Title

Why does Joint Attention Predict Vocabulary Acquisition? The Answer Depends on What Coding Scheme you Use

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors

Sander, Jennifer

Çetinçelik, Melis

Zhang, Yayun

et al.

Publication Date

2024

Why does Joint Attention Predict Vocabulary Acquisition? The Answer Depends on What Coding Scheme you Use

Jennifer Sander (Jennifer.Sander@mpi.nl)
Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525XD Nijmegen, The Netherlands
Max Planck School of Cognition, Stephanstr. 1A, 04103 Leipzig, Germany

Melis Çetinçelik (Melis.Cetincelik@mpi.nl)
Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525XD Nijmegen, The Netherlands
Department of Cognitive Neuropsychology, Tilburg University, Warandelaan 2, 5000 LE Tilburg, The Netherlands

Yayun Zhang (Yayun.Zhang@mpi.nl)
Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525XD Nijmegen, The Netherlands

Caroline Rowland (Caroline.Rowland@mpi.nl)
Max Planck Institute for Psycholinguistics Wundtlaan 1, 6525XD Nijmegen, The Netherlands
Donders Institute for Brain, Cognition and Behaviour, Radboud University,
Thomas von Aquinostraat 4, 6525 GD, Nijmegen, The Netherlands

Zara Harmon (Zara.Harmon@mpi.nl)
Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525XD Nijmegen, The Netherlands

Abstract

Despite decades of study, we still know less than we would like about the association between joint attention (JA) and language acquisition. This is partly because of disagreements on how to operationalise JA. In this study, we examine the impact of applying two different, influential JA operationalisation schemes to the same dataset of child-caregiver interactions, to determine which yields a better fit to children's later vocabulary size. Two coding schemes—one defining JA in terms of gaze overlap and one in terms of social aspects of shared attention—were applied to video-recordings of dyadic naturalistic toy-play interactions (N=45). We found that JA was predictive of later production vocabulary when operationalised as shared focus (study 1), but also that its operationalisation as shared social awareness increased its predictive power (study 2). Our results emphasise the critical role of methodological choices in understanding how and why JA is associated with vocabulary size.

Keywords: Joint Attention; Language Acquisition; Vocabulary Acquisition

Introduction

Joint attention (JA) refers to coordinated attention on a particular object or event between two interaction partners (Gabouer & Bortfeld, 2021), for example a child-caregiver dyad. Past research has indicated that JA influences infants' and children's word learning, and that the quality and quantity of JA episodes correlate with children's later vocabulary development (Abney et al., 2020; Adamson, Bakeman, Deckner, 2004; Deák et al., 2013; Morales et al., 2000; Tomasello & Todd, 1983; Yu, Suanda & Smith, 2019).

However, despite decades of studying the role of JA in language acquisition, we still know less than we would like about *why* JA might facilitate language acquisition, with some claiming that visual attention is key; in particular that the combination of endogenous sustained and object-focused shared visual attention is sufficient (associative accounts, e.g., Yu et al., 2019) and others claiming that an additional active social awareness component is required to facilitate language learning (social accounts; e.g. Tomasello & Farrar, 1986).

The lack of consensus stems at least partially from the problem that we have still not agreed on how to operationalise JA (Tasker & Schmidt, 2008). On the one hand we have coding schemes, primarily used by associative theorists, that operationalise JA in terms of shared visual attention onto the same object at the same time by both interaction partners. These schemes use measures such as the amount of time that dyads spend looking, simultaneously, at the same referent during naming events to quantify JA. We call this a *gaze overlap* scheme (see, e.g., Yu et al., 2019).

On the other hand, we have schemes, primarily used by proponents of the social account, that start from the premise that gaze overlap is not enough to define a joint attentional episode. On this view, one must, in addition, code, in some way, a shared social awareness that indexes active coordination of attention on the part of at least one member of the dyad. We call this a *coordinated joint attention* scheme (see for example, Gabouer & Bortfeld, 2021; Tomasello & Farrar, 1986).

The picture is complicated by the fact that different coding schemes are applied to different datasets by different researchers, making it hard to determine which aspects of JA might be most strongly related to language. And this, in turn,

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

makes it difficult to clearly identify plausible reasons for why JA might facilitate language acquisition.

