HHS Public Access

Author manuscript

Infant Behav Dev. Author manuscript; available in PMC 2021 June 03.

Published in final edited form as:

Infant Behav Dev. 2021 May ; 63: 101566. doi:10.1016/j.infbeh.2021.101566.

Revisiting how we operationalize joint attention

Allison Gabouer, **Heather Bortfeld***

Psychological Sciences, University of California, Merced, United States

Abstract

Parent-child interactions support the development of a wide range of socio-cognitive abilities in young children. As infants become increasingly mobile, the nature of these interactions change from person-oriented to object-oriented, with the latter relying on children's emerging ability to engage in joint attention. Joint attention is acknowledged to be a foundational ability in early child development, broadly speaking, yet its operationalization has varied substantially over the course of several decades of developmental research devoted to its characterization. Here, we outline two broad research perspectives—social and associative accounts—on what constitutes joint attention. Differences center on the criteria for what qualifies as joint attention and regarding the hypothesized developmental mechanisms that underlie the ability. After providing a theoretical overview, we introduce a joint attention coding scheme that we have developed iteratively based on careful reading of the literature and our own data coding experiences. This coding scheme provides objective guidelines for characterizing multimodal parent-child interactions. The need for such guidelines is acute given the widespread use of this and other developmental measures to assess atypically developing populations. We conclude with a call for open discussion about the need for researchers to include a clear description of what qualifies as joint attention in publications pertaining to joint attention, as well as details about their coding. We provide instructions for using our coding scheme in the service of starting such a discussion.

Keywords

Joint attention; Coding scheme; Parent-child interactions; Multimodal communication; Infant social cognition

1. Introduction

Jean Piaget originally suggested that young children's egocentrism prevents them from considering the perspectives of others (Piaget, 1952, 1954). Decades later, Michael Scaife and Jerome Bruner expanded on then-current theories of infant egocentrism (e.g., Butterworth, 1987) by demonstrating that infants can use the direction of another's gaze to purposefully redirect their own gaze (Scaife & Brunner, 1975). Since then, efforts to identify the mechanisms that underlie attention sharing—*joint attention*—have bridged the social and

*Corresponding author at: 5200 Lake Road, Merced, CA 95340, United States. hbortfeld@ucmerced.edu (H. Bortfeld).

Author statement

Allison Gabouer: Conceptualization, Methodology, Writing- Original draft preparation, Visualization. **Heather Bortfeld:** Conceptualization, Methodology, Supervision, Writing- Reviewing and Editing, Funding Acquisition.

perceptual aspects of the processing involved (Siposova & Carpenter, 2019; Stephenson et al., 2021). Here, we characterize two broad perspectives on what joint attention is with the aim of identifying the key operationalization discrepancies that have contributed to confusion in the field. We then outline our own approach to coding joint attention, one that we have developed to flexibly accommodate different aspects of dyadic interactions, including those that vary by children's age and developmental status.

To partake in a shared experience, one member of a dyad attempts to direct the focus of the other. Although this shared focus of attention is now commonly referred to as joint attention, Bakeman and Adamson (1984) used the term "coordinated joint attention" in their seminal paper in which they documented the active coordination of attention between mother-infant and infant-peer dyads and an object of mutual interest. This form of shared attention is considered by many to be the most complex form of dyadic interaction between parents and young children (Bakeman & Adamson, 1984; Siposova & Carpenter, 2019). Several terms have since been used to describe the form of interaction in which a dyad shares attention to an object: joint visual attention (Butterworth & Jarrett, 1991; Scaife & Brunner, 1975), coordinated joint attention (Bakeman & Adamson, 1984), triadic attention (de Barbaro et al., 2016d; Striano & Stahl, 2005), shared attention (Deák et al., 2017; Siposova & Carpenter, 2019), coordinated visual attention (Yu & Smith, 2017a), and coordinated attention (Chen et al., 2019), among others. At present, the most commonly used term, one that's become a catch-all for the concept of two individuals orienting together towards an object or event, is joint attention. The variable terminology and operationalization used by different researchers has resulted in confusion about what exactly constitutes joint attention, both within and outside of developmental research, as well as adjacent forms of attention (i.e., sustained attention). Clarifying the terminology is important because the term *joint attention* has come to have diagnostic implications. For example, as the underpinnings of Autism Spectrum Disorders (ASD) have come into focus, clinicians test abilities like joint attention to help identify children with ASD-specific deficits (Kasari et al., 2012).

1.1. Origins of 'joint attention'

In ontogenetic terms, the ability to engage in joint attention emerges early—at around 9 months of age (Mundy et al., 2007)—and lays the groundwork for subsequent developmental advances. By now there is substantial evidence of relationships between joint attention and other abilities, including language learning (Mundy & Gomes, 1998; Salo et al., 2018; Tomasello & Farrar, 1986), consolidation of social cognition (Mundy & Gomes, 1998), increased visual attention to social partners (Striano & Stahl, 2005), increased pedagogical success (Mundy & Newell, 2007), and mastery of cultural conventions (Bruner, 1974). Given that human development relies enormously on shared experiences and knowledge (MacPherson & Moore, 2017), the fact that so many abilities can be related back to a child's early joint attention skills should not come as a surprise. Indeed, a child's ability to engage in joint attention is now recognized as a strong, predictive indicator of typical development, social and otherwise.

At its inception, the sharing of attention, or "joint visual attention," was described and operationalized simply as gaze following. In their seminal study, Scaife and Bruner (1975)

examined whether and how infants capitalize on the visual attention of an adult in order to locate an object in the immediate environment. These researchers observed that infants could successfully follow an adult's 45° head turn—in either direction—to search for the focus of the adult's interest. Years later, Roger Bakeman and Lauren Adamson (1984) began documenting the gradual developmental progression of infants' ability to coordinate their attention with the attention of others, whether caregivers or peers. In the service of clinical applications, as well as in support of a general interest in how the dyadic partner plays a role in child development, Bakeman and Adamson (1984) defined six states of child engagement, delineated based on free-play sessions during which 6- to 15-month-olds interacted with their mother or a peer. Engagement states ranged from completely unengaged to what the researchers then referred to as "coordinated joint attention," a state of engagement deemed to be the most advanced of the six. Coordinated joint attention was thus introduced to the field as the *active coordination of attention between two people in a dyad and the object one of the two people is focused on or involved with.* In other words, Bakeman and Adamson (1984) operationalized joint attention as the *active* coordination between two people about an object or event of interest.

One interpretation of this *active coordination* terminology is that it centers operationalization of joint attention on the social dimension of dyadic interaction, meaning that joint attention is not passively achieved. This perspective contrasts with a perspective that includes incidental engagement of attention, during which an infant might follow another's gaze without it being the intention of the gazer for his or her gaze to be followed. Indeed, active versus passive coordination between two people towards an object of interest proves to be a crucial distinguishing feature across the different accounts of joint attention, one that underscores disagreements over how joint attention supports infant learning, including language learning, more broadly. This subtle differentiation—between active and passive attention—is also the source of substantial confusion both within and outside of developmental research about what joint attention is (see Emery, 2000, Tomasello et al., 2005 for further discussion). While there would certainly be a benefit to some overall semantic agreement (i.e., differentiation of accounts based on the terminology used), our goal in this overview is not to be dictatorial about how the term "joint attention" is used. Rather, we aim to encourage researchers to be clear in what they mean when they call something joint attention—including how it is identified, the requirements for inclusion in that identification, and ideally consideration of how these requirements impact the research findings themselves.

