"Put Your Hands Up": How Joint Attention Is Initiated Between Blind Children And Their Sighted Peers

Katherine Jones
University of Bristol
Bristol, United Kingdom
kj13690@bristol.ac.uk

Martin Grayson
Microsoft Research
Cambridge, United Kingdom
mgrayson@microsoft.com

Cecily Morrison
Microsoft Research
Cambridge, United Kingdom
cecilym@microsoft.com

Ute Leonards
University of Bristol
Bristol, United Kingdom
ute.leonards@bristol.ac.uk

Oussama Metatla
University of Bristol
Bristol, United Kingdom
o.metatla@bristol.ac.uk

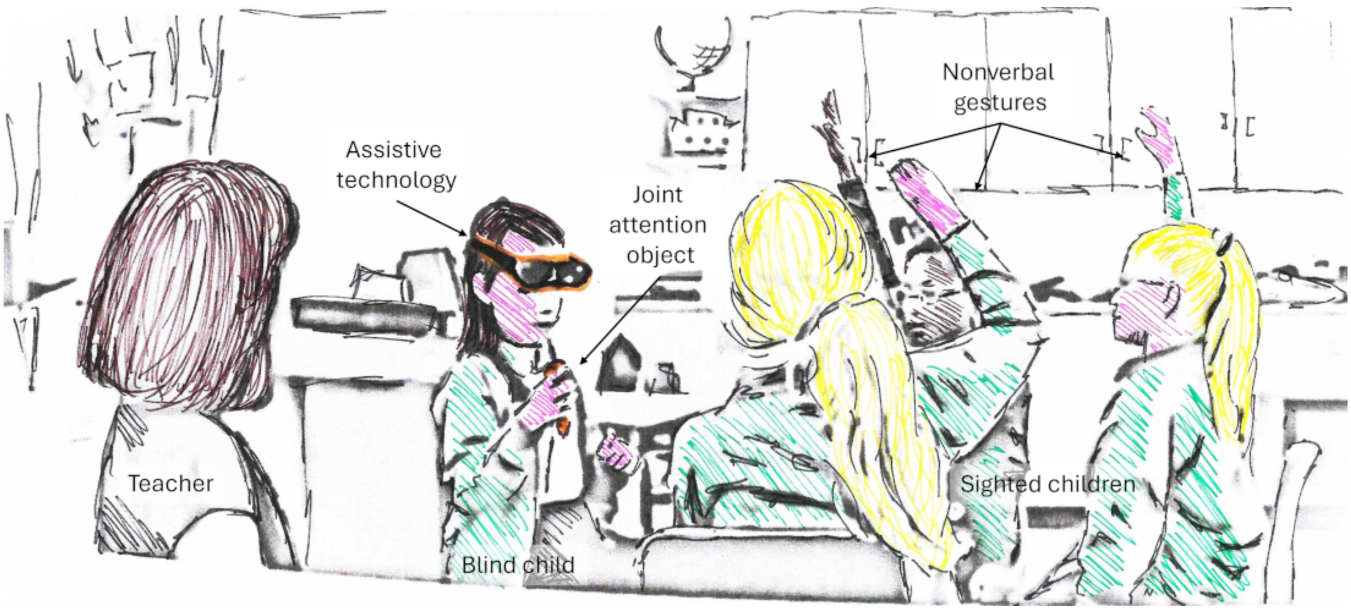


Figure 1: A commonly taught response to the initiation of joint attention, a prerequisite of any social interaction, in sighted school children is to silently raise their hands to indicate readiness for engagement. In this image, the blind child has just told her peers that she is going to choose one of them to demonstrate the rules of a game with, to which the sighted children raise their hands in reply. This automatic nonverbal gesture by the sighted peers does not signal to the blind child that her joint attention initiation attempt was successful, meaning that she has to either try again or wait for a signal of readiness for engagement that she can interpret from her sighted peers, such as a vocalisation.

Abstract

Initiating joint attention (JA) is a fundamental first step in social interactions. In sighted individuals, it relies predominantly on visual cues, such as gaze and hand gestures. These features can reduce opportunities for blind and visually impaired (BVI) and sighted people to interact. Understanding the strategies to navigate these challenges is necessary to develop technology that can facilitate more inclusive JA. To address this, we conducted a longitudinal case study of five children with mixed visual abilities engaging in activities rich with JA opportunities. In a teacher-led classroom, the children experimented with the use of an AI-powered headset designed to support BVI people in social situations. Interaction analysis established that situational complexity affects the children's responses to initiation attempts. Furthermore, the headset adds to this complexity, affecting the frequency and reactions to attempts to initiate JA. The findings informed the creation of a JA initiation framework and suggestions for future design.

CHI '25, Yokohama, Japan

ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3714005

CCS Concepts

• **Human-centered computing → Empirical studies in accessibility**.

Keywords

Joint Attention, Joint Engagement, Social Interaction, Blind, Visually Impaired, Children, Inclusion, Mixed-Visual Ability, AI, Headset

ACM Reference Format:

Katherine Jones, Martin Grayson, Cecily Morrison, Ute Leonards, and Oussama Metatla. 2025. "Put Your Hands Up": How Joint Attention Is Initiated Between Blind Children And Their Sighted Peers. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3706598.3714005

1 Introduction

Joint attention (JA) is a phenomenon where two or more people pay attention to each other and an object or concept [147]. JA is crucial for social and cognitive development. However, initiating JA that leads to meaningful engagement is significantly challenging among mixed-visual groups [68]. We sought to understand how blind, visually impaired[1] (BVI) and sighted children initiate joint attention (JA). We begin by introducing a common scenario observed throughout this work between BVI and sighted children when attempting to initiate JA, when using assistive technology. (Figure 1).

Four seven-year-old children are playing a pass-the-object game called 'melting ice-cream' (4.2). Their sighted teacher, Miss Williams, is facilitating. One of the children (Belle) is congenitally blind and has no light perception, and the other three (Esme, Margaret and Tiffany) are sighted. Belle is wearing a headset designed to support BVI people in social situations by identifying faces and verbalising to the wearer where their peers are in proximity. She also holds the JA object, a wooden ice cream cone with a blob of wooden ice cream on top. Miss Williams initiates the activity through speech:

"OK, so melting ice cream *(the game)* is starting. Girls, what you are doing is showing Belle where you are by telling her and touching her…"

"Oh, don't forget the forewarning!" *Belle interrupts the teacher, her engagement implying that the initiation was a success.* "The forewarning", *Belle repeats.*

"Yeah, remember to forewarn you"*, the teacher agrees,* "which means what, Belle?"

The word forewarning is complex, and Belle becomes motionless for two seconds before saying, "I'll show you".

"You are demonstrating it, aren't you?" *the teacher says,* "So if you are passing something to Belle…"

"So" *Belle interrupts again.* "Imagine that if I was. So, I'm going to demonstrate it, and I'm going to pick somebody to demonstrate."

In unison, Esme, Margaret and Tiffany wordlessly throw their hands in the air in a bid to be chosen by Belle. The headset, designed only to identify faces, does not recognise gesturing with intent to engage and remains silent. The gestures, useful for Esme, Margaret and Tiffany, go unheeded by Belle.

This scenario indicates that there needs to be an understanding of what leads to the successful initiation of JA for social interactions between mixed visual groups before developing technology that facilitates JA for all social partners. The collective reflex of sighted people to silently raise their hands to indicate a desire to be selected for a task, even when a blind child is making the selection, shows that nonverbal gestures are so ingrained in everyday social interactions that despite interaction partners not benefiting from these actions, they are used anyway (see Figure 1). This response is a testament to sighted people's reliance on vision for the complex subtle nuances that social interactions require [68]. Here, we refer specifically to JA, a principal element for social and cognitive development in children [86, 133]. JA can involve physical objects, such as a ball, or non-tangible concepts, such as discussions about friendship, that all interaction partners attend to. Initiating JA is a frequent occurrence in everyday social life. However, due to the strong reliance of sighted people on nonverbal (predominately) visual cues for JA events, sighted people have more opportunities to pick up on or initiate an event, for example, via gaze, than those with visual impairments [120].

