Title
Methodological Considerations For Investigating the Microdynamics of Social Interaction Development

Permalink
https://escholarship.org/uc/item/5fd8q7mc

Journal
IEEE Transactions on Autonomous Mental Development, 5(3)

ISSN
1943-0604 1943-0612

Authors
de Barbaro, K.
Johnson, C. M
Forster, D.
et al.

Publication Date
2013-09-01

DOI
10.1109/TAMD.2013.2276611

Peer reviewed

Methodological Considerations For Investigating the Microdynamics of Social Interaction Development

Kaya de Barbaro, Christine M. Johnson, Deborah Forster, and Gedeon O. Deák

***Abstract*—Infants are biologically prepared to learn complex behaviors by interacting in dynamic, responsive social environments. Although the importance of interactive social experiences has long been recognized, current methods for studying complex multimodal interactions are lagging. This paper outlines a systems approach for characterizing fine-grained temporal dynamics of developing social interaction. We provide best practices for capturing, coding, and analyzing interaction activity on multiple –temporal scales, from fractions of seconds (e.g., gaze shifts), to minutes (e.g., coordinated play episodes), to weeks or months (e.g., developmental change).**

***Index Terms*—Cognitive ethnography, development, infancy, methodology, parenting, sequential analysis, social interaction, systems theory.**

I. Introduction

IT IS now a widely accepted view in developmental science that infants are biologically prepared to learn complex behaviors via their experiences interacting in a dynamic world responsive to their activity. Moment-to-moment changes in infants' activity are codetermined by a complex multidimensional array of input streams including feedback from caregivers, the infant's own sensorimotor behaviors, and various environmental features. Additionally, through their own actions, infants shape the unfolding interaction, both by changing their access to sensory stimulation and by eliciting activity from the caregiver and/or object. Although these claims are not new (see, e.g., [71] and [87]), current efforts in studying the complex processes by which new behavior emerges are lagging (e.g., [3], [28], and [72]). Our goal in this paper is to describe best practices for capturing the natural patterns of social behavior, bridging activity from the fine details of the learning process to the bigger picture of long-scale developmental change.

Traditional experimental methods require "controlling" all but one manipulated variable. When used in developmental psychology, such methods can radically restrict infants' range of behaviors, along with the complexity and variability of their environment [33], [83]. When this range of variability and complexity is so reduced, sampled behaviors will under-specify the everyday processes that embody infants' developing social skills. Additionally, in a complex system like human development, the activity of a single variable often cannot predict the activity of the system as a whole [33], [83].

Since the 1970s a number of studies have focused on the dynamics of naturalistic infant–parent interaction. In these observational studies, little or no manipulation of the setting is imposed by the researchers. Nonetheless, the majority of these studies have been limited in their ability to systematically quantify and model the complex multimodal cycles of activity found in naturalistic behavior. A common technique in this work has been to characterize dynamics across only one or two dimensions of activity, such as the relations between infant gaze (e.g., to or away from the parent), and the parent's smiling (e.g., "on" or "off") [12], [52], [57]. Such studies begin with richly informative interactions, and often code those interactions with moderate temporal precision. However, by reducing the interactions to two running binary streams, such studies mask the high-dimensionality of actual social interactions, and may miss important effects.

For example, when a mother manipulates a toy, her infant might respond by looking towards or away from it, but he might also change his posture (e.g., lean in with interest or pull back with apprehension), reach out to grasp the toy, or reveal emotional changes via vocalizations or facial expressions. By coding only one dimension of activity, such as the infant's gaze, in relation to one other variable, such as the presence or absence of a toy, many important features of the infant's experience, including the caregiver's role in the interaction, are obscured. The sensorimotor and social dimensions that are eliminated are likely to play a critical role in sociocognitive development. These additional dimensions might also help us to disambiguate otherwise puzzling or chaotic results. For example, an infant can look away from an adult due to overstimulation, boredom, or stress, so a more in-depth assessment of physiological, sensorimotor, and social parameters can differentiate between these possibilities.

Possibly for that reason, some early champions of naturalistic observation pointed out that coding schemes that tracked the overall activity of the infant or dyad holistically, considering many dimensions of activity, could better characterize the nature of the interaction [18], [48]. However, such interaction-level codes—characterizing, for example, dyadic states such as "joint attention" or "independent play"—sacrifice the details of the timing and nature of component actions [85]. As a result, studies that provide information only about individual or interaction level activity are unable to address questions of the emergence of developmental shifts in such activity, such as the shift from independent to joint attention.

In this paper, we summarize recent theoretical and methodological advances in capturing and characterizing the social de-

Manuscript received February 23, 2013; revised June 22, 2013; accepted July 19, 2013. Date of publication August 22, 2013; date of current version September 07, 2013.

The authors are with the Department of Cognitive Science, University of California, San Diego, La Jolla, CA 92093-0515 USA (e-mail: kaya@cogsci.ucsd.edu; johnson@cogsci.ucsd.edu; forstermobu@gmail.com; deak@cogsci.ucsd.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAMD.2013.2276611

velopment in all of its complexity. The principles of our account are inspired by dynamical and distributed systems approaches to mind and behavior in humans [20], [34], [49], [50], [58], [68], [83], [84] and nonhuman species [35], [36], [51], [53], [54], [55], [56], [78], [79]. By treating development as system change, these approaches focus attention on how elements of the interaction configure, and reconfigure over developmental time. The methods that arise to address these changes track the emergence of new activity at the system level, as well as elements of the processes from which they emerge. By characterizing how elements of system participate in system-level transitions we gain access to the processes of development.

II. A Systems Approach

One fundamental principle of a systems approach is that interaction is the unit of analysis. That is, even when we track and quantify individual actions (e.g., reaches, looks), we analyze these relative to other participants' actions, shared objects, or the context of the shared setting and activity (e.g., reach or look to partner). This applies even when the system we are studying is intraindividual, such as the work by [82] on changes in the relationships between the joints, muscles, and articulatory control of shoulder, arm, and hand, in the ontogeny of reaching. However, in studies of social behaviors, it is necessary to adopt an interindividual systems approach, wherein the relationships between the dimensions of activity of the agents become the focus of research. To understand development, we must track the activity of individuals as they interact with one another and with their shared environment.

Another tenet of the systems approach is that cognitive activity spans multiple spatio–temporal scales. Unfolding at the scale of fractions of seconds are sensorimotor activities by which participants access and act upon their shared environments. On timescales of seconds and minutes, these sensorimotor activities organize into recognizable dyadic activities. Thus, a look or grasp is positioned within ongoing, coregulated activity, where it may repeat, or adjust following feedback, or organize with other events. For example, the significance of breaking mutual gaze with a partner is very different following a pointing gesture versus following a request. On historical/developmental timescales, the changing dynamics of activity can consolidate into new ways of interacting, establishing long-termpractices that alter the demands on behavior and cognition. Capturing activity at each of these scales provides additional information necessary for understanding developmental shifts in the unfolding sequence of activity.

It is important to note that when studying a complex social system, properties of the activity at each of these scales often do not map directly to one another. For example, the properties of an interaction—such as the presence of a positive feedback loop—will not have a 1:1 mapping to the sensorimotor dynamics involved. That is, the sensorimotor activity paralleling such loops may occur in a wide variety of ways. Further, the same properties may appear in a variety of different interactions. By tracking activity at the level of both the interaction and the sensorimotor activity, we can see how dimensions of sensorimotor activity get configured around shifts or transitions in interaction states.

Finally, in a systems approach development can be seen as configural change. That is, given the multidimensional nature of learning [73] and social discourse (e.g., [44]), it is incumbent to show how various dimensions reconfigure as the mother–infant system develops.

De Barbaro *et al.* [23], investigated the development of triadic attention from this perspective. Prior research had established that a shift from dyadic to triadic (mother–infant–object) play occurs between the ages of 9 and 12 mo (see Fig. 1). By characterizing changing microdynamics at key moments of interaction, de Barbaro *et al.* were able to identify multiple developmental trajectories that participate in the emergence of triadic attention. That is, by tracking the same dimensions of activity—gaze, left, and right hands,—of both infant and mother in relation to objects and one another during free-play interactions, the authors found that it was the infants' configuration of this sensorimotor activity in response to mothers' bids for attention that changed over developmental time.

In de Barbaro *et al.* [23], videotapes of mother–infant dyads at 4,6, 9, and 12 mo were analyzed for interactions called "maternal bids," in which the mother made one of multiple, local objects more accessible to the infant, presumably to engage the infant. At the earliest ages, mothers made objects salient to their infant by looming them near their infants' face or hands; months later, infantsreached for, grasped, and brought objects towards themselves. Additionally, infants' sensorimotor coordination shifted from being highly "coupled," with their gaze and both hands directed in unison to the object manipulated by the mother, to "decoupled," where, for example, infants could gaze at one object while handling another. Eventually, when presented with a novel object, infants would alternate visual attention between his mothers' object and his or her own. This changed the dynamic of the "negotiation" of these toy bids in a way that altered coordinated object play. Specifically, the decoupling of infants' sensorimotor modalities, along with the mother's more frequent manipulation of objects in parallel to the infant's own manipulations, now allowed interactions in which infants cycled between visually attending to their own handling of an object, their mother's handling of an object, and the mother's face (usually in mutual gaze). This change was accompanied by new dynamics of social affect expression: as infants engaged in more mother-congruent (e.g., imitative) object manipulations, mothers would time their positive affect with these actions (e.g., congratulating the infant), presumably reinforcing infants' behavior.