1.2. Perspectives on joint attention

Once it became clear that infants and young children engage in gaze following with adults (Scaife & Bruner, 1975), an onslaught of research—and substantial debate—ensued, much of it focused on when different forms of attentional ability emerge in infancy and early childhood. By now, researchers' assumptions about how joint attention manifests across developmental time reflect their theoretical biases, as does their particular choice of terminology and how they operationalize what they're looking for behaviorally (Racine, 2013). In the process, two distinguishable accounts have emerged, one that characterizes joint attention as fundamentally social, and the other that frames it as associative in nature

(but see (Adamson, Bakeman, Suma, & Robins, 2019) for an alternative framework). In the interest of transparency, our own approach is very much guided by a social perspective. Despite this, we appreciate the implications of various findings from research guided by the associative perspective and include a review of a subset of them here.

1.2.1. Social accounts—As with many complex behavioral constructs, precise operationalization of joint attention has come about gradually. In social accounts, joint attention depends on a triadic interaction that relies on children's contingent interactions with caregivers (Striano & Stahl, 2005). Indeed, Bakeman and Adamson (1984) argued that coordinated joint attention is the most complex form of dyadic interaction between parents and young children (Bakeman & Adamson, 1984; Siposova & Carpenter, 2019). Thus, the social account is founded on the view that joint attention and its corresponding behavioral markers are grounded in social cognition. This operationalization translates behaviorally to careful documentation of back-and-forth social interactions that require social awareness on the part of at least one member of the dyad, with achievement of joint attention also requiring some form of verification of attentional allocation.

In particular, Michael Tomasello has been a strong proponent of joint attention as an active process. In his seminal study on the relationship between joint attention and children's lexical development (Tomasello & Farrar, 1986), he introduced a set of objective guidelines to document joint attention in parent-child interactions. Critically, these criteria included an emphasis on identifying active coordination of understanding between two people. To this end, Tomasello and Farrar (1986) applied a novel coding scheme to parent-child free-play sessions, using a plus (+) and minus (−) system to identify episodes of joint attention. To qualify, an interaction had to contain (a) one member of a dyad initiating engagement with the other, (b) both members of the dyad focusing on the same object for three or more seconds, and (c) the member of the dyad that initiated the interaction showing clear evidence of awareness of the dyadic partner being successfully engaged (i.e., a verifying look towards the other). Only when an interaction received a plus for all three features did it meet the criteria for joint attention. We should note that these behaviors are based in whatever unfolds naturally between a parent and child during a free-play interaction and is in no way guided by the researchers themselves, who are instructed only to play as they normally do. Thus, task demand is not an issue in how parents or children behave.

One critique of such approaches is that components of protocols built to capture these social accounts of joint engagement rely on subjective assessments of the interaction as opposed to the objective, skill-based measures used in assessments prompted by associative accounts. However, concerns about subjectivity should not exclude the application of a more qualitative lens in assessing parent-child interactions; the solution is clear operationalization and a set of objectively based coding criteria. In fact, several implementations of naturalistic coding have been informed by objective measures of joint attention. We return to this point in Section 3, where we suggest how to vary the timing and descriptive components of our coding protocol to fit the needs of the interaction being assessed and the developmental skill-level of the participants involved.

1.2.2. Associative accounts—The reasoning behind the associative perspective is that infants need not demonstrate interpersonal awareness to achieve a state of joint attention with another person (Corkum & Moore, 1998). Rather, joint attention here is considered something largely dependent on infants' visual orienting system, and thus researchers in this vein argue that they do not need to appeal to any social processes (i.e., perspective taking, social referencing) in how such orienting might take place. This non-social perspective reflects ideas about how infants acquire and coordinate their own expectations of how adults behave physically around interesting objects, and thus does not require researchers to postulate any sort of psychological relationship between the adult, the infant, and the object of interest (Moore & Corkum, 1994).

In an example of this perspective, Butterworth and colleagues (Butterworth & Jarrett, 1991) defined joint attention as gaze following (i.e., as happens when a child looks where someone else is looking). Such a definition is inherently non-social, meaning that co-occurrence of gaze is the only criterion for the achievement of joint attention. Rather, these researchers proposed that joint attention is the product of a "geometric mechanism" that enables infants to attend to the same thing as another person, meaning there is no need to invoke any sort of mental-state reasoning on the part of the infant. According to this geometric model, infants are able to use the direction of an adult's head-turn to infer the possibility of an interesting object in that direction. The logic goes that once the infant engages in a correct head-turn, the object is salient enough to attract the infant's attention (e.g., a light bright enough to be noticed). Thus, on this view, rather than being a social process, joint attention relies on an infant's sensitivity to geometric orientation of another's head-turn and on the attention-catching properties of the referent itself, with no need to postulate whether communicative intent underlies an initiator's behavior.

One difficulty for such accounts is the variability observed across parent-child dyads in so-called checking behaviors. This term is used by researchers (e.g., Baldwin, 1991) to refer to the quick look back made by an initiator of joint attention to the dyadic partner, arguably suggesting intentionality in their attempt to engage the other in joint attention. Where such behavior is taken by some researchers as evidence that the initiator is invested in the responsivity of the other dyadic partner to attempts at attentional engagement, a view that is consistent with the social account, others (i.e., Moore & Corkum, 1994) argue that such checking behavior is the result of the completion of an engagement episode, whereby the child's gaze returns to the adult's face with the goal of finding new cues about other visually-engaging referents. Interestingly, these accounts predict different outcomes in terms of who is doing the checking back. Suffice it to say that whether or not the mechanism(s) underlying the checking-back behavior is grounded in any sort of mental state reasoning is itself a topic of debate.

Critical to the points we aim to address here, associative accounts do not dictate that there be intentionality on the part of either member of a dyad for joint attention to occur. Thus, researchers who take this view do not include documentation of verification by one dyadic partner that the other's attentional allocation has been adequately directed to the object of interest (Butterworth & Cochran, 1980; Scaife & Brunner, 1975). Regardless of perspective, the terminology often used to refer to such verification ("checking back") couches

identification of joint attention in the visual modality alone. On this view, other sensory information, whether provided by the infant, the dyadic partner, or by features of the object of mutual interest itself (i.e., rattling noise produced by parent shaking a toy) would not be included in coding of intentionality. This perspective has changed in recent years, as reflected by findings—including our own—that highlight the impact of multimodal input on infant attentional allocation (e.g., Depowski, Abaya, Oghalai, & Bortfeld, 2015; Suarez-Rivera et al., 2019).

1.3. The present approach

This overview should be taken as evidence that the nature of and basis for joint attention will continue to be the topic of debate for the foreseeable future. Rather than engage on it, we introduce our perspective here as a means of accounting for the decisions we have taken in our approach to coding protocol. The protocol is guided by the view that emergence of joint attention is attributable to experience-based advances in social cognition that drive infants' attentional allocation. These advances can include both endogenous factors, principally neurocognitive development itself (Mundy et al., 2003), and exogenous factors, such as explicit integration of experiences with goal-related behavior (Tomasello et al., 2005). In short, our perspective is that the foundation for engagement in joint attention is fundamentally socio-cognitive.

An integrated mechanism that many consider critical to the process of sharing attention was provided by Michael Posner and Mary Rothbart (2007), who identified a neural circuit for attention that includes a posterior attentional system responsible for representation, imitation, and perception in relation to others, and an anterior system responsible for intentional, goal-directed attentional focus. An important and still-outstanding question is how this attentional circuit emerges in the first place. One view from the clinical domain (i.e., Dube et al., 2004) is that consecutive behavioral processes early in development collectively build such that a child acquires the ability to shift interest between a toy and an adult in order to "share the experience" with that person. Such sharing of attention is considered a different class of behavior from gaze shifting.