Any social interaction needs to be initiated by at least one person as well as responded to by another and attempts to share attention with others are based on sensory cues [83]. Initiation can occur through different modalities, such as speech, touch, gaze and nonverbal gestures. For sighted people, the modality to initiate interactions with others relies heavily on visual cues [58]. A wave and a shout show interest in initiating an interaction with someone. A nod or orientation of the head implies a positive response, whereas turning away suggests a lack of interest. Visual cues are even indicators to tell that someone has not heard the initiation attempt, such as looking out for a lack of physical response, which suggests a need to try again [124, 139].

BVI individuals do not necessarily have access to visual cues such as a quick glance or a wave to signal intent to engage in an activity. Instead, they use different signals of intent, such as verbal or tactile signals. Furthermore, sighted individuals may not notice sensory cues sent by a BVI individual as these cues may not be the ones commonly used by the sighted majority. This subtle difference in using signals from different modalities to initiate JA may contribute to a lack of social opportunities for BVI individuals [21, 88]. A lack of opportunities for social engagement is, in turn, the most likely contributor to many BVI individuals' experiences of having fewer friends, lower educational achievement and higher unemployment rates [73, 125].

Increasing social opportunities for BVI people has been identified as an important area of technology design in recent years [16, 56, 115]. This technology development has focused on participatory designs to prioritise BVI people's autonomy in their environment, as well as interdependence in social situations [128]. For example, substantial progress has been made in supporting navigation [72, 74], collaboration in the workplace [90], classroom support [89, 95, 97], access to social media platforms [127] and identifying people in proximity [150]. Yet, research on technology designed to support JA, specifically its initiation and continuation

[1] We recognise that research has suggested that identity-first language is the preferred approach for many individuals [20, 70]. However, people-first language is still recommended academically and medically [50]. We acknowledge this debate by choosing to use a mix of identity-first and people-first language [41, 149].

once the interaction has successfully begun, is still in its infancy [68]. To develop technology that specifically addresses the ability to initiate JA in complex social situations, we need first to understand what leads to the successful initiation of JA for social interactions between BVI children and their sighted peers. The aim of this paper is to start reducing this gap in our knowledge by developing an understanding of this phenomenon through interaction analysis. For this, videos were collected of a mixed visual children's group during teacher-led sessions focusing on PeopleLens technology in a mainstream UK primary school.

As it stands, there is no clear understanding of what causes an absence of JA initiation for children in mixed visual groups. As a result, little clarity exists around what technology needs to address when trying to facilitate the initiation of JA between BVI children and their sighted peers. We specifically look at the initiation of JA so that we can identify how and why it differs from the JA event itself to inform technology design to encourage more opportunities for social interactions. We thus make the following three contributions:

(1) We investigate the initiation of JA in a mixed visual group of children.
(2) We explore how the initiation of JA is affected by the complexity of context, activity type and number of possible interaction partners.
(3) We emphasise how our findings can inform future technology design to support an increase in the frequency of successful initiations of JA in mixed visual groups.

2 Background and related works

2.1 Joint attention

Coordination of behaviour is an essential part of social interaction [8, 133]. This coordination tends to pivot around an object or concept, a phenomenon known as joint attention (JA) [1, 47, 145]. Recognised as an important developmental milestone, JA begins to emerge in a child's first year of life and starts with eye gaze [80, 103, 129]. This ability enables the child to begin to communicate with their caregivers and is a skill that continues its development throughout infancy into childhood and beyond [8, 102, 138]. The concept of JA has been addressed in a wide range of literature, which means that we have a deep understanding of when the skill manifests, what it entails and how it helps with cognitive, cultural, linguistic and social development [2, 104, 131, 146, 148, 152]. As with many developmental skills, JA has a flurry of growth in infancy, mostly around physical objects, which appears to conflate at around three years of age. However, its progression does not stop there. JA progresses from a simple focus on the physical object to encompass concepts, such as using symbols to describe what colour something is, or conceptual, such as exploring the beauty of nature [3, 24, 37]. Whilst relying on multi-sensory cues such as touch during early infancy, the initiation and maintenance of JA in older children and adults move predominately to eye gaze and nonverbal gestures [10, 25, 30]. Verbal skills have also been shown to affect the quality of engagement that children have with their joint attention partner, and as it improves, so too does the interaction [4, 5, 18]. A major takeaway from the literature is that JA requires the ability to coordinate our senses, whether it be through vision, speech, or other modalities, with vision being the dominant sense for most people. If one or more of these cannot be utilised, then the ability to successfully engage in JA with others is affected. This is particularly the case for vision loss due to the reliance that JA has on visual cues [9, 19, 75, 77].

2.2 Development of joint attention in children who have visual impairment

Globally, there are around 90 million children under 18 living with a vision impairment [143]. On a spectrum, vision impairment varies significantly in type and severity across individuals [54, 142, 156]. Whether the condition is congenital, total, partial or progressive, it affects how BVI individuals interact with others and how sighted people interact with BVI individuals [118]. The more severe the vision loss, the more challenges the individual faces, and if an individual is born with no vision, not even light perception, then the difficulty in sharing interactions with others increases significantly [60]. Indeed, research shows that even a small amount of light perception has a significant effect on an individual's ability to share attention [117]. Research shows us that BVI and sighted infants show the same rate of development in JA for the first six months or so, at which point the BVI infant's development appears to slow [140]. Early learning of JA focuses on a person and an object, with touch being a key sensory factor contributing to JA; however, already during these early stages, BVI children demonstrate delays in object permanence, cannot read facial expressions, and struggle with spatial awareness [7, 17]. As the child gets older, their lack of responses to visual JA cues can mirror those of autism and lead to inaccurate diagnoses [23, 153]. A possible contributing factor to BVI children's poorer outcomes is the sighted majority's lack of understanding of BVI children's social signals [33]. BVI children, as a result, are more likely to experience loneliness and bullying, have poor educational outcomes, and, when older, be less likely to marry or have a full-time job than their sighted peers [73, 125]. These outcomes may be impacted by the fact that BVI and sighted people do not initiate or respond to initiation attempts to socially interact in the same way. In particular, BVI people do not seem to have the same type or number of opportunities to develop social and cognitive skills as their sighted peers [68]. This knowledge highlights the importance of establishing the different nuances between the initiation of JA and the JA event itself. Understanding these differences means that future technology design can make clear distinctions between the two phenomena, thus directing its attention to where it would be of most use.

2.3 Initiating joint attention

Recent research has shown that being able to initiate JA is key for successful interactions to take place and that BVI individuals may struggle with the ability to initiate JA successfully with their sighted peers [68]. As they grow older, initiation of JA in sighted children relies more and more on gestures, body language, eye contact and vocalisations such as "hi!", often paired with an accompanying visual gesture such as a wave with their hand [55]. BVI children, however, experience the initiation of JA differently. In adults, for example, sighted people may offer nonverbal visual cues in a typical way to the BVI person, but these initiations will not be perceived and responded to, which results in a void or lack of opportunity to interact [68]. Despite the importance of JA, the literature on the initiation of JA is sparse and focuses either on infants [17, 51, 145], autistic children and the importance of nonverbal visual cues such as a smile [31, 53], or on the JA event itself [105]. This lack of research means that in Human-Computer Interaction (HCI), we do not yet know how to effectively design technology that can give children greater autonomy whilst also experiencing interdependence in social groups.

2.3.1 Situational awareness and joint attention. Typically measured via eye tracking, situational awareness is the individual's ability to perceive entities in the environment, know what they are, what they are for, and how they might affect them[43, 107]. It is the prerequisite for us to be able to perform tasks effectively [44]. Situational awareness not only relies on the readiness of the participant to perform a task but also on being able to understand the environmental context to perform the task successfully. If there is too much sensory information to be processed, decision-making is delayed or based on higher uncertainty, leading to more errors in performing the task [81]. One example of a failure in situational awareness is inattentional blindness, the inability to recognise important objects when completing a task [35]. Inattentional blindness is common in sighted people, who miss information they do not expect to see [134]. In the context of JA, inattentional blindness and disrupted situational awareness might affect attempts by both BVI and sighted parties to initiate JA successfully. Environmental interference would be greater for BVI people because where a sighted person can obtain immediate spatial information of the environment, including potential social interaction partners, in a single glance, the same environment for a BVI person is a continuous learning process that takes time [144, 162].