By relating changes in interaction level activity (e.g., emergence of imitation) to microdynamics of infants' sensorimotor actions, we could characterize more gradual or continuous shifts between several dimensions of activity participating in the emergence of "triadic interactions" around the end of the first year [77]. In particular, we observed changes in the changing hand-eye coordination in the infant, adaptive scaffolding by the mother, and the decoupling of multimodal attention that allows both partner and objects to be a part of the same elongated sequence of activity. These findings are not mutually exclusive with a qualitative shift in dyadic interactions; indeed, a principle of complex systems is that quantitative shifts in components can result in qualitative shifts in system states [76].

Recent technological advances can be leveraged towards this systems based approach to development. The emergence of small and inexpensive high definition cameras, sensors, and data storage promises to expedite the collection of many channels of temporally and spatially precise activity and physiological data. With these advances, we can now synchronize multiple independent streams of precise activity data with shifts in interaction-level activity.

Epistemologically, this increase in dimensionality of datasets allows researchers to shift scientific focus from discrete, present-or-absent *products* of social interactions to *process models* of interactions. Simultaneously capturing both levels of activity allows the researcher to characterizechanges in the timing and manner by which these component dimensions of activity participate across transitions in system-level activity. Assessed longitudinally, such analyses can reveal the developmental course of changes across transitions in mother infant shared activity and coregulation.

III. Getting to Know Your Phenomenon

Perhaps one of the greatest challenges in multidimensional, multiscalar, multiparty research is to define the relevant phenomena and to consider what elements of the activity are most important to capture for constructing a developmental account. Relevant timescales can vary widely. Hormones, for example, can act on timescales of minutes to hours, while other physiological shifts, like rushes of norepinephrine associated with surprise [4], come and go more quickly. An episode of imitation may occur on the timescale of seconds, but individual behaviors within the episode may unfold within 10ths of seconds, and underlying neural processing patterns on the scale of 100ths of seconds. Deciding which timescales are relevant to include depends both on logistical and theoretical constraints. Ultimately, all such decisions depend upon identifying the "phenomenon of interest."

Thus, "getting to know" your phenomenon is of utmost importance. We recommend a period of systematic but qualitative ethnographic analyses of interactions prior to formal data collection. Qualitative analyses allow for an in-depth analysis of the structure of episodes of interest, and generation of hypotheses about those events, without the constraints of predetermined categories that might obscure the developmental process. Careful scrutiny of video samples of the interactions of interest, at varied speeds, with notation and discussion, is also critical. Such foundational qualitative analyses can identify the boundaries of the social system of interest, and determine the spatial, structural, and temporal scales that will be optimal for the theoretical questions of interest. Methodological guidance for conducting extensive qualitative analyses can be found elsewhere (e.g., [59]).

This preliminary qualitative analysis can, and often should, be enhanced by pilot systematic coding. This is an iterative process in which coding schemes are tried, found wanting, modified and retried. It can be particularly helpful to produce visualizations during these preliminary attempts, prior to formalizing the final coding scheme. Different visualizations can abstract over different features of the data (see examples in Fig. 1 and 2) and thus serve as a tool for generating hypotheses [41]. During our research we repeatedly return to scrutinize video and visualizations to generate hypotheses about phenomena, test our formalizations, and interpret our results.

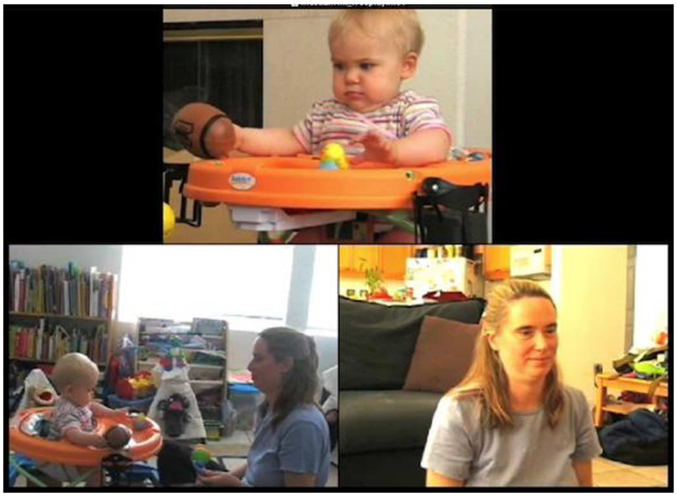

Fig. 1. Screenshot of synchronized videos taken during infant-mother object-play interaction in homes ([see de Barbaro et al. [94], for description). Note that camera angles focus on infant upper-body, mother upper-body, and the dyad from a zoomed-out perspective.

In observational studies like those described here, the processes of identifying relevant interaction level events, and selecting pertinent components of interaction activity to code in detail are tantamount to forming a hypothesis about which dimensions matter for the phenomena of interest. We now consider some factors relevant to defining these variables.

IV. Identifying Events to Frame Interaction Activity

Social interactions subsume such a range of contexts, activities, and behaviors that selecting a high-order category of interaction can be a daunting methodological decision. However, several heuristics can aid that decision. First, an age range should be selected wherein the phenomenon of interest is absent or rare at the younger end of the range, but robust and observable at the upper end. Based on the developmental changes of interest, researchers should select a social event category that "frames" or identifies key moments of the interaction for characterizing shifting microdynamics relevant to the mature phenomenon. In the case of our analysis on the development of triadic attention (de Barbaro *et al.* [23]), the framing event was the maternal bid; in the case of [82] analysis of the emergence of reaching, the framing event was all object oriented movements. These framing events should occur with enough frequency (either naturally or through some mild manipulation of the context) across the full developmental range so that there is adequate data for analysis of developmental change. Finally, the events should reveal sufficient variability across age, individual dyads, or both to support critical hypothesis testing regarding shifts in the phenomenon of interest.

Finding framing events that satisfy the constraints above (age-appropriate, frequent, variable) and other pragmatic

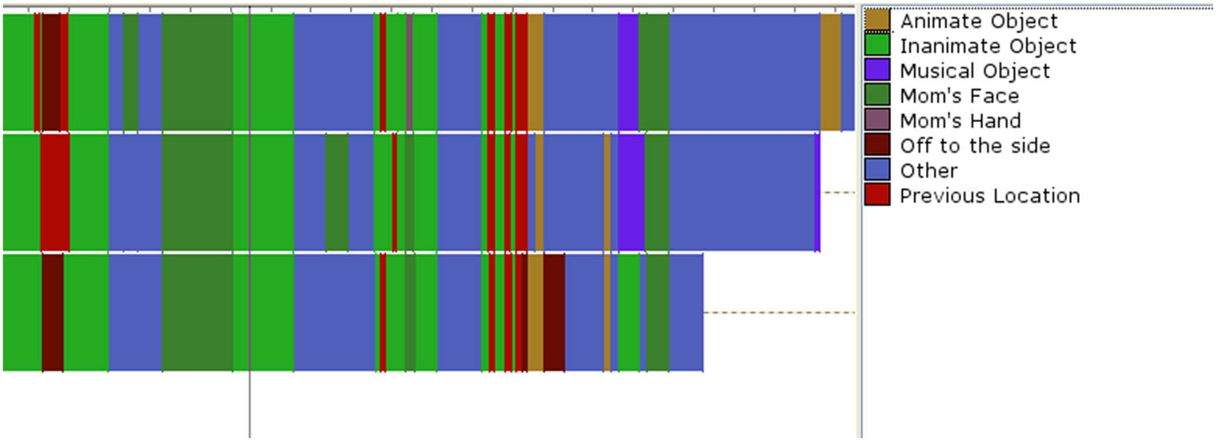

Fig. 2. Visualization (Mangold Interact) of three students' codes of manual activity (stacked vertically) created for a coding meeting. Colors indicate different coding categories as indicated by the key.

constraints (e.g., feasibility for systematic coding) can be challenging. One approach is to generate framing events through bottom-up, qualitative analysis. As described above, for example, de Barbaro *et al.* [23], observed, by extensively watching and discussing videos of mother–infant play, that infants responded very differently with age to mothers' bids to introduce new toys into joint play. As these changes fit and move beyond claims that triadic attention emerges around 9-12 mo, these maternal "toy-bid" events were selected as pivotal framing events. In practical terms, these events also were relatively easy to identify, and had certain common properties (e.g., initiating actions by the mother, shared targets of attention, etc.) that made it possible to compare across the full age range of 4 to 12 mo.

A different sort of interaction-level framing event is illustrated by [48] taxonomy of infant–parent dyadic behaviors, which parses periods of dyadic *symmetrical communication, asymmetrical communication, unengaged attention*, and others. The categories may manifest themselves quite differently at different ages, while still sharing important relational features. For example, in older infants symmetrical states encompass turn taking and imitation, whereas in younger infants they encompass coordinated smiling and bouncing. The use of such high-level constructs may shed light on the nature of important qualities of interaction such as positive engagement, regulation, feedback, and reinforcement. Such schemes are "risky" insofar as the categories are somewhat more subjective and harder to code, and may be difficult to replicate across social contexts or studies. Nevertheless, the use of such abstract categories in a systems analysis offers a powerful method to uncover details as to the changing dynamics of these important activities.

By aiming to find events that will stand as "naturalistic trials" of a particular type of real-world interaction, we will be best positioned to understand how the relevant microdynamics change across shifts in interaction-level activity.

V. Components to Track

A. Embodied Modalities of Access and Activity

At its core, cognition is perception and action, which are inextricably embodied processes. On the millisecond timescale, we move and adjust not just our eyes, but also our hands, our posture, and indeed our whole bodies to gain access to, and act upon, features of our environments [40], [65], [80], including social partners. Infants' rapidly changing bodies and sensorimotor practices affect how and what they can access and act upon in their social environment. The environment is not static, however; caregivers' actions interact with the infant's ongoing activity and perception, modifying their experience and scaffolding learning. Our claim is that by tracking both parties' activity in detail, at a high temporal resolution, we can gain a direct read of developmental processes as they unfold.