The aim of this paper was to begin to redress the balance. We did this by applying two coding schemes (a gaze overlap scheme and a coordinated joint attention scheme) to exactly the same data (video-recorded naturalistic play interactions between mothers and their 12-month-old children) and investigating how strongly each coding scheme predicted later production vocabulary size at 15 and 18 months. We found that JA contributed to vocabulary size regardless of how it was operationalised, but also that the operationalisation of JA affected the strength of its association with vocabulary (study 1). We also found that adding measures of successful shared social awareness improved the fit, over and above measures of shared gaze (study 2).

The paper is structured as follows. We first introduce the video dataset and the two coding schemes. In the Methods section, we detail our data collection and data processing procedure. We then report, in two studies, the results of statistical analyses that tested the contribution of shared focus and shared awareness to production vocabulary using two different coding schemes. We end with a discussion of how decisions about the operationalisation of JA affect the strength of its relationship with vocabulary, and thus should be informed by both the linguistic context as well as the properties of different types of social interactions.

Operationalising Joint Attention

There are, at least, two ways to operationalise joint attention, linked to different theoretical perspectives. Associative theorists, who argue that shared visual focus is key to vocabulary development, tend to operationalise JA as *gaze overlap*, and code behaviours that indicate that the caregiver and child are sharing *visual* attention to an object. These associative coding schemes use measures that quantify the amount of shared attention (e.g. duration of JA events).

Social theorists argue that explicit *awareness* of shared attention on the part of at least one interlocutor is crucial for vocabulary learning. They operationalise JA as requiring some measure of shared social awareness in addition to measures of shared focus. In addition to coding joint attentional behaviours that are not gaze (e.g. touch), these schemes also code behaviours that can be used to identify whether a JA attempt resulted in both parties actively coordinating attention (e.g., by coding the presence/absence of an initiation attempt at JA, and of an appropriate response). This scheme enables us to use not only measures that quantify shared focus but also those that quantify shared awareness (e.g. number of successful JA initiation attempts).

Importantly, these differences in the coding schemes result in different segments of the interaction, with different durations, being identified as JA (see Figure 1). For example, in the present paper, the average JA event as defined by the gaze overlap scheme lasted 1.18 seconds, while the average JA event as defined by the coordinated joint attention scheme lasted 33.86 seconds. Different coding schemes may thus yield different strengths of association between JA and vocabulary.

In study 1 we tested whether different operationalisations of JA lead us to draw different conclusions about the strength of the relationship between shared focus and vocabulary. In study 2 we tested whether shared awareness (in addition to shared focus) improved the model (as suggested by social accounts). For both studies, we focussed on JA around naming events, which are events for which the synchronization of children's attention on the object and its naming by the caregivers is crucial for word learning (Schroer & Yu, 2022).