2.3.2 Dual tasks and joint attention. Indeed, sighted children can assimilate information from the environment, such as its spatial layout, very quickly through their vision (i.e. they get a gist of a scene within a few hundred milliseconds and can dynamically update information through simultaneously available sensory information); this means that when confronted with a simple switch in tasks that require different types of sensory information, they expertly switch between the two [62]. However, BVI children have difficulty switching between tasks, primarily because they do not continuously receive visual information about environmental changes and thus rely on memory about spatial configurations of the environment around them or sequential updating of information through other sensory modalities [126]. For example, it is very hard for them to notice if one of their potential social interaction partners has moved silently to another location or is engaged in another task, i.e. has changed the focus of their attention. This reliance on memory increases cognitive load substantially. Moreover, for each new task, they are essentially novices having to construct the information required to perform this task as they encounter it, which in turn may affect the BVI child's speed of interaction [39].

2.4 Technology to support social interactions for blind and visually impaired children

Early technology for BVI people has prioritised navigation and mobility [15]. The vOICe, a headset that turns images into sounds, or SmartVision, an electronic whitecane used for navigation, both focus on providing verbal information about the environment [46, 66, 92, 93]. Research interest in assistive technology, designed specifically for BVI people to have greater autonomy in social situations, is also increasing [42, 158, 163] with more technical support for BVI people's inclusion within mainstream society [98]. In particular, early intervention to support children's development is key because support in early childhood tends to lead to better outcomes in later life [26, 27, 52, 68, 101].

Furthermore, there is growing interest in utilising the power of AI to support BVI children and adults, such as with wearable bracelets, smart glasses, robot guide dogs, and AI-powered suitcases to aid in navigation [40, 45, 57, 119, 155]. The aim of these designs is to lay the foundation for greater autonomy for BVI people. However, whereas children tend to use mobility devices such as canes, pre-canes, and virtual reality [32], so far, there has been little uptake of other types of technology, perhaps due to concerns around potential negative social stigmas [11, 132].

To address this problem, research is moving away from searching for solutions that focus on only the BVI person to solutions that focus on the interplay between different yet equal partners and thus promoting interdependence between different individuals [12, 34]. The same holds for efforts to support social relationships between BVI and sighted children in classrooms. Examples of work with mixed visual groups of children include co-designing educational games [95] and Incloodle, an inclusive, tablet-based, picture-taking application designed to encourage children to engage with each other [135]. However, such technological support does not focus specifically on the behavioural and environmental complexity that affects social interactions between BVI and sighted people. Furthermore, although there is work for interdependence in social spaces for adults [154], to date, no technology is readily available in the public domain that is specifically designed to encourage interdependence in mixed visual groups of children with regard to interpreting social cues for the successful initiation of social interactions. More importantly, the methods used have been very context-dependent, whereas, for the initiation of JA in social situations, the requirement would be to develop something that can be used across a wide range of different contexts and thus in everyday life.

Even though there have been a number of designs that considered the need for more interdependence opportunities [49, 108, 113, 121], these attempts focused on how to provide "missing" visual information to the BVI person, thus prioritising a sighted person's and thus ableist's perspectives. To be useful and utilised, any technology needs to incorporate both interaction partners' points of view to establish exactly what is needed for both partners to interact with each other as equals [95, 97, 123]. An example of technology that aims at supporting social-spatial awareness to initiate JA is the PeopleLens [28]. The PeopleLens is a pair of AI-powered glasses worn by the BVI person. The headset is in development and has progressed through a number of iterations, most recently using a specialised adaptation of Nreal light glasses [160]. During its development, the researchers and technicians worked with BVI children and adults with the goal of creating a user-centred device that gives autonomy to BVI children and interdependence to mixed visual social groups.

3 Scope

We aimed to develop an understanding of the initiation of JA between BVI and sighted children through an analysis of interactions between a blind child within her peer group of sighted children. The children interacted with each other during teacher-led sessions focused on an AI-powered headset designed to support BVI children in social situations in a mainstream primary school. We wanted to develop an understanding of how blind and sighted children initiate JA with each other, which techniques worked well, which worked less well and why breakdowns in initiation might be happening. We also wanted to identify the effects of initiation when the blind child was wearing the PeopleLens compared to when she was not wearing it. We focused on the following research questions:

- RQ1: What characterises the initiation of JA from the perspectives of both the BVI child and their sighted peers?
- RQ2: What situational factors impact the initiation of JA in mixed visual groups?

4 Methods

4.1 Participants

Four seven-year-old sighted children, one seven-year-old congenitally blind child and two sighted teachers were recruited via direct contact with an inclusive mainstream school in England, UK. This particular school was chosen because its special education needs resource section had acquired an AI-powered headset to help support BVI children in the development of their social skills. All children attended the same year in the same school. In order to protect the anonymity of participants, the following pseudonyms are used throughout: Teachers: Miss Williams and Miss Thomas [T]; Blind child: Belle [B]; Sighted children: Bobbi [S1], Esme [S2], Margaret [S3], Tiffany [S4].

4.2 AI-powered PeopleLens headset technology and accompanying games

The PeopleLens technology has features that aim to indicate to the wearer where their peers are within a social space. The headset employs a specialised AI tool that is pre-programmed with the faces of up to fifteen people. If it detects that one of these people is in front of the wearer, it provides a knock-knock sound that alternates in pitch and tempo as the face of the social partner orients to face the user. If the headset locks onto a person, it will speak their name. The PeopleLens comes with an accompanying set of games designed to encourage social interaction with increasing difficulty:

Melting ice cream: Placement of the children is established prior to the game's commencement and does not change. The game encourages the initial establishment of a social-spatial map and involves the dual task of tracking an ice cream toy object, as well as who has it so that once the ice cream "melts", the BVI person names who has it.

Who stole the cookie?: The children sit in a circle and clap their hands whilst singing, "Who stole the cookie from the cookie jar?" A pre-selected member of the group names another member of the group, "Esme stole the cookie from the cookie jar", she sings. Esme asks, "Who me?" Everyone replies, "Yes, you!" Esme retorts, "Not me", and everyone wonders, "Then who?!" The song is repeated with Esme now accusing someone else of stealing the cookie. As for the melting ice cream game, this game has a comparably low memory load for the blind child as nobody changes location during the game.

The rescue: A safe zone is established by the group, and the teacher sets it in the application so that the headset can indicate to the user where the safe zone is. The children find a spot in the room some distance from each other and the user of the technology. The wearer then uses the headset to find the other members of the group and return them to the safe zone. This game is more complex with regard to updating locations of already searched spaces relative to the safe zone and the blind child's own location within the room.

Everybody here?: The most complex and the most frequently chosen game to play involves the children sitting around a table doing a separate activity, for example, a brick-building activity. The wearer needs to maintain a social-spatial map because the rest of the group moves around the room and talks to each other.

4.3 Procedure

We observed six naturally occurring sessions of social interaction between participants at the school over a period of two and a half months (see table 1 for session content). The sessions were held in a classroom in the school and facilitated by a teacher. At the beginning of each session, the children decided as a group which game they wanted to play. Once the decision was made, the teacher facilitated while the children organised themselves and began to play the chosen game. The teacher retained control of who wore the PeopleLens in each session. We recorded three hours, 16 minutes and 41 seconds of video data. The sessions were very relaxed and had a playful atmosphere, with the main focus being the children interacting with each other, the headset and the JA games. The time wearing the headset varied from 2 to 15 minutes between sessions for Belle and from 1 to 7 minutes for the sighted children. Attendance of the sighted children varied across sessions. However, if Belle was not present, the session did not go ahead. To complement the data collection timeline shown in table 1, below are details of who attended each session, and which game was played.

Session 1: Sighted children present were Bobbi, Esme, Margaret and Tiffany. Facilitated by the teacher, the children chose to play the "rescue" game first, then the "melting ice cream game", both whilst Belle was wearing the headset and again without.

Session 2: Sighted children present were Esme, Tiffany, Bobbi and Margaret. The children first chose "who stole the cookie", followed by "everybody here?". The teacher swapped which child was wearing the headset each time the wearer correctly guessed who had left the room. When the sighted child was wearing the headset, she also wore a blindfold. The blindfold method was repeated across the rest of the sessions. Once "everybody here?" finished, the teacher requested the children sit around the table to choose a new

Table 1: Table showing data collection timeline.