For example, in a shared picture-book reading activity, instead of simply coding gaze, or specifying anamodal "focus of attention," tracking manual actions can provide an important details as to how dyads engage with the material. For example, who is holding the book, or who is turning the pages, will shape where and how gaze is focused, and thus what will be salient and prone to being learned (e.g., [74]). Similarly, as factors like posture and self-sitting also influence access to materials (e.g. [75]), coding the orientation of the infant's torso, hips, and head relative to the book and the parent may be revealing.

Sensorimotor development also changes infants' interactions with social partners [14], [16], [32]. For example, manipulating infants' posture can dramatically change the proportion of time that infants spend looking at social partners' faces [32]. Furthermore, parents adapt their behaviors to their infants' changing sensorimotor skill by producing different patterns of micro-behaviors to encourage increasingly sophisticated shared activity. For example, parents put their hands on their infant's hands to facilitate shared object manipulation [84], [93]. Likewise, gestures that begin as physically enacted forms of engagement can crystallize as symbolic activity, as the participants develop well-practiced routines of interaction [51]. Finally, language and vocalizations are important activities in their own right whose timing in the unfolding activity may have important developmental consequences [43], [89].

B. Measures of Affect and Arousal

Another set of dimensions to consider coding are behavioral correlates of affect and arousal. For example, changing dynamics of facial expressions [27] can be coded in infants [67] in precise detail using "FACS," a system for the micro-analysis of the different sets of facial muscles involved in different expressions. One challenge is that FACS coding is quite labor-intensive, even for trained coders—and training can be intensive. However, advances have also been made in the automated coding of such expressions (see also section IVB; [9], [62], and [64]). FACS coding could be used to capture changing temporal dynamics of musculature to gain insights into the emergence of categories of facial affect expressed by infants (e.g., [10]). Even simple measures of facial affect can be revealing in studies of social interaction. It can be particularly useful to look at the relative timing and extent of facial affect in both participants to address how such behavior facilitates the coordination of other activities. However, it is beneficial to collect a multiplicity of measures of affect and arousal if possible.

Affect and arousal are also embodied in other ways, such as rate of kicking or wiggling, or rate of sucking a pacifier (e.g., [11], [81]). Such measures can serve as theoretically relevant indices as well as converging measures of affect. Another option is to collect measures of autonomic nervous system activity. These include, for instance, galvanized skin responses, pupil dilation, and/or hormonal or other biochemical changes [13]. Some studies are now also looking at mother relative to infant measures, ranging from RSA to HR (e.g., [29]). This work has begun to document how autonomic states relate to social behavior and affect [30]. However, there remain many underexplored relations between ANS activity and learning (e.g., [22]).

C. CNS Activity

While central nervous system activity certainly plays a role in all elements of activity described here, it is still mostly impractical to capture CNS activity during live social interactions. However, recent work from G.D.'s lab [61] has shown that dual EEG recordings from toddlers and parents, submitted to ICA analysis techniques, can be used to capture cortical electrophysiological components of *murhythm* suppression [69] during interaction. This marker of a "mirror neuron" system was detected both when toddlers took a turn in a dyadic game, and when they watched their parent take a turn. Notably, a touchscreen recorded game actions, and an optical motion-tracking system (Natural Point, USA) allowed the synchronization of EEG with both discrete and continuous arm and head movements of both participants. Although such a system is currently quite difficult to design, resource-demanding, and somewhat limiting of natural behaviors, it can integrate otherwise unattainable combinations of data relating CNS activity to minimally-scripted social behaviors.

As our measurements of behavior in interaction gain accuracy at fractions of seconds, they begin to approximate the functional timescale of peripheral and central nervous system activity. This is critical to illuminate the functional significance of central and peripheral nervous system activity for social interactions, and vice versa (e.g., [63]).

D. Extended Timescales of Activity

Finally, embodied interactions also occur within longer timescales of social activity, including relationships, family dynamics, neighborhood, socioeconomic strata, and other aspects of culture [60]. These factors impact the dynamics of unfolding interactions, for example via the history of fathers' play with children [88], or where mothers traditionally place their infants [14]. If the focus of a study is on interactions that occur over minutes, such factors will tend not to change over the course of such an interaction. However, these factors are "contexts" which may powerfully affect the dynamics of interactions. At a minimum, these contexts should be considered in the interpretation and generalizability of the results. In addition, ethnographic methods, including interviews, questionnaires, and sociological data (e.g., census) can provide covariates for more systematic study. Such approaches can also benefit from developing creative measures such as assessing neighborhood wealth in terms of the proportion of cracked sidewalks [47] or assessing stress of air traffic controllers via average visibility during shifts [70]. Used in conjunction with the above assessments, such parameters can generate a much richer account of the system under study.

VI. Systematic Data Collection

A. Choosing An Environment and Paradigm

The decisions of which behavioral systems and which timescales are relevant will partly determine what data will be captured. The physical environments available for recording interactions are an additional determining factor. In general, the environment should have sufficiently high degrees of freedom for the social activity to unfold in a variety of ways. However, somewhat restricting the range of the environment is necessary to avoid problems with data variability.

One decision is whether to record behavior in an environment familiar to the dyad, or a controlled but unfamiliar environment (i.e., laboratory). The advantage of recording data in familiar environments is that they allow us to observe processes of cognition as they occur "in the wild". The affordances of naturalistic environments may provide structure important for understanding developmental processes. As an example, [26] found that 1-year-old infants virtually never followed their mother's gaze in a cluttered environment, indicating that features of the environment can shift outcomes of interaction (see also Deák, Krasno, Triesch, Lewis, and Sepeta, 2013). Our experience recording naturalistic social videos in living spaces—homes, zoos, and native habitats—underscore several important limitations and challenges. For example, in homes it is critical to map out rooms and light sources, and define a location where mother–infant dyads will interact. Deák and colleagues [21]also set constant distances and angles between cameras and participants, and altered lighting when necessary. Other factors to control in home environments may include visual clutter and distraction, ambient sound, and the presence of other human or nonhuman animals.

In laboratories, of course, the advantage is that these features can be controlled. This can make video consistent and controlled enough to support the use of computer vision coding of certain elements of the interaction, thereby greatly reducing video coding time (see reference to [89], below).

Another decision concerns how to allocate recording "bandwidth" to behavioral systems. In home or field settings, it is challenging and sometimes impossible to collect multiple video streams to record micro-behavioral details. Even in laboratory environments it is currently challenging to synchronize multiple full-resolution (NTSC) video streams. Thus, researchers must decide how important it is to have synchronized high-quality or high-definition video. If, for example, close-up face video is a priority, it is also important to collect zoomed-out contextual video of both participants from at least one angle for understanding the nature and timing of the interaction. An example of this, from the Deák laboratory, is shown in Fig. 1.

Moreover, collecting close-up face video will limit how dyads can move around. This restriction can introduce stress and frustration, especially for infants aged 10–18 mo, but even older children tend to move around much more than adults when seated, and restricting motion will elicit emotional responses that alter the quality of the interaction. By contrast, allowing participants to move freely will, given current technology, limits access to fine-grained actions such as changes in facial expression. We thus constantly negotiate a tradeoff between capturing relevant data (i.e., sufficient for analysis) and capturing representative data (i.e., minimally affected by the study design).

B. Human Coding

Given an adequate video dataset, coding by trained observers allows quantification ofa wide range of behaviors, and is potentially the most judicious mediation of nuanced social distinctions. After all, the human system for detecting social information has developed over millions of years, and utilizes "wetware" that is faster and more powerful, by an order of magnitude, than any silicon computer. In particular, human coding is typically necessary for tracking shifts in activity at the level of interaction, which often requires integration across many different channels of activity. The variation across micro-level dynamics makes it nearly impossible to derive such codes from automated sensors or machine coding.

For a basic introduction to behavioral coding methods of social interactions, see [5]. Hand coding can be conducted inexpensively using sophisticated open-source freeware such as ELAN ([86], available at http://tla.mpi.nl/tools/tla-tools/elan/). Alternatively, commercial coding software such as Mangold-Interact (http://www.mangold-international.com) or Noldus Observer (www.noldus.com) provides additional features such as automatic kappa calculation, and simple visualizations, which are useful for attaining high intercoder reliability (see Fig. 2). Given the difficulty of this process with complex naturalistic video records, such functionality is significant.

Other aspects of the human coding process involve laboratory management. Many university laboratories train undergraduate students to code for experience or course credit. Each of the authors has developed a practicum/seminar course to train undergraduate students in the theories and methods of social behavior research, and many work on projects that include video coding. Importantly, these student collaborators often become critical collaborators and informants who assist us in quality-testing and improving nascent coding schemes.

However, there are inherent shortcomings of hand-coding. The most significant is that it is extraordinarily time-consuming. For *each* dimension of activity coded in our dyadic object-play project, developing coding schemes and training to reliability required several months of weekly meetings with 3-4 students, each of whom coded 6-8 hours/week. Thereafter, experienced coders may require up to 1 hour/minute of video to code a single behavioral dimension.

One way to reduce coding time and effort is to use event-based coding rather than continuous coding of the interaction. In event-based coding only those subsections of the interaction defined by the event are coded. Note, however, that it is also necessary to code "control" intervals to compare with the critical framing events. In these cases one can select random, equal-duration intervals that do not contain the critical framing events, or select the intervals preceding or following each framing event.