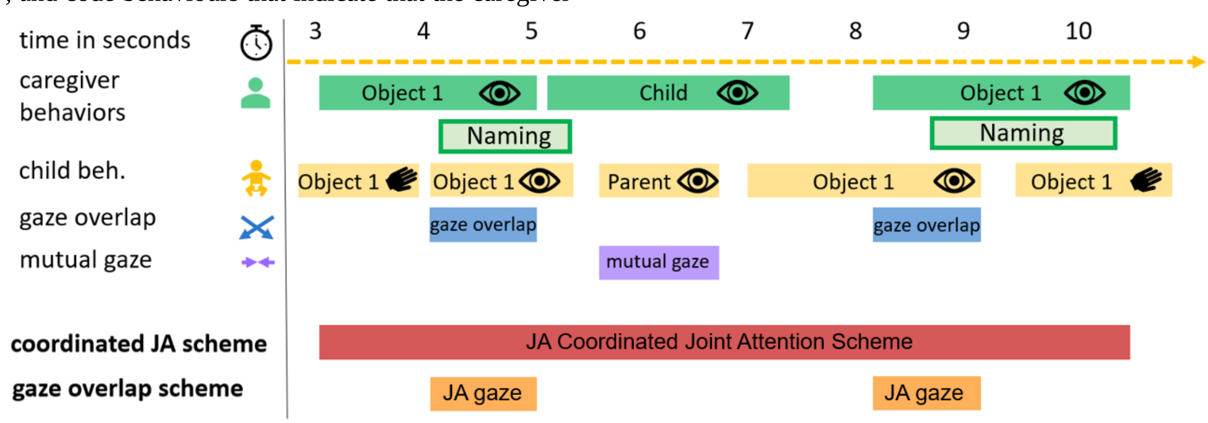

Figure 1: Schematic overview of behaviour annotations (e.g., gaze - represented by the eye symbols, touch - represented by the hand symbols, naming events) for caregiver (green) and child (yellow) in a 10–second sample from the original video data. Overlapping gaze on the object (blue) and mutual gaze (violet) are identified. JA as identified by the gaze overlap scheme is shown in orange and occurs if there is gaze overlap for at least 500 ms. JA as identified by the coordinated joint attention scheme is visualized in red, and occurs if both interaction partners engage in joint attentional behaviours (including gaze, touch, and other attentional behaviours) in a specific temporal relationship that indicates social awareness of both interaction partners within the ongoing shared attentional episode and all minimal duration requirements are fulfilled.

Methods

Participants

Participants were 45 mother-child dyads whose primary language was British English. Children were 12 months old. Dyads were excluded if, for either of the interaction partners, gaze data was only available for less than 20% of the time.

Experimental Setup

Children and their caregiver engaged in a naturalistic toy play situation with one of two possible sets of toys in a lab for about 10 minutes. Interactions were video-recorded by a researcher with a hand-held or tripod-fixated digital camera. Interaction between the dyad and the researcher was kept to a minimum. Caregivers were instructed to interact with the child as they would normally.

Data Processing

Data were processed and annotated in the video analysis program ELAN (Version 6.4, 2022) and post-processed and analysed in R (R Core Team, 2023). The data were annotated for caregivers' speech (categorized into naming events, description of objects, comments on objects, incomprehensible, and other non-relevant speech), and both interaction partners' gaze location (onto relevant objects, their co-participant, and other non-relevant locations), throughout the whole recording. In addition, around naming events and corresponding JA events, the data were annotated for attentional behaviours (pointing, waving, etc.), and touch (to the object or co-participant). These annotations were independent and took place prior to the application of the two JA coding schemes. Intercoder reliability for these annotations was, on average, 81%.

Gaze Overlap Scheme

Naming events were defined as all instances of noun-labelling by the caregiver that identified a present object. JA was defined, based on Yu et al. (2019), as a period of shared and continuously aligned visual attention of both interaction partners onto an object of interest for a minimum of 500 ms, allowing for short looks away by the caregiver for a maximum of 300 ms. We analysed JA events around naming events that overlapped for at least 1ms with naming events.

We used two shared focus measures defined by Yu and colleagues: proportion of JA present during each naming event; naming frequency per minute per dyad; and one more general shared focus measure: duration of JA events surrounding naming events.

Coordinated Joint Attention Scheme

Naming Events were again defined as all instances of noun-labelling by the caregiver that identified a present object. In order to define a measure of shared awareness, as well as measures of shared focus, JA episodes surrounding naming events were coded for the presence of: 1) an initiation attempt, which can consist of a number of behaviours (e.g. relevant speech, touch, or other attentional behaviours addressed at the object or interaction partner), followed by 2) a response from the interaction partner, which can likewise consist of a number of relevant behaviours, followed by 3) a validation behaviour in which the JA initiator has to validate the interaction partner's response by again showing some behaviour addressed at the interaction partner or object. To be coded as a successful JA event, the three behaviours have to be intentional and not accidental. Here, intention was assessed through indicators such as visual attention, physical orientation, haptic interaction, or overt gestures toward the object of interest (Trueswell et al., 2016; Gabouer & Bortfeld, 2021). Moreover, the behaviours have to happen in order within a specific timeframe (e.g. after each step the onset of the following step has to occur within five seconds), and the period of shared attention has to last at least three seconds. JA ends when one interaction partner disengages with the interaction partner and the object in the centre of the current JA period for a period of five seconds or longer (Gabouer & Bortfeld, 2021).