Session	Date	Duration of session	Games played wearing technology	Games played no technology
1	25th Jan	18m, 39s	The rescue (7m 53s) & Ice cream (7m 53.8s)	The rescue (13m 35.8s) & Ice cream (3m 30s)
2	8th Feb	30m, 5.8s	Everybody here? (55s) & the rescue (4m 47.8s)	Cookie Jar (4m 6s) & Everybody here? (8m 7.8s)
3	22nd Feb	32m, 34.3s	The rescue (6m 37.8s)	The rescue (4m 47.8s)
4	29th Feb	37m, 5.9s	N/A	Everybody here? (21m 57.8s)
5	7th March	39m, 5.2s	Everybody here? (8m 32.55s)	Everybody here? 9m 15s)
6	14th March	39m, 13.1s	Everybody here? (15m 5.2s)	N/A

game. They chose the "rescue". They played the rescue with Belle wearing the headset.

Session 3: Sighted children present were Esme, Bobbi and Margaret. The children chose the rescue game and played the game throughout the whole session.

Session 4: Sighted children present were Esme, Tiffany, Bobbi and Margaret. No headset was worn. The teacher chose the game for this session. The group played with magnets whilst also playing "everybody here?". Belle chose not to wear the headset during this session.

Session 5: Sighted children present were Tiffany and Bobbi. The group played with magnets whilst playing "everybody here?" again. Belle alternated between wearing the headset and not.

Session 6: Sighted children present were Esme and Bobbi. The children chose to play with magnets whilst also playing "everybody here?". Belle wore the headset for most of the session.

4.4 Ethical Considerations

This work was conducted by researchers with backgrounds in Education, Human-Computer Interaction, Industry, Psychology and Social Care. The team has personal and professional ties to the BVI community. Learning, reflecting and holding flexible views in light of new knowledge is prioritised in the team's approach. Our collective experiences mean that we apply our personal and professional values, ethics and knowledge to work reflectively with BVI and sighted children and adults. Ethical approval was gained from the University of Bristol Ethics Review Board. Consent was gained from the parents of the children. Assent was gained from the children prior to the study and, with the help of the teachers, re-established periodically throughout [22, 91]. The sessions were held in a separate classroom on the school site and were facilitated by staff specialised in supporting mixed ability groups of children. In a closing feedback session, and mentioned in a follow-up email from one of the teachers, the children expressed a desire to continue with the games introduced during the research after the research study had been completed, which suggests that measures taken to prevent participants from psychological harm were managed appropriately [114, 130].

4.5 Data analysis

To understand the theory-driven concept of JA as a key aspect of group interactions, we conducted a longitudinal empirical case study using interaction analysis of video data of a group of mixed visual children and their sighted teachers in sessions rich with JA opportunities collected by the first author. The first author visited the school on a weekly basis, set up the cameras, and then left whilst the sessions took place. To minimise disruption to the group, the first author did not remain in the classroom during the sessions, which meant that the data were obtained in as natural a way as possible. The video data were analysed using a systematic research technique for coding sequential naturally occurring interaction behaviours, namely an interaction analysis approach [48, 69, 71]. All data were stored securely in the institution's secure data storage facility. The raw data remain confidential, were only looked at by the research team; and only fully anonymised data will be made publicly available. Following these guidelines, the first author watched the videos, transcribed them and familiarised themselves with the data before collaborating with the other authors to iteratively create the codes. The codes were created using inductive and deductive coding. We started with the deductive approach, applying codes established in previous research on JA between BVI and sighted people [68]. These codes focused on the engagement in JA by identifying three primary behaviours, namely:

- Verbal engagement with JA object or person: Using verbal sounds to engage with the object or person.
- Orientation toward JA partner: Looking at, turning head, body, reaching for, touching, holding partner.
- Orientation toward JA object: Looking at, turning head, body, reaching for, touching, holding object.

We expanded the above codes through inductive coding to include a further focus on the initiation of JA. The inductive coding was driven by a distinction between initiation of JA and engagement within the JA event itself. According to previous literature, JA initiation requires a prompt and a response [94]. We thus set out criteria to clarify that for a JA event to occur, the following needs to happen:

1. One member of the group sends out a request to a potential JA partner.
2. The partner replies in a way that the first person can access.
3. The first person acknowledges the reply provided by the initiation partner.

Only when these three actions have taken place and move into a JA episode, can the initiation of JA be acted upon by the JA partners. If any of these three points don't occur, then the initiation does not lead to a JA event. The person who made the initiation attempt, or someone else, needs to try again. This would be considered a new attempt to initiate JA. Based on the understanding that BVI people experience fewer JA opportunities when in a group of sighted people than any of the sighted people amongst each other [68], and in a bid to identify why there are fewer opportunities for these mixed visual groups, we split the codes into the most foundational element with the aim to build from there. Namely, whether an initiation attempt was made, who it was made by, whether the attempt led to a JA episode, and what might have influenced this outcome. Once the initiation attempts were identified, we explored what was happening within each attempt. As a result, we coded two levels of codes: Level 1 codes identified who made the attempt and whether it led to a JA episode. Level 2 codes identified the actions that were taken by the JA partners within individual JA initiation attempts. For detailed descriptions of the level 2 codes, please refer to (table 2). No codes focused on the effect of the headset being worn by the sighted people, because it was worn by them for playful purposes rather than functionality. The video analysis coding and inter-coder reliability were conducted in NVIVO [82]. The analysis of the inter-coder reliability results was conducted in SPSS [65]. All graphs were created in R [122]. The transcripts were written using the transcription glossary provided in Hutchby and Wooffitt (2008) [64].

4.5.1 Inter-coder Reliability. Inter-coder reliability was conducted to assess the coding frame's rigour and how well it can be applied to the data [112]. The first author coded all codes in the first instance, which were reviewed and discussed with the second and third authors. 166 initiation attempts were identified and coded. The first author then coded the actions taken within those initiation attempts. To gain a clearer idea of how reliable these action codes were, two members of the research team not otherwise involved used the code book to code 30% of the data. We calculated Cohen's Kappa to test for inter-rater reliability and then computed the mean to obtain Light's kappa [36, 59, 78]:

- First author and Coder 2: Cohen's $\kappa = .466$, p <.001, moderate agreement
- First author and Coder 3: Cohen's $\kappa = .632$, p <.001, moderate agreement
- Coder 2 and Coder 3: Cohen's $\kappa = .408$, p <.001, moderate agreement

The mean of the above Cohen's kappa provided Light's kappa, which showed that moderate agreement between the researchers' judgements remained, $\kappa = .502$, p < .001.

5 Results

Figure 2 summarises attempts to initiate JA by a) Belle, the BVI child; b) Miss Williams or Miss Thomas, the teachers; and c) Bobbi, Esme, Margaret and Tiffany, the sighted children, for situations with technology use by Belle as compared to situations without technology. With the exception of the teachers who initiated substantially more JA events with the sighted children when technology was used by Belle, technology use consistently reduced the amount of JA attempts within the mixed vision group. Overall, the sighted people made more initiation attempts than Belle. However, this was expected given that at any one time, there were 3-5 sighted people in the session, whereas Belle was the only blind person. When compared to each group member individually, Belle actually made more initiation attempts than each of her sighted peers. The sighted participants' attempts to initiate JA with Belle tended to continue into a JA episode more often than if Belle made an initiation attempt. No relationship was observed between the frequency of initiation attempts leading to JA episodes and the session number. This indicates that practice did not increase the efficiency of initiation attempts over time.

5.1 Actions taken during JA initiation attempts

We next examined the different actions taken by the group (see figures 3 and 4) when they attempted to initiate JA with one another. To simplify the analysis, we treated Bobbi, Esme, Margaret and Tiffany as a single unit, aggregating all JA attempts by any sighted child towards Belle whilst ignoring any JA initiations between sighted children. We established what sequence of events the successful initiation of JA took when led by the entire group, both when Belle was wearing the headset and when she was not. Overall, initiation attempts were dominated by speech across all three steps of the process. Touch and auditory cues were also used but far less frequently. Sighted children had a greater variety in the way they initiated JA, possibly due to the fact that they had access to more information in the environment with dynamic updates possible at a glance [111].

5.1.1 Actions used by Belle to initiate JA. As shown in figure 3, the blind child Belle used speech as the dominant action, followed by orienting toward the JA object with either her body or her hands. The third most prominent action was to initiate JA in response to an environmental cue, such as a knock at the door.