Another way to reduce coding time is to begin with relatively high dimensional coding of a randomly selected subsample of the entire data set. Preliminary analyses of such a dataset is an efficient way to verify which dimensions of activity are critical for quantifying the dynamics of interest. In our analysis of infant-parent object play, we first coded a subset of five dyads across four longitudinal sessions, and then used a more restricted coding scheme on a larger sample of 26 dyads.

Two other problems should be considered: One is that in spite of the best efforts to maintain coders' blindness to prior hypotheses, and to establish high independent intercoder reliability, human coders cannot always "filter out" potentially relevant but extraneous information to focus objectively on a single behavioral channel. For example, a coder's estimate of a caregiver's intelligence or social class might color the coder's judgments of that caregiver's parenting behaviors. Such difficulties may be reduced by judicious use of blinding methods (e.g. covering parts of the video or video without sound). The other limitation is precision. For example, coders cannot typically resolve the gaze-direction of infants in naturalistic interactions within 10° visual angle, even from fairly high-quality video. (Note also that infants' actions are less "sharp" than adults', and usually cannot be coded with equal precision.) If more precise gaze direction measures are needed, head-mounted eye trackers are an increasingly useable option (see below). Otherwise, it will be necessary to use a smaller number of more encompassing looking-targets or location codes.

C. Standards of Reliability

High standards of reliability are important for the validity of human coded data as well as testing results of machine vision data. Reliability is quantified with Cohen's kappa statistic [17], which corrects for agreements expected by chance. Values of Cohen's kappa above. 75 or. 8 have historically been described as excellent [7], [31]. Based on systematic study, Bakeman *et al.* [7] report that this range of kappa corresponds to approximately 90%–95% accuracy given coding schemes with 5 or more categorical distinctions or codes [8]. As the number of codes decline, the variability of prevalence across codes is increasingly relevant for modeling the relations between kappa and accuracy. For example, given a coding scheme with only 2 codes, kappas between. 44-.81 can correspond to a high (90%–95%) degrees of accuracy, where the more equiprobable the codes, the higher the kappa values must be to attain a given accuracy [8]. Thus, kappas below. 75 may be acceptable given the particulars of the coding scheme and the distribution of behaviors observed. For a full table detailing these expected values, please refer to [7]. However, note that, as prevalence of events becomes more variable, lower values of accuracy correspond to increasingly higher rates of false negatives, i.e. failures to identify the presence of rare events, thus, it is preferable to insist on higher values of accuracy.

We use a number of heuristic practices to facilitate high-quality human coding. First, for kappa calculation we utilize a single frame time unit (at 10 or 30 f/s; referred to as time-unit kappa by [7]. This encourages 'sharpening' of temporal resolution, which is critical for accurate assessment of durations of events and timing across multiple modalities of activity. Second, it is advisable to code behavioral dimensions independently, one at a time, and integrate the data post-hoc. When coders must attend to too many dimensions of activity, reliability tends to suffer. Third, when developing a new coding scheme, it is advisable to work with teams of several coders, each coding the same video. This allows a 'triangulation' of difficult classification decisions. A common method during both development and reliability training is to generate kappa tables and identify challenging codes by finding higher off-diagonal cells (i.e., events with different codes from two coders). We have found it immensely useful to organize coding meetings with students around shared visualizations of each coder's codes, stacked to display differences in timing and application of codes between coders (see Fig. 2). This provides a qualitative representation of reliability that makes details about the nature of mismatches apparent, more quickly guiding additional clarification or redefinition.

Ongoing weekly group meetings and regular cross-checking is necessary to avoid "drift" in coders' judgments about the most precise and ambiguous events. In addition, not every student is capable of accurate coding: selectivity of coders is crucial. G.D. has found, in training dozens of student researchers on behavioral coding, that grade point average is the best predictor of coding ability. Finally, a long-term commitment from students (e.g., minimum of 300 h of effort) tends to improve coder expertise, commitment, and work quality.

D. Automated Coding

A growing range of technologies are available for automated data collection in studies of naturalistic social interaction [66]. Here we outline current technologies that may be useful for coding features of embodied attention and action within social contexts, including sensors for eye tracking and motion capture, and computer vision algorithms.

Eye tracking fixations to computer monitors is a well-established methodology, but this precludes almost all naturalistic social interactions. Recently, however, several labs have developed eye-trackers integrated with head-mounted cameras, allowing participants to freely interact while generating gaze–direction data [38] Postprocessing by human coders may be necessary to identify or classify the target of each fixation marker; even this task might be facilitated with automated machine vision (e.g., [74], [90]).

Motion tracking of participants' body parts provides the most precise data available on sensorimotor dynamics. However, it is relatively expensive and it places numerous constraints on data collection: participants must wear suits or wired LEDs, or reflective rigid markers. Increasingly systems are utilizing smaller markers, and systems such as Sony's Kinect offer the possibility of fast marker-free tracking. However, such systems considerably limit the mobility of participants. Additionally, it is computationally complex to recreate dynamic models of the face, or hands, or locomotion bodies. Thus consider judiciously if your question requires yaw, pitch and roll as well as XYZ coordinates of activity in space or whether simply the amount of activity, which could be captured via a simpler and less expensive sensor such as an accelerometer.

As an alternative to sensors, machine vision algorithms can be designed to identify visual features in video. Machine vision algorithms have been used to identify dynamics of facial expression [9], [64], hands and objects [90]. Such algorithms typically apply filters across the 2-D video images to identify, for example, swatches of a certain hue, or illumination patterns of specific frequencies. Generally, more stark contrasts and specific distinctive features (e.g., color, edges, motion) function better. For color or internal edge features, lighting that will reduce any overlaying shadows (i.e., diffuse lighting) is ideal, and backlighting in particular is to be avoided, as it obscures all contrasts but that of the silhouette. To facilitate machine vision analyses, [90] used a completely white room with three different toys, each of a single color. This facilitated automated identification of hands and objects. However, human coders were still necessary to distinguish between hands hovering over to ys and contact with toys. This underscores the points that just as human coders must be trained to reliability on each behavioral dimension, and code in separate passes, a machine vision algorithm often will have to be developed and trained on each behavioral dimension. Moreover, just as human coding benefits from multiple video angles, each machine-coding algorithm will have ideal angles.

Another caveat is that reducing the complexity of context of the environment and range of behavior to facilitate machine coding can reduce the naturalism of the interactions and potentially alter participants' social behaviors. For example, the uniquely colorful objects in [90] paradigm might have been unusually salient attractors of toddlers' attention. As noted above, such features of the environment might affect the outcomes of the interaction.

Ultimately, as machine vision algorithms are highly sensitive to the particular conditions of recording, it is almost always necessary for behavioral researchers to collaborate with machine learning researchers to train machine vision algorithms on a given dataset and, to design video collection paradigms *a priori* to facilitate machine coding. To facilitate use of pretrained automated systems, for example, such as automated FACS coding systems, pilot testing with preliminary videos from the target dataset is critical. Qualities of the video such as illumination, angle, image size, amount of movement, and resolution can make these tools unusable. Other systems may simply lack relevant training for young subjects. For example, the Microsoft Kinect commercial motion sensor's stock software can robustly find and track adults' and children's actions, but our pilot trials indicated that it does not robustly track infants' bodies (presumably due to their different body-part dimensions).

A final point is that in automated collection and coding of social interaction, measures must be synchronized, and this requires planning. One of our best solutions is a clap-board used in movie-making, outfitted with rows of LED lights that flash whenever it is clapped. This produces a punctuate, synchronized light flash and noise that can be detected by multiple cameras and microphones. The light trigger could also generate an electronic output signal for a computer that is recording, for example, motion capture or physiological measures.

VII. Data Analysis

A. Sequential Analyses

Analytic approaches to the dynamic qualities of social interactions fall under the umbrella of *sequential analyses*. Excellent introductions to using these methods for social interaction studies are provided by Bakeman and Gottman *et al.* [5], [7], [41], [45], [46]. Here we provide a brief description of these methods and the questions they can answer.

Many past sequential techniques characterize dynamics within a single dimension of activity. The simplest consider statistics for individual codes, such as rate and duration of gaze fixations, or mean gaps between instances of infant smiling. Sequential contingency analyses can be used to test the likelihood of sequences between events or states in a single dimension of activity. For example, [2] examined which kinds of dyadic states precede states of coordinated joint attention. The simplest contingency analyses examine cooccurrence between states or alternatively transitional probabilities between states at different time lags (e.g., "1-back" states). More complex multiway contingency table analyses (e.g. using log-linear models) can examine such transitional probabilities relative to other variables such as infant gender or mood. The latter variables serve as constants for these analyses.

Other analyses examine relations between pairs of dimensions—for example, the likelihood that mothers will synchronize motion with utterances to infants (e.g., [42]). Dependencies between dimensions can be tested via creating a temporal window (e.g., to quantify the likelihood of infant gaze shifts within 5 s of a parent's point). Subsequent tests can consider the proportion of event sequences, the timing of sequences, or both.

Timeseries analyses can characterize cycles of activity within and between sequences of activity measured at uniform intervals of time. Such analyses were used to identify that cyclic patterns in infants' activity are not periodic but rather fluctuate in response to cycles in the mother's activity (e.g., [57]). Also, changes in infants' activity precede *and* follow changes in mothers' activity, indicating that infants not only respond to but also influence parents' activity [1], [19], [57]. These analyses characterize important dynamics of mother–infant activity.