Precursor behaviors are important in their own right. A recent proposal inspired by dynamic systems theory (Thelen & Smith, 1996) envisions individual growth in attentional skills as a fundamentally social process influenced by both intrinsic and evoked activity whose inputs collectively produce the variable outcomes in children's ability to regulate attention (Yu & Smith, 2016). Another recent proposal focuses on the underpinnings of shared attention, arguing that such states can occur either intentionally or incidentally, but necessarily result in an exchange of information about the environment and the mental states of the parties involved (Stephenson et al., 2021). Our own efforts focus on what happens in parent-child dyads as children become developmentally able to actively engage with objects together with other people, who themselves may adjust their behavior to better support children's attentional focus (Dube et al., 2004). Given this, we do not take a position on the origins of joint attention; however, our approach requires inclusion of attentional verification as a criterion for an interaction to qualify as joint attention. Critically, rather than rely on terminology that prioritizes vision over all other modalities, we adopt the position that any

indicator (i.e., a look, a touch, a nod, a verbal affirmation) that the parent is attuned to the child's attentional allocation qualifies as verification.

1.4. Protocols for assessing joint attention

Several protocols are available to guide assessment of children's engagement in joint attention. Many of these involve an experimenter actively engaged with the child, whose interaction is coded in real-time or post-hoc to characterize children's unprompted interaction with a caregiver in anywhere from a natural to a semi-structured environment. A widely used protocol in the clinical domain is the Early Social Communication Scales (ESCS; Mundy et al., 1996), a structured observation scored in real time by a trained experimenter. The ESCS elicits three specific, quantifiable social-communicative behaviors: joint attention behaviors, requesting behaviors, and social interaction behaviors, while using eight different tasks to assess the target behaviors, including turn-taking, gaze following, book reading, and requesting. Children are scored based on their portrayal of low- and high-level behaviors, where high level behaviors and correct responses result in higher scores. Overall scores are used to produce a social communication profile for the child.

Researchers have also developed protocols to identify and describe engagement in joint attention from the perspective of an onlooker. In these instances, dyads engage in a free-play session with an adult, uninterrupted by prompting or technology. These interactions are typically recorded and coded offline, incorporating specific criteria for characterizing different components of the interaction, including bouts of joint attention (Bakeman & Adamson, 1984; de Barbaro et al., 2016d; Gale & Schick, 2009; Nowakowski et al., 2009; Salo et al., 2018; Tomasello & Farrar, 1986). The different protocols consist of components intended to provide guidance on codable criteria, with agreement between coders assessed via interrater reliability. Not surprisingly given all the issues we have outlined here, these coding protocols can differ substantially on what constitutes joint attention, underscoring the need to better characterize the action of interest rather than simply labeling it. In our efforts to generate a protocol that supports the most objective coding possible, we have found that the various approaches each provides important and unique insights.

Although we focus on video-based manual coding here, we would be remiss if we did not acknowledge the impact that technological advances (i.e., head-mounted eye-tracking) have had on the field by allowing increasingly precise documentation of the focus of dyadic partners' gaze throughout the course of an interaction. Many researchers are now using eye-tracking to establish the statistical properties of the interactions. Because eye-tracking data can be parsed in various ways using machine-based coding, researchers are not only able to identify overlaps in visual gaze, but also relate one or both participants' looking behavior to other events. For example, overlaps between parent and child gaze has been shown to be predictive of vocabulary development (Abney et al., 2017), to relate to hand-eye coordination (Yu & Smith, 2017a) and result from what children see their parents touch (Deák et al., 2014). Additionally, eye-tracking patterns are being used to inform changes in parent behavior that can improve joint attention and child learning during dyadic reading activities (Guo & Feng, 2013). We have also benefited from the insights that these

approaches provide, particularly around issues having to do with the application of time windows for classifying different behaviors.

1.5. Development of coding protocol

Here we introduce a coding protocol that integrates components of the different approaches to joint attention that we have outlined. In particular, the protocol is inspired by the work of Tomasello and Farrer (1986) and of Nowakowski et al. (2009), with timing details founded on insights from both visual observation (Bakeman & Adamson, 1984) and eye-tracking (Abney et al., 2017; Yu & Smith, 2013). Critically, our approach allows coding of both the *initiation* of joint attention (by both/either parent and child), and *maintenance* of joint attention (again, by both members of the dyad). We have found that it is critical to distinguish between the initiation and maintenance of joint attention to accurately characterize the interaction from moment-to-moment. Another important aspect of our approach is that we incorporate the potential to identify the range of sensory cues produced across different modalities by both members of a dyad, which can be tracked alone and in combination. We have added the multimodality dimension of our coding in light of findings from both social (Baron-Cohen, 1991) and associative (Moore and Corkum, 1994) accounts. More recently, our approach is proving to be consistent with results obtained using eye-tracking (Yu & Smith, 2017a, 2017b), as well as those from observational coding (de Barbaro et al., 2016d), whose findings that infants are particularly responsive to their dyadic partners' hands is something we documented early on in our own coding (Bortfeld & Oghalai, 2020; Depowski et al., 2015; Gabouer, Oghalai, & Bortfeld, 2018, 2020). Starting with our initial observation of the importance of the caregiver's hands, we have iteratively fine-tuned our protocol with each data set to better characterize this phenomenon.

1.5.1. Basis for coding protocol—Our joint attention coding criteria are based on active engagement with hundreds of hours of parent-child interaction videos (Bortfeld & Oghalai, 2020; Depowski et al., 2015; Gabouer et al., 2018, 2020), as well as careful reading of others' research. Our focus has been on how to characterize the factors that lead to joint attention and to do so systematically, with an eye towards establishing a mechanistic account of this phenomenon. Our interest originated in efforts to establish guidelines to support joint attention development in children who are deaf or hard-of-hearing and who do not have access to consistent sign language input, a situation that is quite common among deaf children of hearing parents who are candidates for cochlear implantation but have yet to be implanted. These are among the children for whom the establishment of joint attention can provide a critical scaffold to learning about communicative intent prior to their exposure to consistent, structured language input.

The guiding question behind our approach has been whether and how hearing parents can establish joint attention with their deaf children when the typical manner by which this is achieved is (i.e., parental vocalization) is not available to the children. It was in pursuit of answers to this question that we characterized parental behaviors that lead to joint attention in hearing parent-deaf child dyads as well, to compare with behaviors in hearing parent-hearing child dyads. While trying to apply others' coding guidelines to our own parent-child interaction videos, it became clear that we needed to develop a step-by-step guide for

identifying how the two components of joint attention—initiation and maintenance—proceed for both members of a dyad. Moreover, as we developed and applied our own increasingly stringent coding criteria, it was clear that members of our control dyads (i.e., hearing parents of hearing children) were doing things to achieve and maintain joint attention with their children that had not been mentioned in any of the joint attention research reports we were reading.

Originally, we predicted that hearing parents would rely on auditory cues to initiate joint attention with their children, regardless of the children's hearing status, due to their personal familiarity and comfort operating in the auditory modality. We also expected that hearing parents of deaf children would achieve less child engagement in joint attention overall, due to the child's hearing status and the parent's inability to sign. Quite to the contrary, however, we observed that hearing parents, whether of deaf or hearing children, were equally effective in establishing joint attention with their children, and that they did so using a range of different cuing techniques that spanned multiple modalities.

We have now finalized a revised set of criteria (see Fig. 1). First, an initiating partner attempts to direct the other person's attention to a nearby object or shared experience. This initiation can happen through a variety of sensory cues—auditory, visual, or tactile—as well as any combination of the three. These actions must be purposeful and intentional, conveying an overt attempt to engage. Following the initial action from the initiating partner, the non-initiating partner must jointly attend to the object of mutual interest. This "following in" to the initiation cue can also be accomplished through multiple actions such as a gaze-shift, acting on the object, or engaging in verbal conversation regarding the object. The third feature—demonstration of awareness—is done by the initiating partner as a way to ensure the non-initiating partner has noticed their initiation attempt and effectively integrated into the triadic interaction. We have found that researchers often discount this checking behavior, characterizing it as something that happens outside of the joint attention interaction. This oversight is likely the result of the theoretical disagreement over what joint attention is as opposed how the interaction itself unfolds.