5.1.2 Actions used by sighted peers Bobbi, Esme, Margaret and Tiffany during JA initiation attempts with Belle. As shown in figure 4, the sighted children also relied on speech as an action to initiate JA with Belle. At the same time, they used typically visible actions such as orienting themselves toward Belle with gaze, head, body or hands, thus showing behaviour they would usually use with sighted peers to initiate JA, but that Belle could not perceive. Furthermore, the sighted children clearly oriented themselves toward the JA objects in any given situation.

5.2 Impact of situational factors on JA initiation

While the likelihood that an initiation attempt led to JA did not seem to change over time, irrespective of who initiated JA, the presence of the headset technology seemed to matter. To explore this, we examined the interactions across the games (see section 4.2), which varied in terms of situational complexity and identified what appeared to impact the efficiency of JA initiation:

5.2.1 Low situational complexity. In one instance (see figure 5), Belle hangs up her white cane and seeks out her chair. Tiffany watches her and offers guidance where she can, which Belle accepts. This is the easiest type of scenario because there is only one static, concrete JA object (i.e. the chair), only Belle is in motion, Tiffany and Belle are in proximity to each other, and Tiffany's attention is focused on Belle with no distractions.

Table 2: Level 2 codes: Actions taken by joint attention (JA) partners within the initiation attempts

	Code	Description
1	Auditory feedback from the technology	The headset gives the wearer auditory feedback regarding people placement, this is received passively and not actively sought by the wearer. This also takes into account when auditory feedback from the headset was audible when someone else was wearing it.
2	Environmental auditory cues	Such as knocking on a door or table, playing a keyboard or a wind chime.
3	Nonverbal gestures	Such as hand waving, clapping, shrugs or facial expressions.
4	Orientation toward JA object	Turning their head or body towards the JA object.
5	Orientation toward JA partner	Turning their head or body towards their JA partner.
6	Quietening	The child does not respond to a question from their peer or becomes still in response to a question or auditory cue such as a knock on the door.
7	Speaking to JA partner(s)	For example, the child asks or answers a question, makes a sound, or verbally grabs their peers' attention and vice-versa.

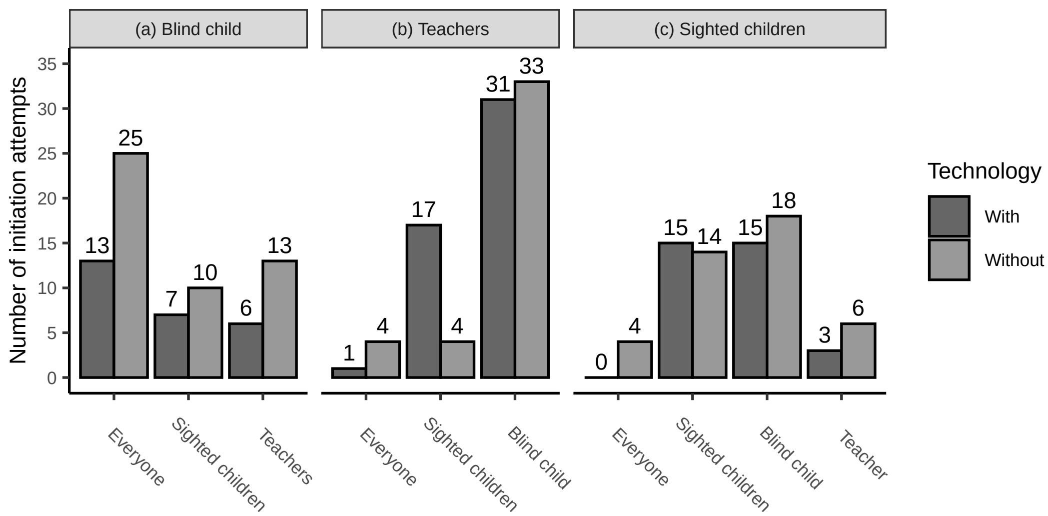


Figure 2: Number of attempts to initiate joint attention with different partners by (a) Blind child, (b) Teachers, (c) Sighted children, whilst the BVI child was wearing the technology (dark grey bars) compared to when she was not (light grey bars).

5.2.2 Medium situational complexity. We observed instances where initiating JA seemed to get more challenging as the environment became more dynamic. This was seen, for example, in the scenario depicted in figure 6. In this scenario, the children were engaged in playing with pegs and pegboards facilitated by the teacher, Miss Williams. Margaret had chosen to fill her board with all orange pegs, and there was some conflict among the sighted children about this. Belle tried to initiate JA by addressing the entire group but expecting Esme to reply. However, Esme was attending to Bobbi and Tiffany, not reacting to Belle's initiation attempt. Instead, Margaret replied, which seemed to surprise Belle, so that she did not react to Margaret but reinitiated JA by addressing Esme directly. In other words, Belle has spoken to the group as a whole and then again to an individual person. This scenario is thus difficult for Belle as she cannot easily find out what her peers are attending to, even though Margaret is next to her, trying to fill her in on some of what is happening around her. Despite the feedback she gets from Margaret and the lack of response from the other children, she is still not able to understand the situation around her and how this might affect her ability to initiate JA with her intended interaction partner. Task-switching (i.e. moving attention from one's own task to understand the situation around oneself affecting potential interaction partners) required to initiate JA in such a way that it moves into a JA episode is reliant on a number of factors that, at this time, Belle does not have enough access to. In more general terms, the factors impacting the environment include a number of potential JA partners and a number of concrete JA objects.

5.2.3 High situational complexity. We observed instances where it was extremely hard to initiate JA. In one such instance (figure 7), the children were playing the "everybody here?" game (4.2) at the same time as playing with magnet sheets and building shapes. In this scenario, it was Belle's turn to wear the headset.

Despite the auditory cues around her and the technology she was wearing, Belle remained oblivious to her friends leaving the room. She talked into the void and only became aware of it when a

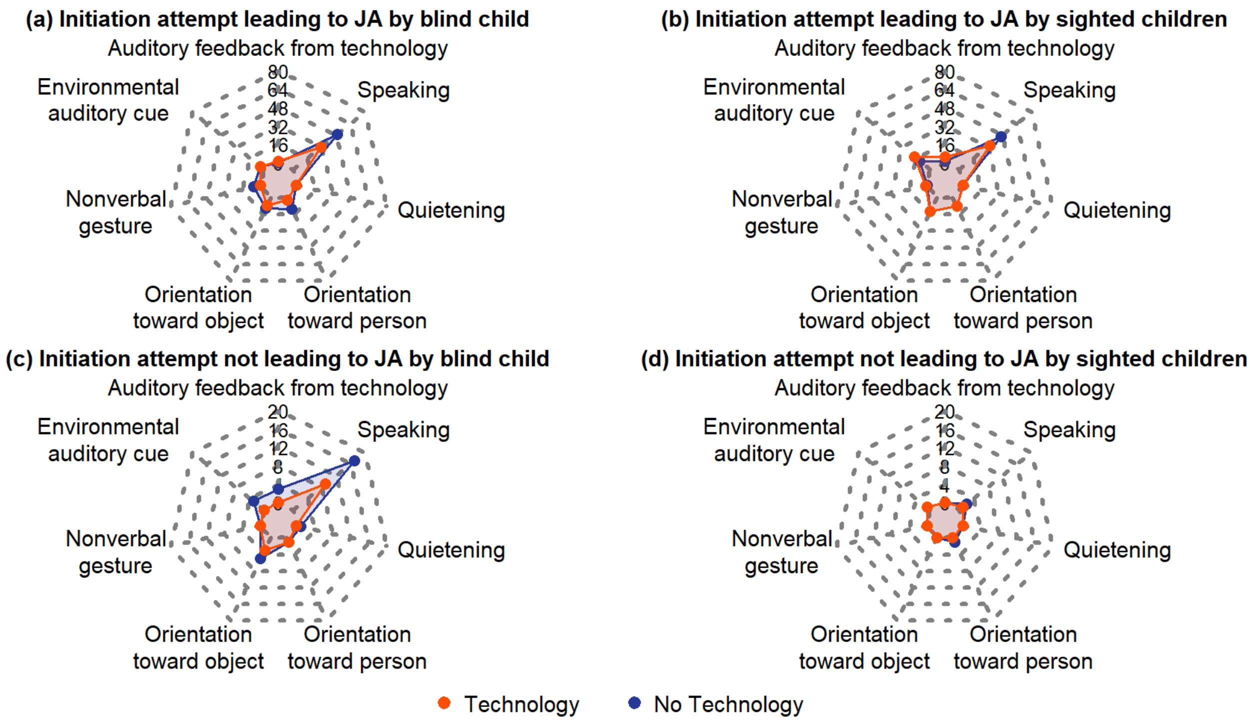


Figure 3: Radial charts showing the frequency of each action (table 2) used by the blind child during JA initiation attempts: leading to JA by the blind child (a), by sighted children (b), not leading to JA by the blind child (c), by sighted children (d).

familiar auditory cue, the knock on the door, grabbed her attention. Here, the factors impacting the environment include a change in the number of potential JA partners, a number of concrete JA objects, and an abstract game involving movement that affects the number of potential JA partners available. Furthermore, Belle has the additional dual-task switching situation of wearing the headset, and no one is updating her about the environmental changes happening around her in the same way that Margaret did in the example above for medium situational complexity.