In sum, the goal of sequential analyses is to find quantitative patterns of the dynamics of infant-parent activity. Within our approach, we utilize these varied techniques to characterize changing dynamics within and across components of interaction before, during, or after framing events relevant to qualitative shifts in system level activity (see Section IV). In particular, we use these tools insofar as they allow us to create quantitative measures of activity that span the entire time period of developmental interest. This is critical for a quantitative assessment of the changing processes accompanying shifts in system level activity.

In the remainder of this section we describe special considerations, common problems, and best practices for using sequential analyses with multiple time-scale, multimodal datasets of naturalistic interaction.

B. Special Considerations for Analyzing Multiscale, Multidimension Social Data

The statistical tools for analyzing developmental social behavioral systems are derived from standard sequential analyses. However, such data merit additional considerations.

Sequential analyses of one or two binary dimensions of activity are relatively straightforward. However, when additional dimensions of activity with multiple values are added, there can be a computational explosion of analytic possibilities. For example, the number of possible time series relations to be explored increases exponentially. Even a somewhat reduced state space of our mother–infant object-play data, considering all combinations of five channels with three values each, yields 225 possible 1-back transitions between states. Given the expense of collecting and coding large social-interaction data sets, such a state-space will typically be too large to explore. If a state-space is too sparse, analyses will be weak, and results will not plausibly generalize to other contexts.

The problem of computational explosion can be exacerbated by the fact that regularities in the activity may differ or shift across the interaction. These changes are unlikely to be random: for example, the timing of infants' gaze-hand coordination may be quite different when reaching for "novel" objects versus cycling attention to objects in their lap. Mothers' responses to infants' smiles early in an interaction might affect infants' later bids and the dyad's unfolding action sequences. Patterns of activity can also shift during an interaction due to dynamic factors such as infant fatigue or hunger.

Given these concerns, preliminary focused examination of video samples and preliminary data analyses and visualizations are necessary to restrict the main hypotheses and the search space. We have two main strategies in this vein.

First, we use our framing events (see Section IV) to parse the stream of interaction into naturalistic events like "trials." This parsing is analogous to time-locking cortical activity to stimuli in event-related-potential (ERP) analyses. Within these trails we utilize classic sequential analysis techniques to quantify the microdynamics of activity of component dimensions of activity. These can range from the presence or absence of activity, event contingencies, and temporal relations, both within and between behavioral dimensions.

Thus, in our approach, tracking system-level activity in combination with micro-components has multiple functions. Epistemologically, characterizing the configurations of component dimensions at key moments of interaction relevant to shifts in system activity can reveal the nature of developmental processes. Analytically, these framing events can make identifying regularities in complex data more tractable. Specifically, framing events function as hypotheses for narrowing the search space. Of course, the strength of this approach depends on preliminary qualitative examination of the data.

Additionally, at the stage of analysis, iteratively cycling between various visualizations of data and inferential statistics will be critical. As [7] note, sequential analyses are not "off the shelf." No single analytic tool can characterize dense, multichannel behavior dynamics of social interaction. Investigators should anticipate a laborious process of converging on a set of tools that will capture the temporal dynamics of interest. The importance of this process is recognized and has been articulated by others [39], [41], [92]. In the next section, we provide examples of this process from our own analyses.

Finally, before moving to the next section, we note that analyses may also be facilitated by a growing range of bottom-up or data-driven approaches to finding regularities in dense, high-dimensional social-behavioral data [15], [91]. A full review is beyond the current scope of our paper; but we anticipate continued expansion these approaches.

C. From Data Visualization to Summary Measures

Iterative cycles of creating visualizations and preliminary measures to capture dynamics surrounding framing events can inspire insight as to the specific measures or models which will best capture the dynamics of activity observed surrounding the framing events, as well as to precisely identify parameters of framing events or summary measures.

For example, our qualitative analyses of mother–infant object play indicated that whereas 4-mo-olds directed gaze and manual activity to objects manipulated by their mother, older infants were increasingly able to cycle gaze and manual contact between objects manipulated by mothers and other objects. To characterize this longitudinal change, we coded all of infants' looking and manual actions to each object, relative to mothers' periodic bids to introduce new objects.

As described in Section IV, the start and stop boundaries of framing events may be defined via independently coded states of interaction. Alternatively, the boundaries of framing events may need additional operationalization to best captured changing dynamics across the developmental period.

For example, observations of data indicated that the bid negotiation resolved very quickly in younger infants relative to older infants. In particular, older infants often distributed activity between bid toy and previous toys for extended periods of time, a feature of the response to the bid we found theoretically salient. Typical event-related analyses identify a fixed "window" of activity following the catalyzing event. However, in order to characterize the changing dynamics surrounding the bid event quantitatively, we needed to operationalize the response to the bid in a manner that would capture both dynamics but would also not overshoot the "true" negotiation, obscuring the dynamics following the bid. This required us to use particular features of the unfolding activity following each presentation of a novel toy, within each interaction, to "set" the end of each bid negotiation.

To identify the parameters that would best satisfy our conditions for the framing event, we went through multiple iterations of algorithms to identify potential end conditions. Bid negotiation periods from each iteration were overlaid upon visualizations of mother and infant sensorimotor activity (see Fig. 3). This process allowed us to operationalize the event to best capture our intuitions regarding the relevant period of negotiation. As with more formal machine learning optimization approaches, it is critical to use only a subset of data (we chose 5 of 26 dyads) to use during this process of operationalizing framing events or outcome measures. This ensures we are not over fitting our measures to our data.

Next, a variety of visualizations of component dimensions before, during, or after the framing event can be used to qualitatively characterize the dynamics of activity. As the microdynamics of multimodal, multiparty activity are only beginning to be studied, beginning this process with raw data is critical so that dynamics of activity can be accurately represented and characterized. Summary measures can obscure important unknown dynamics. Note that a variety of visualizations, each abstracting or highlighting a different element of the interaction, will be useful for identifying and characterizing temporal dynamics of activity.

For example, Fig. 4 shows time-line visualizations representing how infants directed visual and haptic activity to newly offered toys, versus any other object during the negotiation following each maternal bid (see Fig. 4; see [23]). We created these timeseries by summing the number of infant modalities directed to the bid object (in solid green) and all objects (in dotted red). From this we were inspired to derive an outcome measure, *total modality time* (TMT), by summing these curves across all frames occurring within the boundaries of the bid negotiation (i.e., taking the area under each curve). Our final summary measure is the relative proportion of the TMT to the bid toy relative to the TMT to all toys.

However, TMT curves and measures gloss over various other patterns, such as the timing of directingmodalities to objects, or the dynamics of continued mother activity during the bid effect. To capture these dynamics additional measures and perhaps novel framing events might be created and defined, and these would be best facilitated via other visualizations (e.g., Fig. 3).

D. Software for Data Analysis

Standard coding software (Mangold, Noldus, Elan), as well as specialized analysis programs such as GSEQ, developed by [6] provide tools for accessing, manipulating, visualizing and analyzing sequential analysis data. Each of these programs allow researchers to combine dimensions of activity and perform contingency and simple event-sequential analyses, with GSEQ offering the most flexible and powerful tools for state based analyses. Additionally, most can create simple visualizations. However, the inflexibility of these visualizations, and the overall focus of these programs on event-based analyses in

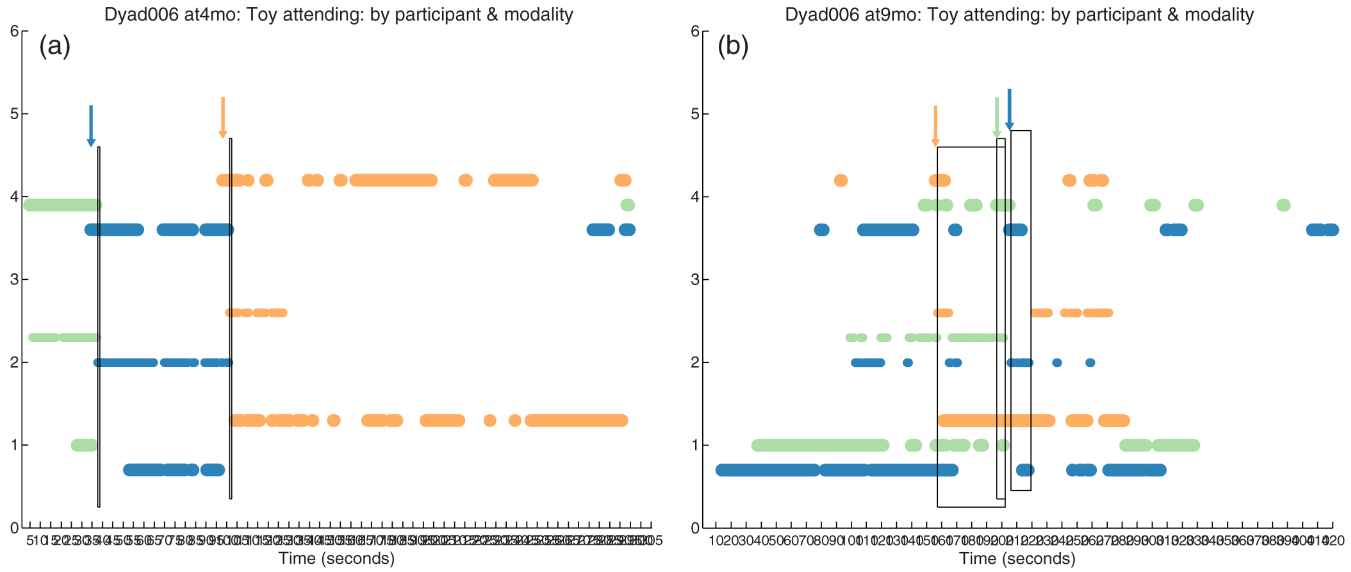

Fig. 3. Typical multimodal activity in (3a) 4mo and (3b) 9 mo interacting dyads. Note that interaction in (3a) is 45 s in duration, while (3b) is 70 s. Segments along the y-axis indicates a specific partner (infant or mom) and sensorimotor modality (gaze, hands). Colored marks indicate, frame-by-frame, moments of sensorimotor contact with any of three objects (indicated in green, blue, and orange). Maternal bids are indicated via arrows; boundaries for bid negotiation periods are indicated via black boxes. Fig. (3a) shows a 4 mo dyad with complete transition of gaze and hands from previous object to maternal bid object. By contrast Fig. (3b), at 9 mo, shows cycling of infant modalities between maternal bid toy and previously attended toys (within each of three visible bids). Bids are labeled with the TMT outcome measures.