We used the three measures of shared focus described above: average duration of naming JA events, proportion of JA present during each naming event, and naming frequency per minute per dyad (see gaze overlap scheme). We also used one measure of social awareness: the number of successful JA initiations per dyad, which is defined by Gabouer and Bortfeld (2021) as sequences of interaction that contain 1) a JA initiation attempt, 2) a response by the interlocutor and 3) a verification of that response by the initiator, and which last at least three seconds.

Language measures

Children's concurrent and subsequent receptive and production vocabulary size was assessed at 12-, 15-, and 18 months of age through a parental report checklist, the British English CDI (Alcock et al., 2020), which is an adaption of the Mac-Arthur-Bates Communicative Development Inventory (CDI, Fenson et al., 2006). Parents tick off the words their children know (receptive vocabulary) and say (production vocabulary).

Results

Study 1

In study 1 we tested whether the two schemes, that operationalised JA differently, affected the strength of the association between shared focus and vocabulary. For each coding scheme, we ran a binomial regression predicting the proportion of known items in the CDI at 15 and 18 months from 3 measures of shared focus at 12 months. The three measures we used were: the *average proportion of overlap of naming events and JA per dyad*, the *number of naming events per minute*, and *the mean duration of JA episodes*. The regression model included an interaction between the three predictors and age.

Results for the gaze overlap scheme (see Figure 2) show that the average proportion of overlap of naming events and

JA per dyad had a positive effect on production vocabulary at 15 months ($\beta = 0.009$, $z = 3.37$, $p < .001$). This predictor had a significant interaction with age ($\beta = -0.01$, $z = -4.52$, $p < .001$), indicating a difference between the two age groups. The number of naming events per minute had a positive effect on production vocabulary at 15 months ($\beta = 0.34$, $z = 21.96$, $p < .001$), suggesting that a higher number of naming events was associated with a larger production vocabulary size at that age. The interaction with age was again significant ($\beta = -0.062$, $z = -3.12$, $p = .002$), this time because the effect of the number of naming events on production vocabulary was slightly weaker at 18 months. JA duration had a negative effect on production vocabulary at 15 months ($\beta = -0.11$, $z = -3.31$, $p = .001$), but this effect was reversed for vocabulary at 18 months ($\beta = 0.18$, $z = 5.03$, $p < .001$).

For the coordinated joint attention scheme (see Figure 3), the number of naming events per minute had a positive effect on production vocabulary at 15 months ($\beta = 0.35$, $z = 21.78$, $p < .001$), suggesting that a higher number of naming events was associated with a larger production vocabulary size at that age. The interaction with age was significant, because the effect of the number of naming events on production vocabulary was slightly weaker at 18 months ($\beta = -0.085$, $z = -4.15$, $p < .001$). Neither the proportion of overlap of naming events and JA, nor the average duration of JA events show a significant effect on later language production scores.

Critically, for our purposes, model comparison showed that the gaze overlap scheme fitted the vocabulary data better than the coordinated joint attention scheme (AIC gaze overlap = 5217.9, AIC coordinated joint attention: 5254, ΔAIC: 36.09).

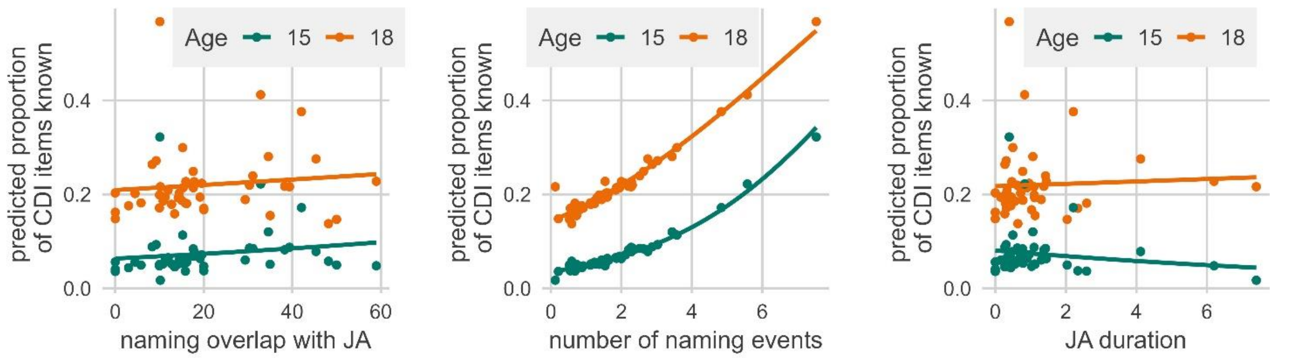

Figure 2: Predicted proportion of CDI items known (y-axis) plotted against the observed values (x-axis) for the three predictors based on the gaze overlap coding scheme: the average proportion of naming events that overlapped with JA (left), the average number of naming events per minute (middle), and the average JA event duration (right) per dyad.