5.3 A Conceptual Framework of the Complexity Levels of JA Initiation

In Figure 8, we present a hierarchical framework that aims to visualize the different levels of situational complexity identified in section 5.2. We do this to clarify the challenges mixed visual groups might experience in achieving and maintaining JA based on aspects necessary for interdependence. The framework demonstrates that the initiation of JA becomes more challenging for the BVI person than it does for the sighted person as the situational complexity of the environment increases because it requires awareness of situational dynamics. As shown in figure 8, the low complexity scenario has the highest chances of success. This environment has few moving parts, with just one concrete object or game, such as a ball. Both BVI and sighted partners are focused on the same object and each other. Each initiation attempt by either partner is likely to succeed because there are no external distractions interrupting attention, and they remain in the same location.

Referring back to figure 8, a medium situational complexity scenario shows a divergence in the chance of success experienced by BVI and sighted people. The situation becomes more complex as the dynamic changes within the environment increase. For example, an additional concrete object, such as a jigsaw puzzle, is introduced, which divides the attention of the JA partners toward different objects or tasks. A second contributing factor is an increase in potential JA partners. This means that the environment will start changing faster, and there will be more options with whom to initiate JA. The visual information received by the sighted members of the group provides rapid updates on who is where, and what individual people are doing. However, as the BVI child does not have instant access to updated information, they are more likely to make initiation attempts at a time when others are focused on something else and are thus not aligned with the BVI person's focus of attention. Whilst sighted people may experience a slight drop in the likelihood of an initiation attempt leading to a JA episode, for the BVI child, this reduction is far more notable.

A high complexity scenario shows a significant divergence in the chances of success experienced by BVI and sighted people (see figure 8). In highly complex social situations, there are a number of both stable and dynamic features atHi Pet play. Consider a birthday or
work party with different types of games, in an unfamiliar space. People are moving in and out of rooms, and chatting with various

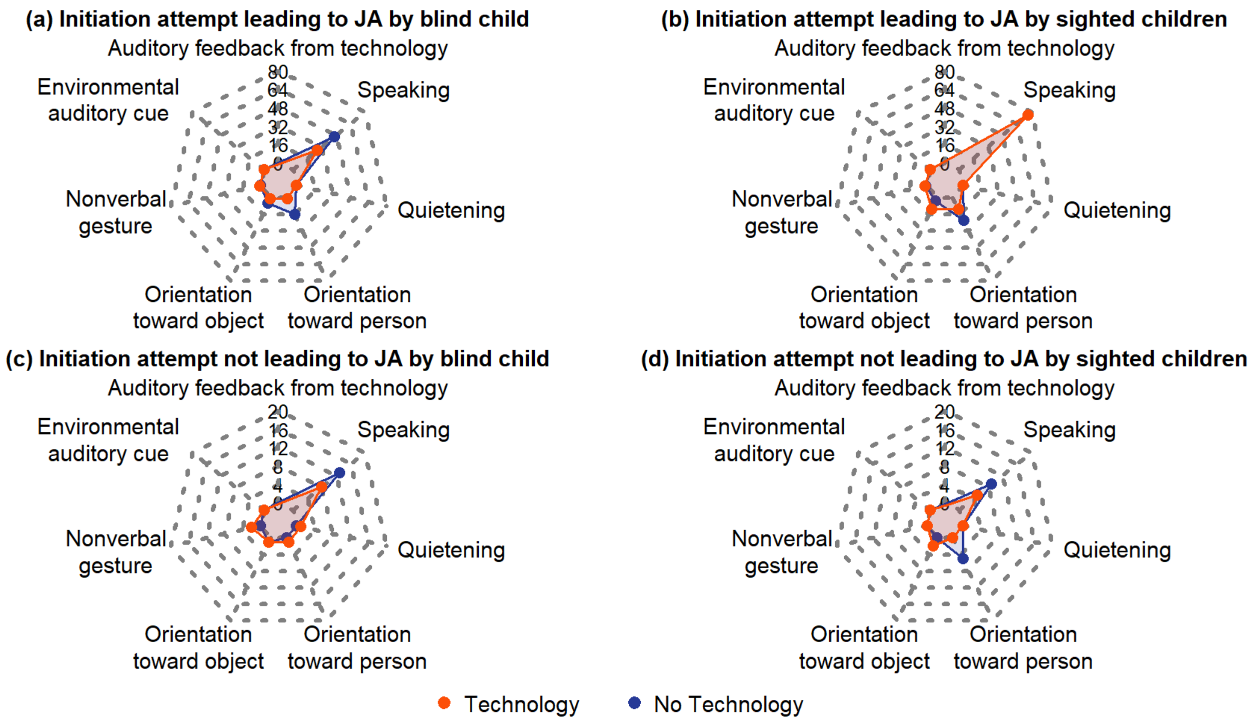


Figure 4: Radial charts showing the frequency of each action (Table 2) used by the sighted children during JA initiation attempts: leading to JA by the blind child (a), by sighted children (b), not leading to JA by the blind child (c), by sighted children (d).

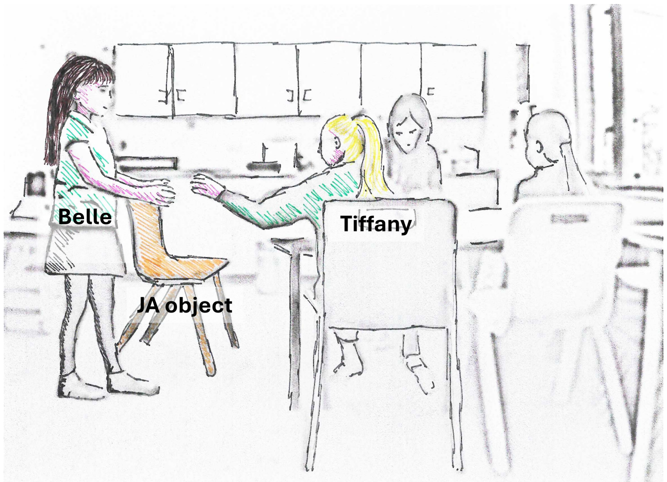
friends or colleagues. There are multiple games in play that everyone can take part in. There is music playing, just loud enough for conversation to be difficult. In environments such as these, sighted people are able to continuously update their environment through a single glance, knowing their JA options within microseconds. Their initiation attempts may not always lead to JA, but they are able to assess the outcome quickly and move on. The social situation is tricky, and everyone is task-switching and likely to experience cognitive overload to some extent. Therefore, they are only able to manage their own experience of the event and are more likely to fail to update a peer on what is going on around them. A BVI person might be making attempts to initiate JA with people they still assume are in their vicinity. These attempts are not being heard because of, for example, noise levels or inattentional blindness [84]. The person they are attempting to engage with might not even be in the room anymore, but the BVI person has no way of knowing. Each time they make an attempt to engage in a new JA event, they are taking a risk. Although on the surface, the party is one of interdependence, in actuality, the autonomy needed to be able to navigate the dynamically changing environment is not available to the BVI person, and so their attempts to initiate JA with a partner of their choice are less likely to be successful.