1.5.2. Implementation of coding protocol—In previous sections, we identified sources of disagreement about what is an is not necessary for a dyad to be considered as engaged in joint attention. Indeed, we now recognize that the terminology different researchers use to characterize different attentional states, including joint attention itself, is highly variable and often contradictory. Consistent with our intuitions, a recent paper (Siposova & Carpenter, 2019) calls attention to this problem and attempts to characterize various states of attention in dyadic interactions in light of different degrees of common knowledge between dyadic partners. Thus, there is a critical need for an open-ended discussion about discrepancies—both theoretical and mechanistic—in joint attention research. To this end, we share our systematic coding approach. Our micro-coding scheme is grounded in the following definition of joint attention: the active and intentional engagement between two people regarding an object of mutual interest. In outlining our approach, we address discrepancies that we have identified in others' approaches and provide reasoned resolutions that can be objectively applied in future research.

2. Proposed coding scheme

Our scheme consists of three component steps that involve identifying an attempt at initiation of joint attention and determining whether the attempt was successful or failed. Here, we use the term “bid” to describe such attempts, which are purposeful actions on the part of the initiator with the intent of directing the target’s attention to an object of interest. Each step of the identification process is represented in a decision tree (see Fig. 1). Although our previous work specifically focused on parent-initiated joint attention, the coding scheme can be used to identify instances of both parent- and child-initiated joint attention (see Section 3 for suggestions about how to use the protocol).

2.1. Intention

The initial step in identifying joint attention is determining whether a bid has taken place or not. The bid must be intentional and non-random, where intention is defined as events which the coders perceived as non-accidental and which the initiator acted purposefully to share attention with the target. Intention was largely gauged using the following indicators: (a) ostensive visual focus on the object of interest, (b) physical orientation toward the object of interest, (c) haptic interaction with the object of interest, or (d) an overt gesture toward the object of interest (Trueswell et al., 2016). Specifically, accidental actions (e.g., sneezing, grazing, tripping over a toy, etc.) are not motivated by intention to engage in joint attention, and thus do not meet the standard of a bid in the current coding protocol. Perceived intention additionally hinges on a third piece of criteria presented in the decision tree – active verification. Active verification (see 2.3) is required on the part of initiator as it indicates concern as to the outcome of the bid, suggesting an intentional nature.

The requirement for a bid to be intentional is in contrast to most research conducted using eye-tracking in which data are classified via machine learning techniques (Guo & Feng, 2013; Yu & Smith, 2017b). These approaches generally include all instances in which the parent and child both attend visually to the same object, whether intentionally or not. Yu, Smith, and colleagues have demonstrated that these instances of coordinated looking are important for predicting vocabulary development (Abney et al., 2017) and the engagement in sustained attention (Suarez-Rivera et al., 2019), among other things. We acknowledge that learning can occur in these non-intentional situations (see Yu & Smith, 2013), but the lack of a verification component in these approaches contrasts with the “coordinated joint attention” described originally by Bakeman and Adamson (1984). The implications of these intentional episodes, in contrast to incidental engagement, is an open question that is worthy of investigation. We return to this point in Section 3 regarding additional applications for the outlined protocol.

To engage in joint attention, one member of the dyad must have the intention to share attention with the other. Thus, intention is the first step in Tomasello and Farrar’s coding (1986), a critical component of Racine’s (2013) definition of joint attention, and the first decision in our proposed coding scheme (see Fig. 1). Without intention, joint attention can happen by accident or just through coincidence when members of a dyad happen to focus their gaze or place their hands on the same thing at the same time. Without intention, what is being called joint attention is really just the result of happenstance, rendering what leads to it

no longer interesting. In our approach, we emphasize the following question: How do parents direct their child's attention in a purposeful way?

2.2. Response

2.2.1. Engaged response—Our two main criteria for characterizing the response to a bid are the type and duration of the response (see Fig. 2). To this end, we employ three different rules of engagement. The first is as follows. Once the initiator has finished an initial bid for attention, the target has a five second window of time in which to respond (Fig. 2, yellow bar; Carpenter et al., 1998; Tomasello & Farrar, 1986), thus resulting in a successful bid. There are various actions the target can perform that we consider indicative of a successful bid. The non-initiating, or target, member can respond to a bid by pointing, gaze following, tapping or touching the initiator, engaging with the object of mutual interest, deliberately gesturing within the initiator's visual field, changing affective demeanor, and/or producing language. The application of the three-second rule requires the target engage in one or several of the above responses for at least three seconds within five seconds (Fig. 2, yellow bar; Bakeman & Adamson, 1984; Nowakowski et al., 2009). This three second engaged response is often referred to as sustained attention in the eye-tracking literature (see Suarez-Rivera et al., 2019; Yu & Smith, 2016), and is suggested to have different implications, compared to shorter bouts of engagement in joint attention.

Once a target has done this—a process often referred to as "integrating" with an object and dyadic partner—the target can fluctuate between various states of engagement or disengagement (Fig. 2, green bar; Abney et al., 2017), provided that any disengagement does not exceed five seconds. The third timing component focuses on the timing for the initiator to actively acknowledge the engagement state of the target (Fig. 2, blue bar). The current protocol employs a five-second window of opportunity starting from the time the target responds to the bid.

2.2.2. Failed bids—We have yet to find any research on what characterizes bids that do *not* result in successful engagement in joint attention (Fig. 1). In an effort to differentiate and compare successful versus failed attempts to establish joint attention, the proposed coding scheme provides the option to include codes for failed bids. Here, a failed bid is any intentional bid (as described in 2.1) that does not result in successful engagement on the part of the target. Using the five-second response window (Fig. 2, yellow bar), we can identify successful bids, as described above, as well as failed bids. When a target fails to attend/integrate with the dyadic partner within the five second window, this qualifies the initiator's bid as unsuccessful. For example, a parent may use gaze as a bid to initiate joint attention with a child. Often a parent will follow such a bid with labeling. However, if the child is preoccupied directing their gaze in another direction or engaging with a different object, the child may miss this visual bid. Likewise, a parent may be consistent in cue timing, but the cues themselves may not work to guide the child's attention. If a parent is insensitive to this fact, the opportunity will be missed for the parent to adapt bid strategies and thus accommodate a child's specific sensitivities. We realize this perspective may be controversial, but this highlights how critical it is to agree on the importance of the verification behavior in joint attention. We encourage researchers to use this coding scheme

to investigate which intentional actions result in successful and failed bids, how one can "repair" a failed bid, and what role failed bids play in the long-term progression of dyadic interactions. We also encourage investigation into the prevalence and impact of verification behaviors more generally.

2.3. Active verification

As we have argued, in cases in which a target is receptive to an initiator's bid, the initiator must confirm the target's focus of attention by showing some form of verification of the change attentional allocation (Carpenter et al., 1998; Tomasello & Farrar, 1986). Because dyads use several sensory modalities during interactions, several behavioral responses on the part of the initiator can qualify as verification. For example, a visual gaze change to the target to gauge reception to the initiation act, a vocalization from the target that is heard and responded to by initiator, or a manual/tactile action that is seen or felt by the initiator (i.e., a visual gaze change to the target's hand). Such *active coordination* between the dyad and with object or event is the essence of social interaction and is quite distinct from so-called coordinated looking to the same object, which may or may not be intentionally engaged. In other words, verification provides a clear indicator that the bid was intentional.

Ours is not the first joint attention protocol to employ a verification component. For example, Bayliss et al. (2013) referred to return-to-face saccades, in which one partner reorients to the other when sharing attention. These researchers characterized this as a form of social feedback that the initiator of joint attention uses to verify the outcome of his or her behavior. The requirement of such verification is also documented in research with children who are deaf or hard-of-hearing (Nowakowski et al., 2009; Prezbindowski et al., 1998) Finally, Striano and Stahl (2005) argued that previous assessments of joint attention were lacking the "monitoring component" (their term for verification of a bid's success or failure).