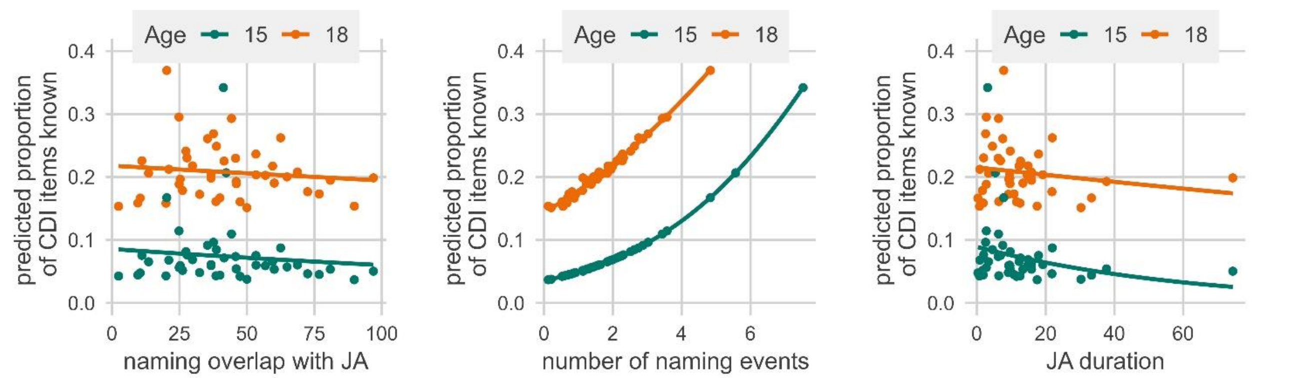

Figure 3: Predicted proportion of CDI items known (y-axis) plotted against the observed values (x-axis) for the three predictors based on the coordinated joint attention coding scheme: the average proportion of naming events that overlapped with JA (left), the average number of naming events per minute (middle), and the average JA event duration (right) per dyad.

Study 2

In study 1 we showed that the gaze overlap scheme yielded a better fit to the vocabulary data than the coordinated joint attention scheme. In study 2 we tested whether shared awareness (in addition to shared focus) would further improve the fit, as suggested by social accounts.

For each coding scheme, we once again ran binomial regressions predicting the proportion of known items in the CDI at 15 and 18 months. We started with the initial models in study 1 and then created minimally different models by adding one measure of shared awareness: *the number of* *successfully initiated JA events*. As before, the regression model included the interactions with age.

The results are presented in Figure 4. When we consider the gaze overlap scheme, most importantly the number of successful JA initiations (our shared awareness measure) had a positive effect on production vocabulary (β = 0.16, z = 7.83, p < .001). The interaction with age was significant (β = −0.15, z = −6.23, p < .001), suggesting that successful JA initiation is associated with later production vocabulary. With regard to our shared focus measures, findings were similar to study 1. The number of naming events per minute had a positive effect on vocabulary at 15 months (β = 0.16, z = 6.04, p < .001), but the interaction with age was significant, suggesting that the effect was slightly stronger at 18 months (β = 0.1, z = 3.08, p < .01). Finally, we found a significant negative effect of JA duration on production vocabulary at 15 months (β = −0.10, z = −2.93, p = .003), but, again, this effect was reversed for vocabulary at 18 months (β = 0.17, z = 4.65, p < .001). We did not find a significant effect of the overlap of naming events and JA events for 15 months (β = 0.003, z = 1.1, p = .27), but the interaction between this predictor and age was significant (β = −0.0085, z = −2.54, p = .011), because the effect was significant for 18 months (β = −0.005, z = −3.26, p = 0.01).