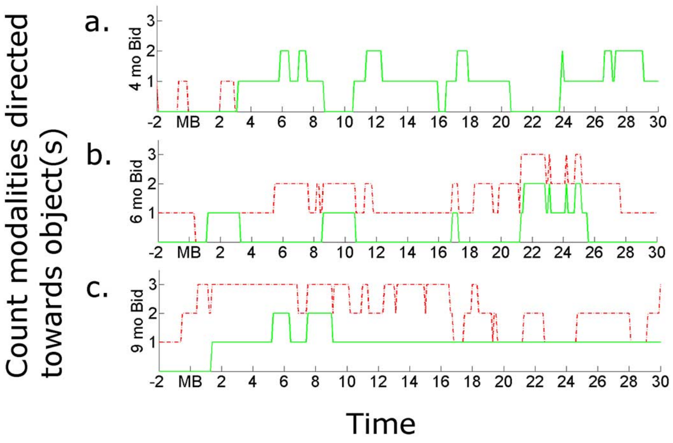

Fig. 4. Each panel indicates a fixed window of interaction 2 s prior and 30 s following an instance of mother's presentation of a novel bid object (MB, set to $\mathrm{time} = 0$), (4a) at 4 mo, (4b) 6 mo, and (4c) 9 mo. Solid green timeseries indicate the total count (i.e., the sum) of infant modalities (including gaze, right hand, and left hand) directed at bid object at each frame of interaction. Dotted red timeseries indicate the sum of infant modalities directed toward all objects (including bid object) at each frame of interaction. Note the increasing proportion of modalities to nonbid toys after 4 mo, and the more frequent and higher amplitude activity (corresponding to more rapid cycling of multiple modalities, e.g., gaze and hands rather than simply gaze) to the bid toy between 6 and 9 mo.

1-2 dimensions may limit the cyclical process of exploration, analysis, and discovery that will often be necessary to discover complex multimodal dynamics. Nevertheless, such tools may efficiently guide simpler analyses such as identifying cooccurrence or latencies between modalities.

For more complex dynamics, programming languages such as Matlab are preferable in that they are much more flexible in visualizations and in the range of analyses supported. However, they require some coding skill. Alternatively, in the case of visualizations, ChronoViz (http://chronoviz.com/index.html) provides Mac OS users with a user friendly open-source toolkit for flexibly integrating and plotting many types of heterogeneous time-coded data simultaneously with video [37]. For complex analyses, the use of flexible programming software may be necessary.

If Matlab or a similar programming language is used for analyses, we recommend storing data in the form of timeseries. Integrating all variables in a common temporal structure is important for looking across different temporal scales of activity, and timeseries provide a highly flexible format for the various sequential analyses that may be of utility [7]. Categorical data with multiple categories (e.g., targets of gaze) can be transformed to an equal number of binary time series indicating gaze activity to each target. For state based analyses, time series data can easily be transformed into events.

In an integrated dataset of this sort, all dimensions of activity should be entered on the scale of the dimension with the smallest temporal unit. This means that slower-sampled variables will be represented as if they were sampled at higher frequency. This can be misleading. It is therefore important to make a reasoned decision about whether to assign the beginning or endpoint of a slower-sampled variable event to the time-point of the first, or maximum, or average datapoint of the highest-frequency variable's concurrent event. For example, if we are sampling EEG at 256 Hz and eye tracking at 100 Hz, a decision must be made about which EEG sample will cooccur with the corresponding eye direction vector that is sampled for every two to three EEG time points. The potential variability in sampling and reduction procedures means that when reporting methods for publication, the sampling rates for each dimension should be reported, as well as the procedures by which samples are reduced.

E. Statistics

Once features capturing dynamics of activity are identified and calculated, standard statistical tools such as t-tests and repeated measures analysis of variance (RM-ANOVA) can be used to assess developmental changes. For example, comparing the TMT for bid toys relative to other toys across the first year indicated that while infants attend to maternal bid objects across all sessions, increasingly over the first year they allocate more modality-time to other objects relative to the maternal bid object. In other words, the shift to triadic attending at the end of the first year was paralleled by an increasing tendency to distribute attention between objects manipulated by mom and those in the infants' own possession. This finding lead us to the hypothesis that the infants' progressive distribution of attention allows them to attend to and incorporate some elements of their mother's actions into their own play, as in episodes of imitation or games.

Predictive modeling approaches derived from machine learning models provide flexible methods that can test for relations between many types of variables simultaneously. For example, supervised classification models combine weights on sets of "input features" to predict an outcome measures. Such models are of particular importance for understanding the features most critical for predicting an outcome. Potential outcomes can range from continuous measures (e.g., percent modality time for each bid negotiation within an interaction) to binary or categorical measures (e.g., whether or not a given word was successfully learned) depending on the model used. The flexibility in input features allows one to include features that capture activity from fractions of seconds (e.g., who was holding the object of interest at the moment of naming) to months surrounding the event of interest (e.g., days since onset of reaching). To examine whether certain features differentially predicted the outcomes across a developmental period or across dyads, we could create multiple separate models each with restricted instances of activity.

We recommend using classification models limited to linear combinations of the input features. Nonlinear classification can be more powerful, but the results can also be more difficult to interpret. Next, the number of event instances included in the model affects the number of input variables that can be tested. Without a sufficient number of examples, it is necessary to either restrict the number of inputs, use more directed statistical analyses, or identify a more common event.

Moreover, while the "blind" nature of machine learning approaches is appealing because such models provide theoretically transparent demonstrations of discovering patterns, these methods cannot replace the process of getting to know your data. This is critical for interpreting model results. For example, due to the optimization process, models will favor features with even just slightly stronger regularity than theoretically relevant features that are slightly less regular. Thus, machine learning analyses require judicious interpretation.

VIII. Conclusion

The proliferation of new methods and tools for the collection, coding, and analysis of social interactions permit studies of unprecedented sophistication and power. In particular, novel technologies allow the collection of highly precise, high-density records of multiple participants' activity. By synchronizing multiple streams of activity data at multiple time-scales of activity, we shift the epistemological framework for thinking about interaction from simple products to complex processes.

This insight from systems approaches allows us to revisit accounts of abrupt or qualitative development. In particular, we can characterize regularities in the timing and sequences by which components of interaction activity—the modalities of arousal and affect, as well as the modalities by which our subjects access and act upon their environments—participate across transitions in system-level activity, and thereby reveals the trajectories of developmental change.

Whereas previous behavioral coding schemes often indicated sudden, qualitative shifts in interaction activity, we describe quantitatively the dynamics of the system that can result in qualitative changes in behavior [76]. In this way we will begin to converge on more powerful predictive and even explanatory accounts of the origins of social skills.

References

[1] L. Adamson and R. Bakeman, "The development of shared attention during infancy," *Annal. Child Develop.*, vol. 8, pp. 1–41, 1991.

[2] L. B. Adamson and R. Bakeman, "Mothers' communicative acts: changes during infancy," *Infant Behav. Develop.*, vol. 7, no. 4, pp. 467–478, 1984.

[3] K. E. Adolph, S. R. Robinson, J. W. Young, and F. Gill-Alvarez, "What is the shape of developmental change?," *Psychol. Rev.*, vol. 115, no. 3, p. 527, 2008.

[4] G. Aston-Jones and J. D. Cohen, "An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance," *Annu. Rev. Neurosci.*, vol. 28, no. 1, pp. 403–450, 2005.

[5] R. Bakeman and J. Gottman*, Observing Interaction: An introduction to sequential analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1997.

[6] R. Bakeman and V. Quera*, Analyzing Interaction: Sequential analysis with SDIS and GSEQ*. Cambridge, U.K.: Cambridge Univ. Press, 1995.

[7] R. Bakeman and V. Quera*, Sequential Analysis and Observational Methods for the Behavioral Sciences*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[8] R. Bakeman, V. Quera, D. McArthur, and B. F. Robinson, "Detecting sequential patterns and determining their reliability with fallible observers," *Psychol. Methods*, vol. 2, no. 4, pp. 357–370, 1997.

[9] M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, "Measuring facial expressions by computer image analysis," *Psychophysiology*, vol. 36, no. 2, pp. 253–263, 1999.

[10] D. S. Bennett, M. Bendersky, and M. Lewis, "Does the organization of emotional expression change over time? Facial expressivity from 4 to 12 mos," *Infancy*, vol. 8, no. 2, pp. 167–187, 2005.

[11] R. Bigsby, W. Coster, B. M. Lester, and M. R. Peucker, "Motor behavioral cues of term and preterm infants at 3 mos," *Infant Behav. Develop.*, vol. 19, no. 3, pp. 295–307, 1996.

[12] T. B. Brazelton, B. Koslowski, and M. Main, "The origins of reciprocity: The early mother-infant interaction," in *The Effect of the Infant on its Caregiver*. New York, NY, USA: Wiley, 1974, pp. 49–76.