This criterion is relevant to both the parent and the child as initiator too, a view that is consistent with other domains of research, such as that employing the Still-Face Paradigm (Cohn & Tronick, 1983), in which infants demonstrate themselves to be sensitive to relevant social cues in a triadic interaction. Even very young infants (between 3 and 9 months of age) will spend a significant amount of time looking toward an adult when the adult coordinates both her affect and attention between the infant and an object, as opposed to simply coordinating affect or attention only with the infant (Striano & Stahl, 2005). Specifically, this pattern of gaze behavior (actively switching between the infant and the object of mutual interest) on the part of the adult results in the infant spending more time looking toward the adult. Overall, this suggests infants' exceptional sensitivity to relevant socio-communicative acts performed by the parent. Interestingly, these researchers (Striano & Stahl, 2005) suggested that converging, multimodal cues may also influence the engagement in joint attention, a view that has guided the focus of our own research (Bortfeld & Oghalai, 2020; Depowski et al., 2015; Gabouer et al., 2018, 2020), and is a topic that we will return to below.

Overall, we view joint attention through a socio-cognitive lens. In that spirit, we consider the initiator's confirmation of a target's reaction to an object a critical component of joint

attention. For example, parents' use this reaction in deciding how to further the interaction to support learning; this checking behavior highlights the active and triadic dimensions of joint attention, in which monitoring of the infant's psychological relation to the object is critical (Campos & Stenberg, 1981). Without this acknowledgment, it is difficult to determine the intentionality of bids in the first place. This prompts a different yet related question regarding the implications of the initiator's sensitivity to the bid's success or failure. Currently, there is little evidence regarding such measures of sensitivity.

2.3.1. Not coded—Given the restrictions we have outlined, there are interactions that are not coded as bids for joint attention, which may seem very intentional in all other respects. If an initiator does not verify engagement, we do not code it as a bid (neither successful nor failed) for initiation of joint attention. This category differs from the failed bid categories in that a failed bid is the result of the lack of the target's integration but does include checking-back. If the initiator fails to ensure that the target followed the bid and is also attending to the new object, then the bid is not coded, regardless of the target's behavior. Future research will need to pursue a means of comparing outcomes of non-coded instances with coded instances, as it is an open question whether they influence children's communicative development in a manner similar to those that are verified.

2.4. Tracking multimodal cues

In an effort to further characterize successful and failed bids for joint attention, we also examine how different sensory modalities are used in these attempts, both alone and in various combinations (i.e., (Bortfeld & Oghalai, 2020; Depowski et al., 2015; Gabouer et al., 2018, 2020). By coding and quantifying an initiator's use of multimodal cues, we can further inform a social account of joint attention and move away from a strictly visual interpretation and mechanism. Deák and colleagues (Deák et al., 2017) employed a microcoding scheme to investigate how parent-initiated joint attention is supported by gaze and manual actions. The researchers found that, across months' worth of interactions, development of joint attention in parent-child dyads is the result of co-modulation of behaviors. From this, they argued that joint attention is complex, interactive, and is supported by the maturation of the child's sensorimotor networks, which afford engagement in multimodal communication. Not surprisingly, microcoding parent-child interactions has become increasingly popular in joint attention research and is leading to new hypotheses and theoretical directions.

As an extension of the coding scheme outlined above, we also provide guidelines here for identifying the sensory components of bids—including auditory, visual, and tactile cues. Multimodal cues are a powerful source of information for newborns and young infants in that auditory cues commonly result in visual attention (Kaplan & Werner, 1991; Mendelson et al., 1976). More recent studies support the general idea that multimodal information supports vocabulary development (Trueswell et al., 2016), establishment of category labels (Clark & Estigarribia, 2011), sustained attention (Suarez-Rivera et al., 2019), and joint attention (Gabouer et al., 2018, 2020). By interrogating whether bids that consist of one sensory modality or various combinations of sensory modalities result in more or less joint attention, we can expand our understanding of infant development in general, and the influence of different interaction styles in particular.

2.5. Coding scheme in practice

In the research described above, our lab employed a coding template built using EUDICO Linguistic Annotator (ELAN). ELAN is a custom language annotation software program created by the Max Planck Institute for Psycholinguistics (The Language Archive, Nijmegen, The Netherlands). ELAN allows for multimodal analyses of language and other behaviors (Wittenburg et al., 2006) and is available free of charge (http://tla.mpi.nl/tools/tla-tools/elan/). The template was built to allow for transcription of any auditory information, as well as a dependent "layer" to identify the modality information. Each of these layers is called a tier. The template is built by starting with two tiers labeled "Parent Initiated Joint Attention" and "Child Initiated Joint Attention." Then a new "Controlled Vocabulary" is added. A controlled vocabulary creates a forced-choice dropdown list of all the bid outcome (e.g., successful, failed, no active verification, incidental) and modality options (e.g., auditory-object noise, visual-tactile) that can be selected to describe an interaction. The controlled vocabulary is the content of a "Linguistic Type", which specifies the parameters of the controlled vocabulary. The linguistic types, which were label as "Bid Outcome" and "Modality", can be added to a tier to provide the dropdown list when a certain tier is selected and annotated. After the creation of the controlled vocabulary and using the controlled vocabulary to specify the linguistic type, you can then apply the linguistic type to tiers. Two new dependent tiers are added to attach the instance of bid outcome to the modality used by the initiator—these tiers are labeled "Parent-Initiated Modality" and "Child-Initiated Modality". These tiers are then associated with the linguistic type "Modality", which will prompt the dropdown of the modality or combinations of modalities when this tier is selected. Additionally, these tiers are assigned a "parent tier" with creates a hierarchical structure in the template.

3. What is joint attention?

If our goals are to identify the mechanisms underlying the development of joint attention abilities, as well as to understand how parents can better support children's attentional development, researchers must first agree on what joint attention is, or at the very least clearly define the construct of interest. Given the developmental implications of successful engagement in joint attention, the phenomenon has become an important developmental milestone. However, it is difficult to measure such a milestone when its behavioral manifestation is characterized in so many different way. We hope we have made it clear that what qualifies as joint attention in parent-child interactions has myriad forms (Siposova & Carpenter, 2019). The operationalization of joint attention will likely continue to be informed by a mechanistic understanding of what supports development of the skill itself—a topic for future research. We commit to providing a clear definition of what we consider joint attention to be, and urge other researchers to do so as well.

3.1. Clinical and other applications

The delay or deficiency of joint attention abilities is a key diagnostic indicator for a range of atypically developing populations. In particular, children with Autism Spectrum Disorders (ASD) commonly exhibit deficits in joint attention (Mundy, 1995). Greater understanding of joint attention in infancy promises to yield important insights into the development of

language and social cognition, and directly informs developmental models of autism (Elison et al., 2013), and further, can inform interventions for children at-risk for or diagnosed with autism (Kasari et al., 2012). In addition to children at-risk for or diagnosed with ASD, deaf children who are born into hearing families can experience similarly impaired joint attention abilities (Nowakowski et al., 2009). Interestingly, impairment in this population highlights the social basis for joint attention (i.e., due to the mismatch in hearing status between the parent and the child) rather than its basis in a neurodevelopmental delay. In other words, while the mechanism is hard to get at, regardless of population, information from a range of populations can help complete the picture of the component parts underlying joint attention. For example, because joint attention can serve as an important scaffold for children to learn about communicative intent, one can imagine greater deficits in joint attention in deaf children of hearing parents who do not use sign language. This is an empirical question, the answer to which will rely on systematic implementation of an objective coding protocol. In short, a greater understanding of the construct of joint attention must include an agreed-upon definition and clear operationalization of the construct itself.