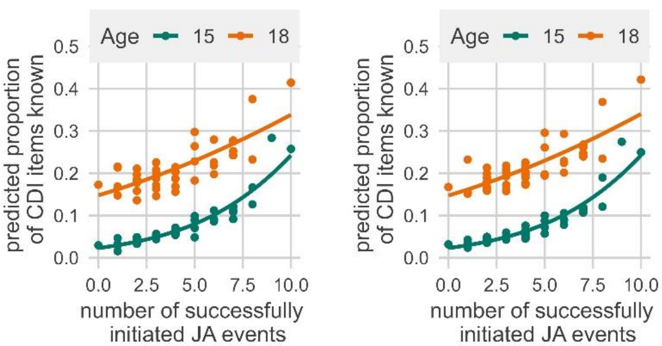

Figure 4: Predicted proportion of CDI items known (y-axis) plotted against the observed values (x-axis) for the number of successfully initiated JA events per dyad in the models that included shared awareness predictors in addition to shared focus predictors (study 2). Gaze overlap scheme (left) and coordinated JA scheme (right).

Looking at the model for the coordinated joint attention scheme, once again, critically, the number of successful JA initiations (our shared awareness measure) had a positive effect on production vocabulary (β = 0.17, z = 8.47, p < .001), suggesting that a higher number of successful JA initiations was associated with a larger production vocabulary. The interaction between this predictor and age was significant (ß = −0.18, z = −7.26, p < .001) because the effect was absent at 18 months. The effect of our shared focus measures was similar to that in study 1, though with some differences. The overlap of naming events and JA events also had a positive effect on production vocabulary (β = −0.005, z = −2.32, p = .02). The interaction between this predictor and age was significant (β = 0.006, z = 2.57, p = .01), because the effect was absent at 18 months. Finally, the number of naming events had a positive effect on vocabulary at 15 months (β = 0.15, z = 5.38, p < .001). The interaction with age was significant (β = 0.12, z = 3.32, p < .001), because the effect was stronger at 18 months. We found no significant effect of JA duration on production vocabulary scores.

Once again, the critical results are those from the model comparisons. Results are presented in Table 1. Overall, as in study 1, the gaze overlap coding scheme fit the data better than the coordinated joint attention coding scheme (AIC gaze overlap = 5159.7, AIC coordinated joint attention: 5185.6, ΔAIC: 25.9). However, adding the shared awareness predictor improved model fit for both models (ΔAIC: 58.2, ΔBIC: 53.2 for the gaze overlap scheme; ΔAIC: 68.4, ΔBIC: 63.4 for the coordinated joint attention scheme). Overall, the best fitting model used the gaze overlap coding scheme and included both measures of shared focus and the measure of shared awareness.

Table 1: AIC and BIC (and Δ) values for all models

	gaze overlap scheme	**coord. JA scheme**	
Study 1: shared focus models	AIC 5217.9	AIC 5254	ΔAIC 36.1
	BIC 5238	BIC 5274	ΔBIC 36
Study 2: shared focus + awareness models	AIC 5159.7	AIC 5185.6	ΔAIC 25.9
	BIC 5184.7	BIC 5210.6	ΔBIC 25.9
	ΔAIC 58.2 ΔBIC 53.3	ΔAIC 68.4 ΔBIC 63.4	

Discussion

In two studies, we investigated the relationship between joint attention (JA) and production vocabulary development. In Study 1, we tested if different operationalisations of JA—one implemented based on the gaze overlap scheme and another implemented based on the coordinated joint attention scheme—led us to draw different conclusions about the strength of the relationship between shared focus and vocabulary. In Study 2, we tested whether adding measures of shared awareness improved how well JA predicts vocabulary development—a major claim of social accounts. Study 1 showed that the shared focus measures applied to the gaze overlap scheme explained more of the variability in later vocabulary size than the same measures applied to the coordinated joint attention scheme. Study 2 showed that including measures of shared awareness always improved model fit compared to the models that only containing shared focus measures. The finding that adding measures of shared awareness improved the models suggests that investigations of the relationship between JA and later language abilities benefit from including a measure of social awareness in addition to measures of shared focus.