6 Discussion

In this paper, while considering models of assistive technology such as those discussed by Hersh (2008) [61], as well as technological designs like the ABBI, PeopleLens, or Meta Ray-Bans [87, 100, 155], we build upon previous research that examines the challenges of interdependence experienced by children in groups, particularly in mixed visual ability settings [68, 97, 151]. Our focus is on identifying the challenges faced by children with mixed visual abilities while engaging in joint attention (JA) rich games and investigating how these challenges are influenced by the use of an AI-powered headset (the PeopleLens), which is designed to support the identification of social partners.

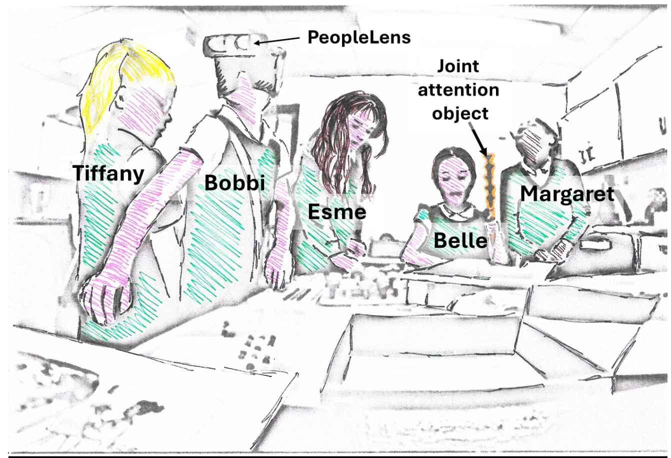
Using interaction analysis of videos, we found that the initiation of JA between children with different visual abilities is shaped by the situational complexity of social interactions, which differentially impacts both the frequency of initiation and the transition into JA episodes for blind and sighted children. Specifically, four main environmental characteristics were identified as influencing the fluidity of engagement: the number of potential JA objects, the type of JA activities, the number of potential JA partners, and the overall stability or dynamics of the JA environment.

Finally, while we observed that the technology has the potential to alleviate access challenges, its usefulness is diminished by the requirement for the blind child to actively seek information from the device, making it an additional cognitive task. Rather than promoting interdependence by equalizing opportunities for sighted and blind JA partners, it inadvertently exacerbates the existing inequality. In the following sections, we discuss these two main findings under the overarching topics of 'maintaining situational awareness in increasingly complex situations' and 'design insights,'



	Participant	Action	Level 2 code
01	Belle	((Reach: looking for her chair))	4
02	Tiffany	((Gaze))	3
03		((Reach: toward Belle's hands))	5
04		((Touch: takes hold of Belle's hand))	5
05		Your chair is. here	7
06	Belle	((Touch: allows Tiffany to guide her hand))	5
07	Tiffany	((Touch: pulls Belle's hand down so it touches the chair))	4
08	Belle	((Lets go of Tiffany's hand and sits))	

Figure 5: Vignette depicting an example of low situational complexity (see 5.2.1) : Above: Corresponding video illustration, Below: Transcript and corresponding level 2 codes (see also, table 2).



	Participant	Action	Level 2 code
01	Belle	Guys, put your hand up (.) put your hand up. who wants this tow↓er. ((Belle says to the room))	7
02	Margaret	Not me:?	7
03	Belle	Do you want it Esme.	7
04	Belle	((Belle leans over and sweeps her hand across the empty seat))	5

Figure 6: Vignette depicting an example of medium situational complexity (see 5.2.2): Above: Corresponding video illustration, Below: Transcript and codes (see also, table 2).

before addressing limitations and suggesting directions for future research.

6.1 Maintaining situational awareness in increasingly complex social situations

As we have demonstrated in sections 5.2 and 5.3, the opportunities for mixed visual groups to maintain situational awareness in JA-rich activities are skewed in favour of sighted children as the environment becomes more dynamically complex. Research has previously confirmed that greater perceptual sensitivity supports more complex play for sighted children and that the less attention that can be paid to social partners, the more reliant children are on a pre-established schema [13, 157]. It has also long been understood that blind children have to build up a schema of their environment through tactile and auditory senses and that, even when used together, touch and sound do not capture all the information that might be required for dynamically changing social interactions at all times [79]. This knowledge leads us to understand why language development, friendships and interdependence, as well as autonomy and independence, are so important when considering the initiation of JA activities in mixed visual groups [99, 109, 141].

This increase in situational complexity increases the likelihood of disruption to the three-step process of initiating JA between BVI and sighted people [29]. We established that for the mixed visual ability group, the three steps to successfully initiate JA, namely an initial hail of some kind, a response to that hail, and an acknowledgement of that response by the instigator, consisted primarily of speech as a modality of interaction. This substantially restricted more subtle turn-taking scenarios that are an important part of sharing attention in social interactions [137]. For sighted people, this turn-taking does not usually rely on speech [38, 67], but on eye gaze and visual signalling, to know when it is one's turn to speak [159]. For the BVI child, turn-taking is difficult because she does not have access to the visual cues to know when her turn is, and the place holding auditory cue that she used to procure attention is not necessarily utilised or recognised as being important by others [85]. Together with the lack of situational awareness, the blind child, therefore, might accidentally interrupt others' social interactions, or their initiation attempts might go unnoticed.

A lack of situational awareness is exacerbated by the dual demands of tasks and group dynamics. In other words, as demonstrated by our results, the primary issue for the blind child is the inability to quickly update their understanding of the surrounding environment when shifting attention from one focus to another, a

	Participant	Action	Level 2 code
01	Belle	((There is a knock on the door. Belle tilts her head in response.))	2
01	Belle	Ms Williams	7
02	Ms Williams.	Yea:h?	7
03	Belle	Did you tap somebody (.) somebody on the head↓	7
04	Ms Williams	What. to say they could go:?	7
05	Belle	Yea:h	7
06	Ms Williams	Well, did I? Is everyone her↑e or not.	7
07	Belle	Hello? ((she asks the room))	7
08	Belle	((After no reply she muses)) I feel like (.) both of them left.	7

Figure 7: Vignette depicting an example of high situational complexity (see 5.2.3): Above: Corresponding video illustration, Below: Transcript and level 2 codes (see also, table 2).

skill essential for fluid social interactions [161]. Social situations tend to grow more complex during childhood development [14]. The decline in the success rate of initiating JA observed with increasing situational complexity, is likely attributable to the heightened cognitive load placed on the blind child. Specifically, this refers to the amount of working memory required to track changes in the (social) environment without the benefit of visual dynamic updating available to sighted children. The more challenging and time-intensive this updating process becomes, the lower the likelihood that the child will respond to JA bids from sighted peers or send appropriate JA requests out [116].

A notable characteristic of more complex environments is the divergence of tasks performed by potential interaction partners. This divergence necessitates task switching for the initiators of JA, who must update themselves on environmental changes. For a blind child, this task is significantly more challenging, as environmental updates can only be obtained sequentially through active non-visual "scanning." Unlike sighted children, who can process visual updates in milliseconds to capture the gist of a scene, blind children require several seconds for such updates [110]. This delay reduces the opportunities for a blind child to engage in JA with their peers.

Moreover, failures to initiate JA opportunities may stem from the inaccessibility of signals provided by sighted children, such as nonverbal gestures. These cues, effective for those with vision, are not accessible to blind children [106]. This challenge may be further compounded by the potential misinterpretation of cues from the blind child. For instance, a blind child might inadvertently provide misleading verbal prompts to sighted peers, such as suggesting they raise their hands, despite being unable to perceive such nonverbal responses.

6.2 Design insights for adaptive technologies supporting inclusive joint attention in mixed visual ability groups

In conversations with the participant group, we discovered that Belle is articulate, friendly and chatty and takes a keen interest in the people around her and her environment. She spends her lunch and play breaks in school with her friends and is very sociable. Her lessons are somewhat different from those of her sighted peers; she has a big braille machine that takes almost half of her desk and a plethora of physical objects to learn with. She often starts her lessons in the main classroom with everyone else in her class. However, she expresses that the classroom environment is often too noisy, and she can't concentrate on her work with her teaching assistant because there is too much going on around her. Prior research has dubbed this situation as leading to a potential "assistance bubble" [97]. Perhaps as a result, Belle tends to request to spend many of her lessons in a room away from her classmates, where she is alone with her teacher. However, for the PeopleLens sessions, Belle asked that they continue. She expressed how much she enjoyed using the headset, and equally enjoyed her friends using it. In fact, the whole group expressed how much they enjoyed and wanted to continue the sessions. The aim of the PeopleLens was to support more opportunities for BVI children to engage with their peers, and in this instance it succeeded. It is with this consideration in mind that we explore ways that future designs could be adjusted and adapted to support interdependence among mixed visual groups of people across the whole range of vision.