3.2. Adjusting the protocol to fit different needs

There are several ways in which the protocol outlined here can be modified to address specific questions. Here we suggest a handful of adjustments that can be made to accommodate the different perspectives researchers bring to the topic, particularly with regard to time windows for different criteria, as well as the investigation of other types of engagement.

As we have noted, there are several ways to determine the amount of time an episode of joint attention should span (Abney et al., 2017). The research paradigm used to develop the current coding protocol is commonly a parent-child free play, in which the dyad is left alone with several developmentally appropriate toy options and recorded without interference. However, these free-play tasks are not the only context in which joint attention is worthy of assessment. Additionally, the child participants in the studies that prompted our coding protocol are of varying ages but range from infant to toddler. Instances of joint attention outside of this age range are much less predictable, largely due to the lack of research in this population (see Bean & Eigsti, 2012; Nowakowski et al., 2011 for exceptions). As the type of task and the participant age vary, the coding protocol can also vary either informed by prior research or based on a data-driven approach.

Recent research also varies in whether to include the intentionality component that we outline here. The current protocol uses intentionality as a requirement for bid success. However, findings from eye-tracking suggests that intentionality need not preclude language learning (Yu & Smith, 2013) or influence an infant's ability to sustain attention (Suarez-Rivera et al., 2019). The importance of intentionality in relation to joint attention remains a point of debate, particularly as it depends on the age of the child and type of task a dyad is engaging in. As such, the current protocol can be adapted to track and compare bids with simple modifications to the decision tree. To implement identification of incidental engagement, or successful engagement that does not include active verification on the part of the initiator, one can simply code for the bids labeled here as "Not Coded" (Fig. 1). The

label of choice can then be added to the Controlled Vocabulary in ELAN, resulting in the forced dropdown menu containing a code for bids that would traditionally fail at step 3b (Fig. 1). The incorporation and assessment of these different types of engagement can help move forward our understanding of dyadic relationships. Of importance, is to successfully define and differentiate each type of attentional state for which we are interested. In the above example, adding a label for incidental engagement should not be immediately collapsed with those interactions prompted by an intentional bid.

We have developed our joint attention coding scheme based on our perception that explicit and objective guidelines were needed. We are aware that much joint attention research, including our own, is centered on white, suburban, upper-middle class families, and that joint attention may not look in the same across diverse populations. This remains an open question. A recent cross-cultural investigation found that the western model of joint attention —one that emphasizes the visual modality—does not generalize to ethnically diverse dyads (Little et al., 2016). Moreover, joint attention can be achieved through other modalities, such as via touch (Botero, 2016). Thus, we encourage application of our multimodal coding scheme on a range of populations. There is much to be learned on that front. We are also excited about new approaches to answering long-standing mechanistic questions about joint attention that are possible with neuroimaging techniques compatible for use with infants and toddlers. For now, we plan to continue pursuing systematic characterizations of that initiation and maintenance of joint attention in hearing parent-deaf child and hearing parent-hearing child dyads, with the goal of developing best-practice guidelines for parents of deaf children who are candidates for cochlear implantation.

4. Limitations and future directions

We are the first to admit that our coding scheme cannot serve as the basis for making claims about the origins of joint attention abilities. Rather, we have used our careful reading of seminal research, together with our own experience observing parent-child interactions, as a guide to developing a coding scheme that is objectively useable. We are adjusting our own approach as findings emerge from research that capitalizes on new techniques and technological advances (e.g., eye-tracking). However, we are compelled by the nuanced behavior that we are able to characterize through our own manual coding approach and urge researchers to consider the important findings behavioral coding continues to produce. Indeed, comparing these approaches will lead to fruitful research. For example, one issue we see as needing to be addressed is whether overlapping looking behavior—whether coincidental or intentional—produces the same developmental outcomes. Moreover, because overlaps and contradictions in terminology have been the source of substantial confusion (i.e., Chen et al., 2019; Racine, 2013), we encourage researchers to clarify their choice of terms. All of us should define what we mean by each term we use, and provide clear operationalization of the behavior(s) we are considering in any given study. Given the current confusion, any move in this direction will increase understanding and contribute to research advances. Ideally, a single, coherent definition of joint attention will be agreed upon, although at present this possibility seems somewhat remote.

Acknowledgements

We are grateful to Eve Clark for her insights and technical advice. We also thank Rose Scott and Eric Walle for their helpful discussions about issues addressed here.

Funding

This research was supported by research funds from the University of California, Merced to the first author and by the National Institute on Deafness and Other Communication Disorders under Grants R01 DC010075 and R01 DC018701 to the second author.

References

Abney DH, Smith LB, & Yu C (2017). It's time: Quantifying the relevant timescales for joint attention. The 39th Annual Meeting of the Cognitive Science Society, 1489–1494.

Adamson LB, Bakeman R, Suma K, & Robins DL (2019). An expanded view of joint attention: Skill, engagement, and language in typical development and autism. Child Development, 90(1). 10.1111/cdev.12973.

Bakeman R, & Adamson LB (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. Child Development, 55(4), 1278–1289. 10.2307/1129997. [PubMed: 6488956]

Baldwin DA (1991). Infants' contribution to the achievement of joint reference. Child Development, 62(5), 875–890. [PubMed: 1756664]

Baron-Cohen S (1991). Precursors to a theory of mind: Understanding attention in others. In Whiten A (Ed.), Natural theories of mind: Evolution, development and simulation of everyday mindreading (pp. 233–251). Basil Blackwell.

Bayliss AP, Murphy E, Naughtin CK, Kritikos A, Schilbach L, & Becker SI (2013). Gaze leading: Initiating simulated joint attention influences eye movements and choice behavior. Journal of Experimental Psychology General, 142(1), 76–92. 10.1037/a0029286. [PubMed: 22800442]

Bean JL, & Eigsti IM (2012). Assessment of joint attention in school-age children and adolescents. Research in Autism Spectrum Disorders, 6(4), 1304–1310. 10.1016/j.rasd.2012.04.003.

Botero M (2016). Tactless scientists: Ignoring touch in the study of joint attention. Philosophical Psychology, 29, 1200–1214. 10.1080/09515089.2016.1225293.

Bortfeld H, & Oghalai JS (2020). Joint Attention in hearing parent–deaf child and hearing parent–hearing child dyads. IEEE Transactions on Cognitive and Developmental Systems, 12(2), 243–249. 10.1109/TCDS.2018.2877658. [PubMed: 33748419]

Bruner JS (1974). From communication to language - A psychological perspective. Cognition, 3, 255–287. 10.1016/0010-0277(74)90012-2.

Butterworth G (1987). Some benefits of egocentrism. In Bruner JS, & Weinreich-Haste H (Eds.), Making sense: The child's construction of the world (pp. 62–80). London: Methuen.

Butterworth G, & Cochran E (1980). Towards a mechanism of joint visual attention in human infancy. International Journal of Behavioral Development, 3, 253–272. 10.1177/016502548000300303.

Butterworth G, & Jarrett N (1991). What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy. British Journal of Developmental Psychology, 9, 55–72. 10.1111/j.2044-835x.1991.tb00862.x.

Campos JJ, & Stenberg CR (1981). Perception, appraisal, and emotion: The onset of social referencing. In Lamb ME, & Sherrod LR (Eds.), Infants social cognition: Empirical and social considerations (pp. 273–314). Erlbaum.

Carpenter M, Nagell K, Tomasello M, & Butterworth G (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. Monographs of the Society for Research in Child Development, 63, 1–174. 10.2307/1166214.

Chen C, Castellanos I, Yu C, & Houston DM (2019). What leads to coordinated attention in parent–toddler interactions? Children's hearing status matters. Developmental Science. 10.1111/desc.12919.

Clark EV, & Estigarribia B (2011). Using speech and gesture to introduce new objects to young children. Gesture, 11, 1–23. 10.1075/gest.11.1.01cla.