Overall, the gaze overlap scheme provided a better fit to the vocabulary data in both studies. However, adding the social awareness measure to the models not only improved the model fit of the coordinated joint attention scheme, but also of the gaze overlap scheme.

In the coordinated joint attention scheme JA has been operationalised through intentionality of behaviours and awareness of shared attention between the interaction partners. This has been done by introducing a three-step initiation procedure into the coding scheme. This inevitably adds subjectivity to the coding scheme, as intentionality and awareness have to be assessed by a human and cannot be automated. This subjectivity may add additional noise to the data, which could explain why in both studies, the overall fit of the models for the gaze overlap scheme were better than the models for the coordinated joint attention scheme. Meanwhile, the gaze overlap scheme can be considered more objective and easier to automate, as it does not rely on subjective considerations of intentionality or awareness.

The findings of this study have a number of implications. Most importantly, they show that how we define JA determines not only how we operationalise JA but also how we measure it. These definitions can vary widely, from definitions that rely only on gaze overlap (i.e., the associative accounts) to ones that also include a variety of behaviours that can signal mutual awareness of the interactional partners' attention. As our results show, the choice of JA definition, operationalisation and measures influences which conclusions can be drawn about the relationship of JA and language development. Thus, studies investigating these differences, such as the current one, help identify the aspects of JA that facilitate language acquisition.

The coding schemes we compared differ in several fundamental ways, which went beyond their different operationalisations of JA as shared focus or shared focus and shared awareness:

First, they differ in the *number and variety of behaviours* they include in their coding schemes. In the gaze overlap scheme, JA is operationalised by gaze overlap only, excluding non-gaze related behaviours and mutual gaze between the interaction partners. In the JA operationalisation of the coordinated joint attention scheme, gaze is only one of the several possible behaviours that dyads engage in during JA, in addition to touch or general attentional behaviours such as pointing and tapping—behaviours that have been shown to influence a child's attention (Deák et al., 2017). Further, this operationalisation of JA does not limit eye gaze to sharing gaze onto the same object, but also allows for mutual gaze. This flexibility allows a variety of behaviours to co-occur and overlap as the interaction partners maintain or establish JA by the means of these behaviours. However, it introduces an element of subjectivity into the coding, giving the coder more degrees of freedom, and thus adds more noise to the data. This may be the reason why it did not capture vocabulary size as accurately as the gaze overlap scheme.

Second, the schemes differ in their *temporal granularity*. In particular, they differ in the required minimal duration of JA (gaze overlap scheme: 500 ms vs. coordinated JA scheme: 3,000 ms) and the duration of disengagement which terminates an ongoing JA event (gaze overlap scheme: 300 ms vs. coordinated JA scheme: 5,000 ms). This results in a different time scale of JA events. Note that this difference does not result from differences in how JA is defined but is very much in line with it. As the gaze overlap scheme only considers gaze, short moments of interruption in gaze overlap translate to interruptions in the JA episode. In contrast, the broader definition of the coordinated joint attention scheme requires us to consider a longer timescale, including longer periods of interruption. Allowing for longer periods of interruption is in line with the dynamics of the social situation. Even though granularity is not a consequence of the JA definition, future research should look at the effect of defining JA at different temporal granularities.

Third, the schemes differ in their *suitability for interactions between blind and/or deaf individuals*. Because the gaze overlap scheme operationalises JA as shared gaze onto an object of interest, it cannot be used with blind individuals. Whether blind individuals can engage in JA at all depends on the applied definition of JA. Definitions and coding schemes of JA that require gaze overlap would then have to draw the conclusion that blind individuals do not engage in JA. Definitions and coding schemes of JA that allow for different behaviours and define JA as coordinated joint engagement can potentially consider joint interactions of blind individuals as JA, for example through joint touch.

Similar issues arise when we consider applying the coding schemes to interactions between individuals using a sign language. It has been shown that signing individuals engage more often in mutual gaze and in higher frequent gaze switching between the object or action of interest and their interaction partner, compared to speaking dyads (Lieberman et al., 2011, 2014). Even though shared gaze onto an object of interest is possible in signing dyads, mutual gaze is required for a sign to be successfully communicated.