[13] J. T. Cacioppo, L. G. Tassinary, and G. Berntson*, Handbook of Psychophysiology*. Cambridge, U.K.: Cambridge Univ. Press, 2007.

[14] J. J. Campos, D. I. Anderson, M. A. Barbu-Roth, E. M. Hubbard, M. J. Hertenstein, and D. Witherington, "Travel broadens the mind," *Infancy*, vol. 1, no. 2, pp. 149–219, 2000.

[15] H. Choi, C. Yu, L. Smith, and O. Sporns, "From Data Streams to Information Flow: Information Exchange in Child-Parent Interaction Experiment and Data Preprocessing," presented at the 33nd Annu. Conf. Cogn. Sci. Soc., .

[16] M. W. Clearfield, "Learning to walk changes infants' social interactions," *Infant Behav. Develop.*, vol. 34, no. 1, pp. 15–25, 2011.

[17] J. Cohen, "A coefficient of agreement for nominal scales," *Edu. Psychol. Measur.*, vol. 20, no. 1, pp. 37–46, 1960.

[18] J. Cohn and E. Tronick, "Mother–infant face-to-face interaction: The sequence of dyadic states at 3, 6, and 9 mos," *Develop. Psychol.*, vol. 23, no. 1, p. 68, 1987.

[19] J. F. Cohn and E. Z. Tronick, "Mother-infant face-to-face interaction: Influence is bidirectional and unrelated to periodic cycles in either partner's behavior," *Develop. Psychol.*, vol. 24, no. 3, p. 386, 1988.

[20] M. Cole, *Cultural psychology: A once and future discipline*. Boston, MA, USA: Harvard Univ. Press, 1996.

[21] J. Danly, J. A. Acuña, and G. O. Deák, "Infant attention-following at home: a longitudinal study from 4-9 mos," presented at the Int. Conf. Develop. Learn., Shanghai, China, 2009.

[22] K. de Barbaro, A. Chiba, and G. O. Deák, "Micro-analysis of infant looking in a naturalistic social setting: insights from biologically based models of attention," *Develop. Sci.*, vol. 14, no. 5, pp. 1150–1160, 2011.

[23] K. de Barbaro, C. M. Johnson, and G. O. Deák, "Twelve-mo "Social Revolution" Emerges from Mother-Infant Sensory-Motor Coordination: A Longitudinal Investigation," *Human Develop.*, 2013.

[24] K. de Barbaro, C. M. Johnson, D. Forster, and G. Deák, *Infant Sensory-Motor Decoupling Contributes To 12 mo Social "Revolution": A Longitudinal Investigation Of Mother-Infant-Object Interactions*, submitted for publication.

[25] G. O. Deák, A. Krasno, J. Triesch, J. Lewis, and L. Sepeta, "Watch the hands: Infants learn to follow gaze by seeing adults manipulate objects," *Develop. Sci.*, to be published.

[26] G. O. Deák, T. A. Walden, M. Yale Kaiser, and A. Lewis, "Driven from distraction: How infants respond to parents' attempts to elicit and re-direct their attention," *Infant Behav. Develop.*, vol. 31, no. 1, pp. 34–50, 2008.

[27] P. Ekman and E. L. Rosenberg, *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. London, U.K.: Oxford Univ. Press, 1997.

[28] J. Elman, "Development: It's about time," *Develop. Sci.*, vol. 6, no. 4, pp. 430–433, 2003.

[29] R. Feldman, R. Magori-Cohen, G. Galili, M. Singer, and Y. Louzoun, "Mother and infant coordinate heart rhythms through episodes of interaction synchrony," *Infant Behav. Develop.*, vol. 34, no. 4, pp. 569–577, 2011.

[30] T. M. Field, "Infant gaze aversion and heart rate during face-to-face interactions*," *Infant Behav. Develop.*, vol. 4, pp. 307–315, 1981.

[31] J. Fleiss, *Statistical measures for rates and proportions*. Hoboken, NJ, USA: Wiley, 1981.

[32] A. Fogel, "Movement and communication in human infancy: The social dynamics of development," *Human Movement Sci.*, vol. 11, no. 4, pp. 387–423, 1992.

[33] A. Fogel, *Developing Through Relationships*. Chicago, IL, USA: Univ. Chicago Press, 1993.

[34] A. Fogel and E. Thelen, "Development of early expressive and communicative action: Reinterpreting the evidence from a dynamic systems perspective," *Develop. Psychol.*, vol. 23, no. 6, p. 747, 1987.

[35] D. Forster, "Consort turnovers as distributed cognition in Olive baboons: A distributed approach to mind," *Cogn. Animal*, pp. 163–171, 2002.

[36] D. Forster and P. F. Rodriguez, "Social Complexity and Distributed Cognition in Olive Baboons (Papio anubis): Adding System Dynamics to Analysis of Interaction Data," *Aquatic Mammals*, vol. 32, no. 4, pp. 528–543, 2006.

[37] A. Fouse, N. Weibel, E. Hutchins, and J. D. Hollan, "Chronoviz: A system for supporting navigation of time-coded data," presented at the 2011 Annu. Conf. Extended Abstracts Human Factors Comput.Syst., 2011.

[38] J. M. Franchak, K. S. Kretch, K. C. Soska, and K. E. Adolph, "Head Mounted Eye Tracking: A New Method to Describe Infant Looking," *Child Develop.*, 2011.

[39] D. Fricker, H. Zhang, and C. Yu, "Sequential pattern mining of multimodal data streams in dyadic interactions," presented at the 2011 IEEE Int. Conf. Develop. Learn., 2011.

[40] E. J. Gibson, "Exploratory behavior in the development of perceiving, acting, and the acquiring of knowledge," *Annu. Rev. Psychol.*, vol. 39, no. 1, pp. 1–42, 1988.

[41] A. Gnisci, R. Bakeman, and V. Quera, "Blending qualitative and quantitative analyses in observing interaction: Misunderstandings, applications and proposals," *Int. J. Mult. Res. Approaches*, vol. 2, no. 1, pp. 15–30, 2008.

[42] L. J. Gogate, L. E. Bahrick, and J. D. Watson, "A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures," *Child Develop.*, vol. 71, no. 4, pp. 878–894, 2003.

[43] M. H. Goldstein and J. A. Schwade, "Social feedback to infants' babbling facilitates rapid phonological learning," *Psychol. Sci.*, vol. 19, no. 5, p. 515, 2008.

[44] C. Goodwin, "The co-operative, transformative organization of human action and knowledge," *J. Pragmatics*, 2012.

[45] J. Gottman, *Time-Series Analysis: A Comprehensive Introduction for Social Scientists*. Cambridge, U.K.: Cambridge Univ. Press, 1981.

[46] J. Gottman and A. Roy, *Sequential Analysis: A Guide for Behavioral Researchers*. Cambridge, U.K.: Cambridge Univ. Press, 1990.

[47] C. M. Hoehner, L. K. Brennan Ramirez, M. B. Elliott, S. L. Handy, and R. C. Brownson, "Perceived and objective environmental measures and physical activity among urban adults," *Amer. J. Prev. Med.*, vol. 28, no. 2, pp. 105–116, 2005.

[48] H. C. Hsu and A. Fogel, "Stability and Transitions in Mother-Infant Face-to-Face Communication During the First 6 mos: A Microhistorical Approach," *Devel. Psychol.*, vol. 39, no. 6, p. 1061, 2003.

[49] E. Hutchins, *Cognition in the Wild*. Cambridge, MA, USA: MIT Press, 1995.

[50] E. Hutchins, "Distributed Cognition," *Int. Encyclopedia Soc. Behav. Sci.*, pp. 2068–2072, 2001.

[51] E. Hutchins and C. M. Johnson, "Modeling the emergence of language as an embodied collective cognitive activity," *Topics Cogn. Sci.*, vol. 1, no. 3, pp. 523–546, 2009.

[52] J. Jaffe, D. N. Stern, and J. C. Peery, ""Conversational" coupling of gaze behavior in prelinguistic human development," *J. Psycholinguist Res.*, vol. 2, no. 4, pp. 321–329, 1973.

[53] C. M. Johnson, "Distributed primate cognition: A review," *Animal Cogn.*, vol. 3, no. 4, pp. 167–183, 2001.

[54] C. M. Johnson, "The micro-ethology of social attention: 'Brightness' in bonobos.," *Folia Primatologica*, vol. 75, no. S1, p. 175, 2004.

[55] C. M. Johnson, "Observing Cognitive Complexity in Primates and Cetaceans," *Int. J. Comparative Psychol.*, vol. 23, pp. 587–624, 2010.

[56] C. M. Johnson and M. R. Karin-D Arcy, "Social attention in nonhuman primates: A behavioral review," *Aquatic Mammals*, vol. 32, no. 4, p. 423, 2006.

[57] K. Kaye and A. Fogel, "The temporal structure of face-to-face communication between mothers and infants," *Develop. Psychol.*, vol. 16, no. 5, p. 454, 1980.

[58] J. Kelso, *Dynamic Patterns: The Self Organization of Brain and Behaviour*. Cambridge, MA, USA: MIT Press, 1995.

[59] B. J. King, *The Dynamic Dance: Nonvocal Communication in African Great Apes*. Boston, MA, USA: Harvard Univ. Press, 2004.

[60] T. Leventhal and J. Brooks-Gunn, "The neighborhoods they live in: The effects of neighborhood residence on child and adolescent outcomes," *Psychol. Bulletin*, vol. 126, no. 2, p. 309, 2000.