Cohn JF, & Tronick EZ (1983). Three-month-old infants' reaction to simulated maternal depression. Child Development, 54(1), 185–193. 10.2307/1129876. [PubMed: 6831986]

Corkum V, & Moore C (1998). The origins of joint visual attention in infants. Developmental Psychology, 34, 28–38. 10.1037/0012-1649.34.1.28. [PubMed: 9471002]

de Barbaro K, Johnson CM, Forster D, & Deák GO (2016d). Sensorimotor decoupling contributes to triadic attention: A longitudinal investigation of mother-infant-object interactions. Child Development, 87, 494–512. 10.1111/cdev.12464. [PubMed: 26613383]

Deák GO, Krasno AM, Jasso H, & Triesch J (2017). What leads to shared attention? Maternal cues and infant responses during object play. Infancy, 23, 4–28. 10.1111/infa.12204.

Deák GO, Krasno AM, Triesch J, Lewis J, & Sepeta L (2014). Watch the hands: Infants can learn to follow gaze by seeing adults manipulate objects. Developmental Science, 17(2), 270–281. [PubMed: 24387193]

Depowski N, Abaya H, Oghalai J, & Bortfeld H (2015). Modality use in joint attention between hearing parents and deaf children. Frontiers in Psychology, 6, 1–8. 10.3389/fpsyg.2015.01556. [PubMed: 25688217]

Dube WV, MacDonald RPF, Mansfield RC, Holcomb WL, & Ahearn WH (2004). Toward a behavioral analysis of joint attention. Behavior Analyst, 27, 197–207. 10.1007/BF03393180.

Elison JT, Wolff JJ, Heimer DC, Paterson SJ, Gu H, Hazlett HC, Styner M, Gerig G, Piven J, Piven J, Hazlett HC, Chappell C, Dager S, Estes A, Shaw D, Botteron K, McKinstry R, Constantino J, Pruett J, … Wright F (2013). Frontolimbic neural circuitry at 6 months predicts individual differences in joint attention at 9 months. Developmental Science, 16(2), 186–197. 10.1111/desc.12015. [PubMed: 23432829]

Emery NJ (2000). The eyes have it: The neuroethology, function and evolution of social gaze. Neuroscience and Biobehavioral Reviews, 24, 581–604. www.elsevier.com/locate/neubiorev%0Ahttps://www.sciencedirect.com/science/article/pii/S0149763400000257. [PubMed: 10940436]

Gabouer A, Oghalai J, & Bortfeld H (2018). Hearing parents' use of auditory, visual, and tactile cues as a function of child hearing status. International Journal of Comparative Psychology, 31, 1–27.

Gabouer A, Oghalai J, & Bortfeld H (2020). Parental Use of Multimodal Cues in the Initiation of Joint Attention as a Function of Child Hearing Status. Discourse Processes, 57(5-6), 491–506. 10.1080/0163853X.2020.1759022. [PubMed: 32669749]

Gale E, & Schick B (2009). Symbol-infused joint attention and language use in mothers with deaf and hearing toddlers. American Annals of the Deaf, 153, 484–503. 10.1353/aad.0.0066. [PubMed: 19350956]

Guo J, & Feng G (2013). How eye gaze feedback changes parent-child joint attention in shared storybook reading? In Nakano YI, Conati C, & Bader T (Eds.), Eye gaze in intelligent user interfaces: Gaze-based analyses, models and applications (pp. 9–21). London: Springer, 10.1007/978-1-4471-4784-8_2.

Kaplan PS, & Werner JS (1991). Implications of a sensitization process for the analysis of infant visual attention. Newborn attention: Biological constraints and the influence of experience (pp. 278–307). Ablex Publishing.

Kasari C, Gulsrud A, Freeman S, Paparella T, & Hellemann G (2012). Longitudinal follow-up of children with autism receiving targeted interventions on joint attention and play. Journal of the American Academy of Child and Adolescent Psychiatry, 51(5), 487–495. 10.1016/j.jaac.2012.02.019. [PubMed: 22525955]

Little EE, Carver LJ, & Legare CH (2016). Cultural variation in triadic infant–caregiver object exploration. Child Development, 87(4), 1130–1145. 10.1111/cdev.12513. [PubMed: 27018870]

MacPherson AC, & Moore C (2017). Attentional control by gaze cues in infancy. Gaze-following: Its development and significance (pp. 53–75). Psychology Press. 10.4324/9781315093741-3.

Mendelson MJ, Haith MM, & Gibson JJ (1976). The relation between audition and vision in the human newborn. Monographs of the society for research in child development. JSTOR.

Moore C, & Corkum V (1994). Social understanding at the end of the first year of life. Developmental Review, 14, 349–372.

Mundy P (1995). Joint attention and social-emotional approach behavior in children with autism. Development and Psychopathology, 7(1), 63–82. 10.1017/S0954579400006349.

Mundy P, & Gomes A (1998). Individual differences in joint attention skill development in the second year. Infant Behavior & Development, 21(3), 469–482. 10.1016/S0163-6383(98)90020-0.

Mundy P, & Newell L (2007). Attention, joint attention, and social cognition. Current Directions in Psychological Science, 16, 269–274. [PubMed: 19343102]

Mundy P, Block J, Delgado C, Pomares Y, Vaughan Van Hecke A, & Venezia Parlade M (2007). Individual differences and the development of joint attention in infancy. Child Development, 78, 938–954. 10.1111/j.1467-8624.2007.01042.x.Individual. [PubMed: 17517014]

Mundy P, Fox N, & Card J (2003). EEG coherence, joint attention and language development in the second year. Developmental Science, 6(1), 48–54. 10.1111/1467-7687.00253.

Mundy P, Hogan A, & Doehring P (1996). A preliminary manual for the abridged early social communication scale (ESCS).

Nowakowski ME, Tasker SL, Cunningham CE, McHolm AE, Edison S, Pierre JS, Boyle MH, & Schmidt LA (2011). Joint attention in parent-child dyads involving children with selective mutism: A comparison between anxious and typically developing children. Child Psychiatry and Human Development, 42(1), 78–92. 10.1007/s10578-010-0208-z. [PubMed: 20960051]

Nowakowski ME, Tasker SL, & Schmidt LA (2009). Establishment of joint attention in dyads involving hearing mothers of deaf and hearing children, and its relation to adaptive social behavior. American Annals of the Deaf, 154, 15–29. 10.1353/aad.0.0071. [PubMed: 19569301]

Piaget J (1952). The origins of intelligence in children. In Cook M (Ed.), The origins of intelligence in children. W W Norton & Co., 10.1037/11494-000

Piaget J (1954). The construction of reality in the child. In Cook M (Ed.), The construction of reality in the child. Basic Books. 10.1037/11168-000.

Posner MI, & Rothbart MK (2007). Research on attention networks as a model for the integration of psychological science. Annual Review of Psychology, 58, 1–23. 10.1146/annurev.psych.58.110405.085516.

Prezbindowski AK, Adamson LB, & Lederberg AR (1998). Joint attention in deaf and hearing 22 month-old children and their hearing mothers. Journal of Applied Developmental Psychology, 19, 377–387. 10.1016/S0193-3973(99)80046-X.

Racine T (2013). Getting beyond rich and lean views of joint attention. In Seeman A (Ed.), Joint attention: New developments in psychology, philosophy of mind, and social neuroscience (pp. 245–251).