The gaze overlap scheme would reject sign interactions as JA. Thus, this scheme is not the best choice for investigating JA in the environment of interactions using a sign language. The coordinated joint attention scheme, allowing for mutual gaze, touch, and other attentional behaviours, shows a higher level of flexibility and has successfully been applied to interactions in a sign language before (ASL, Sander, Lieberman & Rowland, 2023).

Finally, even though other behaviours beyond gaze are of special interest in some populations, in fact non-gaze related JA behaviours occur in all interactions. Thus, future work might look at extending schemes like the gaze overlap scheme to broaden their definition of what behaviours could be included in JA episodes. Choosing an appropriate scheme for the given participant group, making explicit decisions about the granularity of JA, as well as choosing the relevant measures of JA represent a challenging and understudied aspect of research on JA. Future research is needed to understand which aspects of joint attention might facilitate language acquisition.

References

Abney, D. H., Suanda, S. H., Smith, L. B., & Yu, C. (2020). What are the building blocks of parent-infant coordinated attention in free-flowing interaction? *Infancy*, 25(*6*), 871-887.

Adamson, L. B., Bakeman, R., & Deckner, D. F. (2004). The development of symbol-infused joint engagement. *Child development*, 75(4), 1171-1187.

Alcock, K. The UK communicative development inventory database: words and gestures ages 8-18 months 2012-2016.

Deák, G. O., Krasno, A. M., Jasso, H., & Triesch, J. (2018). What leads to shared attention? Maternal cues and infant responses during object play. *Infancy*, 23(1), 4-28.

Deák, G. O., Triesch, J., Krasno, A., de Barbaro, K., & Robledo, M. (2013). Learning to share: The emergence of joint attention in human infancy. *American Psychological Association*.

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2006). MacArthur-Bates Communicative Development Inventories, Second Edition (CDIs) [Database record]. *APA PsycTests.*

Gabouer, A., & Bortfeld, H. (2021). Revisiting how we operationalize joint attention. *Infant Behavior and Development*, *63*, 101566.

Lieberman, A. M., Hatrak, M., & Mayberry, R. I. (2011). The development of eye gaze control for linguistic input in deaf children. In *Proceedings of the 35th annual Boston University conference on language development* (Vol. 108, pp. 391-404). Somerville, MA: Cascadilla Press.

Lieberman, A.M., Hatrak, M., & Mayberry, R. I. (2014). Learning to Look for Language: Development of Joint Attention in Young Deaf Children. *Language Learning and Development*, *10*(1), 19-35.

Max Planck Institute for Psycholinguistics Nijmegen, The Language Archive (2022). *ELAN* (Version 6.4). Retrieved from https://archive.mpi.nl/tla/elan

Morales, M., Mundy, P., Delgado, C. E. F., Yale, M., Messinger, D., Neal, R., & Schwartz, H. K. (2000). Responding to Joint Attention Across the 6-Through 24-Month Age Period and Early Language Acquisition. *Journal of Applied Developmental Psychology*, *21*(3), 283-298.

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/.

Sander, J., Lieberman, A., & Rowland, C. F. (2023). Exploring joint attention in American Sign Language: The influence of sign familiarity. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 45, No. 45).

Schroer, S. E., & Yu, C. (2022). The real-time effects of parent speech on infants' multimodal attention and dyadic coordination. *Infancy*, *27*(6), 1154-1178.

Tasker, S. L., & Schmidt, L. A. (2008). The "dual usage problem" in the explanations of "joint attention" and children's socioemotional development: A reconceptualization. *Developmental Review*, 28(3), 263-288.

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child development*, 1454-1463.

Tomasello, M., & Todd, J. (1983). Joint attention and lexical acquisition style. *First language*, 4(12), 197-211.

Trueswell, J. C., Lin, Y., Armstrong III, B., Cartmill, E. A., Goldin-Meadow, S., & Gleitman, L. R. (2016). Perceiving referential intent: Dynamics of reference in natural parent–child interactions. *Cognition*, 148, 117-135.

Wickham H (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

Yu, C., Suanda, S. H., & Smith, L. B. (2019). Infant sustained attention but not joint attention to objects at 9 months predicts vocabulary at 12 and 15 months. *Developmental Science*, *22*(1).