[61] Y. Liao, S. Makeig, Z. Acar, and G. Deák, "EEG imaging of toddlers during "live" dyadic turn-taking: Mu-Rhythm modulation and source clusters in natural action observation and execution," presented at the Human Brain Mapping, Beijing, China.

[62] G. Littlewort, J. Whitehill, T. Wu, I. Fasel, M. Frank, J. Movellan, and M. Bartlett, "The computer expression recognition toolbox (CERT)," presented at the 2011 IEEE Int. Conf. Automatic Face Gesture Recognit. Workshops (FG 2011), 2011.

[63] S. Makeig, K. Gramann, T.-P. Jung, T. J. Sejnowski, and H. Poizner, "Linking brain, mind and behavior," *Int. J. Psychophysiol.*, vol. 73, no. 2, pp. 95–100, 2009.

[64] D. S. Messinger, M. H. Mahoor, S.-M. Chow, and J. F. Cohn, "Automated measurement of facial expression in infant–mother interaction: A pilot study," *Infancy*, vol. 14, no. 3, pp. 285–305, 2009.

[65] A. Noë, *Action in Perception*. Cambridge, MA, USA: the MIT Press, 2004.

[66] D. O. Olguın, P. A. Gloor, and A. S. Pentland, "Capturing individual and group behavior with wearable sensors," presented at the AAAI Spring Symp. Human Behav. Model., 2009.

[67] H. Oster, D. Hegley, and L. Nagel, "Adult judgments and fine-grained analysis of infant facial expressions: Testing the validity of a priori coding formulas," *Develop. Psychol.*, vol. 28, pp. 1115–1115, 1992.

[68] S. Oyama, *The Ontogeny of Information: Developmental Systems and Evolution*. Durham, NC, USA: Duke Univ. press, 1985/2000.

[69] J. A. Pineda, "The functional significance of mu rhythms: Translating "seeing" and "hearing" into "doing"," *Brain Res. Rev.*, vol. 50, no. 1, pp. 57–68, 2005.

[70] R. L. Repetti, "Short-term and long-term processes linking job stressors to father–child interaction," *Social Develop.*, vol. 3, no. 1, pp. 1–15, 1994.

[71] B. Rogoff and J. Lave, *Everyday Cognition: Its Development in Social Context*. Boston, MA, USA: Harvard Univ. Press, 1984.

[72] R. S. Siegler, *Emerging Minds: The Process of Change in Children's Thinking*. London, U.K.: Oxford Univ. Press, 1996.

[73] L. Smith and E. Thelen, "Development as a dynamic system," *Trends Cogn. Sci.*, vol. 7, no. 8, pp. 343–348, 2003.

[74] L. B. Smith, C. Yu, and A. F. Pereira, "Not your mother's view: The dynamics of toddler visual experience," *Develop. Sci.*, vol. 14, no. 1, pp. 9–17, 2011, 10.1111/j.1467-7687.2009.00947.x.

[75] K. C. Soska, K. E. Adolph, and S. P. Johnson, "Systems in development: Motor skill acquisition facilitates three-dimensional object completion," *Develop. Psychol.*, vol. 46, no. 1, p. 129, 2010.

[76] J. P. Spencer and S. Perone, "Defending qualitative change: The view from dynamical systems theory," *Child Develop.*, vol. 79, no. 6, pp. 1639–1647, 2008.

[77] T. Striano and P. Rochat, "Developmental link between dyadic and triadic social competence in infancy," *Brit. J. Develop. Psychol.*, vol. 17, no. 4, pp. 551–562, 1999.

[78] S. Strum and D. Forster, "Nonmaterial artifacts: A distributed approach to mind," in *Proc. In the mind's eye: Multidisciplinary Approaches to the Evol. of Human Cogn. I: International Monographs in Prehistory*, Ann Arbor,, MI, USA, 2001, pp. 63–82.

[79] S. Strum, D. Forster, and E. Hutchins, "Why Machiavellian intelligence may not be Machiavellian," *Machiavellian Intell. II: Extensions Evaluations*, pp. 50–85, 1997.

[80] L. A. Suchman*, Plans and situated actions: The problem of human-machine communication*. Cambridge, U.K.: Cambridge Univ. Press, 1987.

[81] E. Thelen, "Rhythmical behavior in infancy: An ethological perspective," *Develop. Psychol.*, vol. 17, no. 3, p. 237, 1981.

[82] E. Thelen, D. Corbetta, K. Kamm, J. P. Spencer, K. Schneider, and R. F. Zernicke, "The transition to reaching: Mapping intention and intrinsic dynamics," *Child Develop.*, vol. 64, no. 4, pp. 1058–1098, 1993.

[83] E. Thelen and L. B. Smith*, A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, MA, USA: MIT Press, 1994.

[84] L.L. S. Vygotsky*, Mind in Society: The Development of Higher Psychological Processes.*. Cambridge, MA, USA: Harvard Univ. Press., 1978.

[85] M. K. Weinberg and E. Z. Tronick, "Beyond the face: An empirical study of infant affective configurations of facial, vocal, gestural, and regulatory behaviors," *Child Develop.*, vol. 65, no. 5, pp. 1503–1515, 1994.

[86] P. Wittenburg, H. Brugman, A. Russel, A. Klassmann, and H. Sloetjes, "ELAN: a professional framework for multimodality research," presented at the LREC, .

[87] D. Wood, J. S. Bruner, and G. Ross, "The Role of Tutoring in Problem Solving*," *J. Child Psychol. Psychiatry*, vol. 17, no. 2, pp. 89–100, 1976.

[88] J. J. Wood and R. L. Repetti, "What gets dad involved? A longitudinal study of change in parental child caregiving involvement," *J. Family Psychol.*, vol. 18, no. 1, pp. 237–249, 2004.

[89] C. Yu, L. B. Smith, and A.F. Pereira, ". In grounding word learning in multimodal sensorimotor interaction," in *Proc. 30th Annu. Conf. Cogn. Sci. Soc.*, 2008, pp. 1017–1022.

[90] C. Yu, L. B. Smith, H. Shen, A. F. Pereira, and T. Smith, "Active information selection: Visual attention through the hands," *IEEE Trans. Autonom. Mental Develop.*, vol. 1, no. 2, pp. 141–151, Aug. 2009.

[91] C. Yu, T. G. Smith, S. Hidaka, M. Scheutz, and L. B. Smith, "A data-driven paradigm to understand multimodal communication in human-human and human-robot interaction," *Adv. Intell. Data Anal.*, vol. IX, pp. 232–244, 2010, Springer.

[92] C. Yu, D. Yurovsky, and T. L. Xu, "Visual data mining: An exploratory approach to analyzing temporal patterns of eye movements," *Infancy*, vol. 17, no. 1, pp. 33–60, 2012.

[93] P. Zukow-Goldring and M. A. Arbib, "Affordances, effectivities, and assisted imitation: Caregivers and the directing of attention," *Neurocomputing*, vol. 70, no. 13-15, pp. 2181–2193, 2007.

Kaya de Barbaro received the Ph.D. degree from University of California, San Diego, La Jolla, CA, USA, n 2012.

She is currently a Postdoctoral Fellow in the Cognitive Science Department at the University of California, San Diego, La Jolla, CA, USA. She uses qualitative and quantitative methods to characterize the microdynamics of embodied attention during complex interactions. Additionally, she integrates these approaches with theories and measures of the physiological systems that modulate infants' attention. Her work has been funded by the Temporal Dynamics of Learning Center, National Science Foundation Science of Learning Center.

Christine M. Johnson received the Ph.D. degree in psychology from Cornell University, Ithaca, NY, USA in 1990.

She is currently a Lecturer in the Cognitive Science Department the University of California, San Diego, La Jolla, CA, USA. She joined the department in 1991. Her research focus is in the evolution of social cognition and in taking a comparative approach to its study. She utilizes observational analyses of real-world interactions between socially sophisticated animals, including bonobos, dolphins, elephants, and humans.

Deborah Forster received the Ph.D. degree in cognitive science from University of California, San Diego, La Jolla, CA, USA, in 2012.

She is currently a Project Scientist at the Qualcomm Institute of Calit2 at the University of California, San Diego, (UCSD) La Jolla, CA, USA, and a Researcher in the Machine Perception Laboratory at the Neural Institute for Computation. She trained in Behavioral Ecology and Cognitive Science at UCSD. She spent many years studying wild baboons in Kenya and worked with other primates at the San Diego Zoo. In her research, she tracks behavioral dynamics on multiple spatio–temporal scales, in multiagent settings. She applies dynamical systems principles framed in state-space and timeseries analyses that are prevalent in other computational and machine learning approaches in the cognitive sciences.

Gedeon O. Deák received the Ph.D. degree in child psychology from the University of Minnesota (Twin Cities), St. Paul, MN, USA, in 1995.

He was an Assistant Professor at Vanderbilt University from 1995 to 1999. Since 1999, he has been in the Department of Cognitive Science and the Human Development Program at the University of California, San Diego, La Jolla, CA, USA, where he is a Full Professor. He has been a Visiting Professor at Southwest University, Chongqing, China. He is author or coauthor of over 50 articles in peer-reviewed journals and proceedings, and 10 book chapters and encyclopedia entries.

Dr. Deák is a Member of the IEEE Technical Committee on Autonomous Mental Development and an Associate Editor of the IEEE TRANSACTIONS ON AUTONOMOUS MENTAL DEVELOPMENT. He has been a Spencer Foundation/National Academy of Education Postdoctoral Fellow and a Hellman Fellow. He has received research grants from the National Science Foundation (USA), the National Institute of Child Health and Development, the Kavli Institute for Mind and Brain, and the MIND Institute.