-verbal communication measures. *Journal of child psychology and psychiatry* 27, 5 (1986), 657–669.

31. Eva Murillo and Mercedes Belinchón. 2012. Gestural-vocal coordination: Longitudinal changes and predictive value on early lexical development. *Gesture* 12, 1 (2012), 16–39.

32. Julie Osterling and Geraldine Dawson. 1994. Early recognition of children with autism: A study of first birthday home videotapes. *Journal of autism and developmental disorders* 24, 3 (1994), 247–257.

33. Opal Ousley, Rosa Arriaga, Gregory Abowd, and Michael Morrier. Jan 2012. Rapid Assessment of social-communicative abilities in infants at risk of autism. *Technical Report CBI-100, Center for Behavior Imaging, Georgia Tech* (Jan 2012).

34. Catherine Plaisant, Brett Milash, Anne Rose, Seth Widoff, and Ben Shneiderman. 1996. LifeLines: visualizing personal histories. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 221–227.

35. James M Rehg. 2011. Behavior Imaging: Using Computer Vision to Study Autism.. In *MVA*. 14–21.

36. Michael Rutter, Ann Le Couteur, and Catherine Lord. 2003. Autism diagnostic interview-revised. *Los Angeles, CA: Western Psychological Services* (2003).

37. Bridget A Taylor and Hannah Hoch. 2008. Teaching children with autism to respond to and initiate bids for joint attention. *Journal of Applied Behavior Analysis* 41, 3 (2008), 377–391.

38. Karen Toth, Jeffrey Munson, Andrew N Meltzoff, and Geraldine Dawson. 2006. Early predictors of communication development in young children with autism spectrum disorder: Joint attention, imitation, and toy play. *Journal of autism and developmental disorders* 36, 8 (2006), 993–1005.

39. Amy M Wetherby, Juliann Woods, Lori Allen, Julie Cleary, Holly Dickinson, and Catherine Lord. 2004. Early indicators of autism spectrum disorders in the second year of life. *Journal of autism and developmental disorders* 34, 5 (2004), 473–493.

40. N.D. Wiseman. 2009. *The First Year: Autism Spectrum Disorders: An Essential Guide for the Newly Diagnosed Child*. Da Capo Life Long. `https://books.google.com/books?id=Z8hAF5CkGugC`

41. Krist Wongsuphasawat, John Alexis Guerra Gómez, Catherine Plaisant, Taowei David Wang, Meirav Taieb-Maimon, and Ben Shneiderman. 2011. LifeFlow: visualizing an overview of event sequences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1747–1756.

42. Zhefan Ye, Yin Li, Alireza Fathi, Yi Han, Agata Rozga, Gregory D Abowd, and James M Rehg. 2012. Detecting eye contact using wearable eye-tracking glasses. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM, 699–704.

43. Chen Yu and Linda B Smith. 2013. Joint attention without gaze following: human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PloS one* 8, 11 (2013), e79659.