Salo VC, Rowe ML, & Reeb-Sutherland BC (2018). Exploring infant gesture and joint attention as related constructs and as predictors of later language. Infancy, 23(3), 432–452. 10.1111/infa.12229. [PubMed: 29725273]

Scaife M, & Brunner JS (1975). The capacity for joint visual attention in the infant. Nature, 253, 265–266. 10.1038/253265a0. [PubMed: 1113842]

Siposova B, & Carpenter M (2019). A new look at joint attention and common knowledge. Cognition, 189, 260–274. 10.1016/j.cognition.2019.03.019. [PubMed: 31015079]

Stephenson LJ, Edwards SG, & Bayliss AP (2021). From gaze perception to social cognition: The shared-attention system. Perspectives on Psychological Science. 10.1177/1745691620953773, 174569162095377.

Striano T, & Stahl D (2005). Sensitivity to triadic attention in early infancy. Developmental Science, 8, 333–343. 10.1111/j.1467-7687.2005.00421. [PubMed: 15985067]

Suarez-Rivera C, Smith LB, & Yu C (2019). Multimodal parent behaviors within joint attention support sustained attention in infants. Developmental Psychology, 55, 96–109. 10.1037/dev0000628. [PubMed: 30489136]

Thelen E, & Smith LB (1996). A dynamic systems approach to the development of cognition and action. MIT press.

Tomasello M, & Farrar MJ (1986). Joint attention and early language. Child Development, 57, 1454–1463. http://www.jstor.org/stable/1130423. [PubMed: 3802971]

Tomasello M, Carpenter M, Call J, Behne T, & Moll H (2005). Understanding and sharing intentions: The origins of cultural cognition. The Behavioral and Brain Sciences, 28(5), 675–735. 10.1017/S0140525X05000129. [PubMed: 16262930]

Trueswell JC, Lin Y, Armstrong B, Cartmill EA, Goldin-Meadow S, & Gleitman LR (2016). Perceiving referential intent: Dynamics of reference in natural parent-child interactions. Cognition, 148, 117–135. 10.1016/j.cognition.2015.11.002. [PubMed: 26775159]

Wittenburg P, Brugman H, Russel A, Klassmann A, & Sloetjes H (2006). ELAN: A professional framework for multimodality research. In Proceedings of the 5th International Conference on Language Resources and Evaluation (pp. 1556–1559).

Yu C, & Smith LB (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. PLoS One, 8(11). 10.1371/journal.pone.0079659.

Yu C, & Smith LB (2016). The social origins of sustained attention in one-year-old human infants. Current Biology, 26(9), 1235–1240. 10.1016/j.cub.2016.03.026. [PubMed: 27133869]

Yu C, & Smith LB (2017a). Hand–eye coordination predicts joint attention. Child Development, 88(6), 2060–2078. 10.1111/cdev.12730. [PubMed: 28186339]

Yu C, & Smith LB (2017b). Multiple sensory-motor pathways lead to coordinated visual attention. Cognitive Science, 41, 5–31. 10.1111/cogs.12366. [PubMed: 27016038]

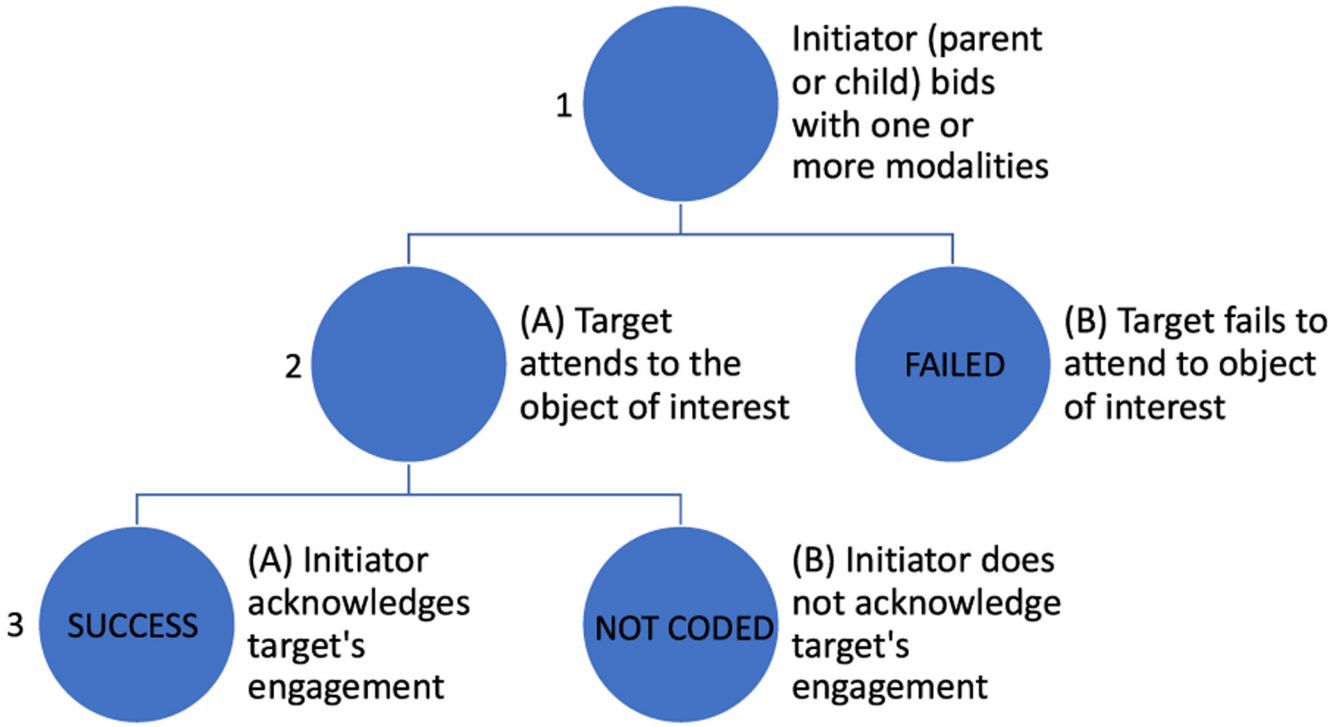

Fig. 1.

Coding tree used to identify successful and failed joint attention. Step one reflects the criteria described in 2.1 Intention. If the bid is determined to be intentional, the second level refers to the non-initiator's response to step one (described in 2.2). Lastly, the third level represents verification by the initiator (described in 2.3). Each level of the decision tree requires a yes or no decision that either ends the identification process and provides the appropriate label for such a situation or meets the criteria for a successful initiation of joint attention.

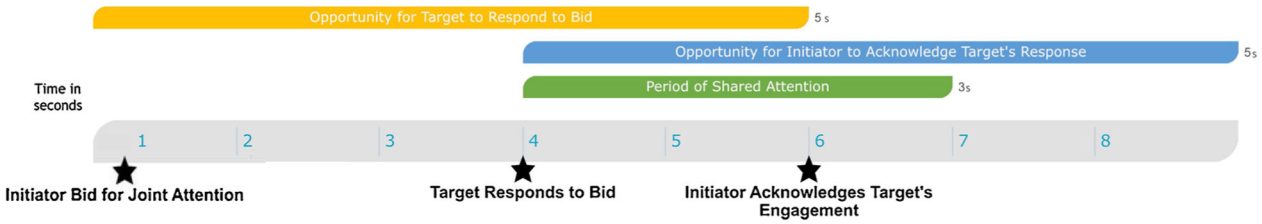

Fig. 2.

Timeline of joint attention initiation in seconds. Seconds 0-5 indicate the time after the initial bid for joint attention (1). After the onset of a bid, the non-initiator needs to respond to the bid within 5 s (yellow bar) for the bid to be considered a success (2). If the target does not respond, it is classified as a failed bid (Fig. 1: 2b). Once the target responds, the pair must engage with the object of mutual interest for at least 3 s (green bar). During this time, the initiator has a 5 s period to acknowledge the target's response (blue bar). If the initiator fails to acknowledge the engagement of the target, the bid is not coded (Fig. 1: 3b). Note: these times can be adjusted based on the specific population of interest, as can the requirement for initiators to verify joint attention (the final step) (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).