The key insight of this work for the design of technology is, that to be useful, it will need to adapt to the fluidity of social situations. We suggest that this could be captured through the following design insights:

Consider the complexity of the environment. If initiation attempts are being made in environments with low situational complexity, the need for technology to support the initiation between mixed visual JA partners is not there in the same way as it is for a dynamic situation involving a group of people engaged in multiple

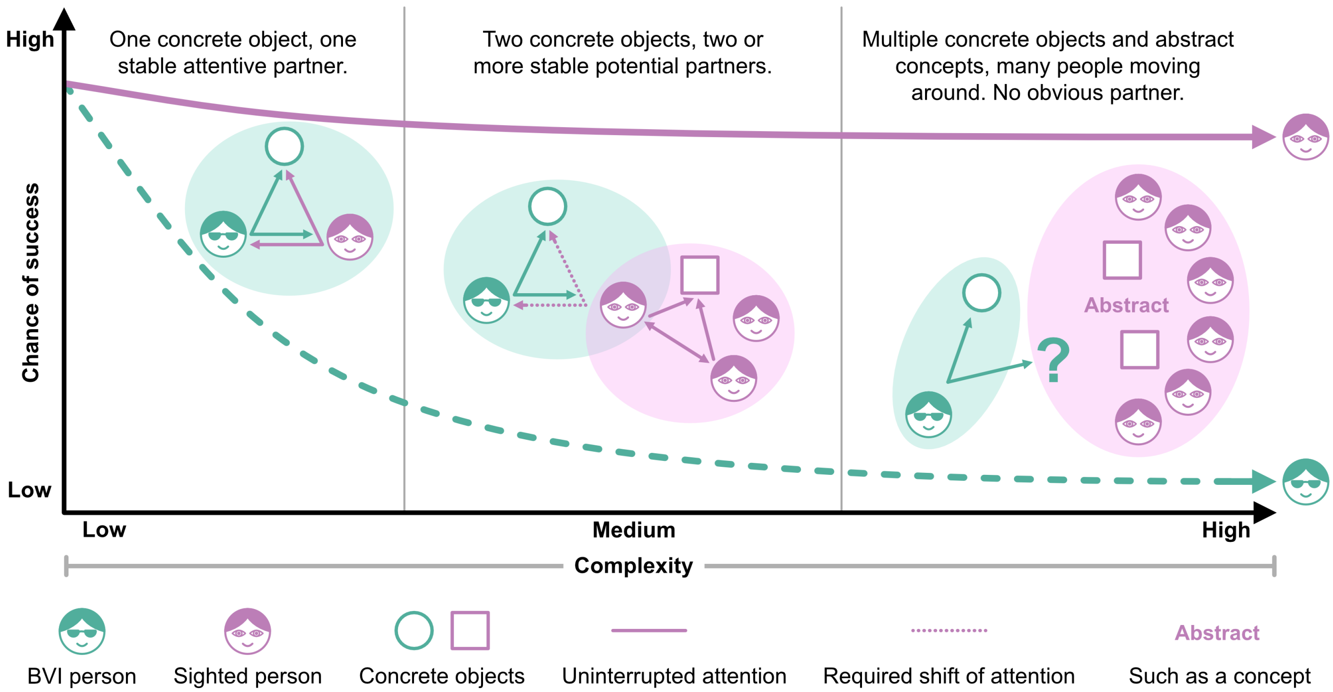


Figure 8: Line graph with integrated images showing the different elements of initiating JA, with the opportunities for a BVI person diminishing with each increase in situational complexity.

JA potential roles. In medium-complexity environments, the BVI person could potentially be updated regarding changes to the environment by their sighted JA partner. Also, current models such as the vOICe or SmartVision [46, 66, 92, 93] could be used to update the BVI person about their environment if desired. It is as the dynamic complexity of the environment increases that mixed visual groups would benefit more from technology that could facilitate the initiation of JA. In more complex situations, the technology could update on request the BVI people on environmental changes important for their social interaction intent and any nonverbal information they might be missing. This could, for example, include information about the current location and focus of attention of their intended JA partner. At the same time, the sighted people should be updated or signalled when the BVI person makes an initiation attempt in a way that they might not have perceived. Furthermore, there could be a feature that alerts the sighted potential JA partners to inform the technology of bigger or more dynamic changes in the environment, such as them leaving the room or a new person entering.

Consider what constitutes "relevant" information, then provide relevant updates on environmental changes. The use of the PeopleLens was an insightful first step into figuring out how one could increase situational awareness in the blind child. The PeopleLens currently relies on sequential (undirected) active scanning of who is present and where in the environment. This could be taken to the next step by exploring how one might be able to enable extraction of information about the *relevant* changes in the social environment only, to increase the speed and specificity of the initiation attempt required in busy social settings.

In this context, the term *relevant* is used for the intended interaction. However, an awareness needs to be maintained that some intentions arise only when one knows who is present and what opportunities there are to interact. With that in mind, to understand better what kind of environmental change is relevant and when and how to provide this information requires more research. Prior research has pointed out that being able to tease out which information is relevant is crucial when designing awareness support in non-visual environments since overloading users with too much information can be detrimental when users engage in activities that compete for attention and memory resources [96]. The use of technology is a good step forward from the perspective of the children, as can be seen in the fact that the whole group requested to continue using the PeopleLens after the study was over. Indeed, already at the current stage of technological design, they may be benefiting from the social element of the technology (such as the sighted children gaining more insight into the world of the blind child when being blind-folded and having to rely on the PeopleLens and non-visual cues) in a way that we could not derive from this study.

Design for discretion, to reduce distraction from the main JA activity. As highlighted before, if not used carefully, technology becomes a complicated dual task too slow to be of use and too difficult to manage with regard to memory load. Instead of providing information indiscriminately through active scanning, the next stages in technology development should focus on tackling the problem of how one can enable access to information relevant to increase social interaction-specific situational awareness only [96]. Such filtered information would ensure that any potential negative side effects of information overload could be avoided, helping the blind child with object-based and people-based task switching. We are mindful of previous research that found blind children can experience a void when attempting to interact with others seeming to become invisible to their friends due to not sending out the signals the friends expect [68]. Based on this understanding of a "void", a second important factor to explore is how the blind child can signal interaction intents to sighted potential JA partners without having to use language, and how sighted children's visual signals can be translated into useful signals for the blind child. This might require two or more separate pieces of technology or a multi-part piece of technology that is linked and usable by all potential JA partners.

Consider both independence and interdependence when supporting the establishment of the JA triad. Our observations revealed that the blind child enjoyed it when her friends used the AI-powered technology. The group as a whole seemed to benefit from learning about each other's different experiences, fostering a deeper understanding among members. Shifting the responsibility of using the technology from the BVI child to the sighted children could reframe the dynamic, emphasising the collective efforts of the group rather than placing the burden on an individual. When reflecting on interdependence in the initiation of JA, it is important to acknowledge that any technology design must account for the entire JA triad. This triad includes not only the BVI and sighted individuals but also the potential JA objects, concepts, and the technology that mediates their interactions [11]. A balanced approach should consider the individual's independence and autonomy within the group [76], the group's interdependence with one another [154], and the interplay between the individual, the group, and the JA objects or concepts [164]. Such an approach could effectively address the concerns outlined above.

Once again, one size does not fit all. Vision and vision loss occur on a spectrum, and the initiation of JA, and JA itself have a wide range of complexity. As a result, designers might want to consider that one piece of technology might not be sufficient to meet the needs of everyone in all situations. We recognise that this proposal is not a new one [6, 63, 136]. However, our conclusions and the proposed framework of JA initiation that is grounded in them, coincide with the premise that adaptable technology that users can pick up and put down at the appropriate moments would be of more use than trying to fit a square peg into a round hole [132]. In other words, trying to fit a single device into such a multi-faceted, constantly changing phenomenon, such as inclusive JA, does not justifiably benefit all members of a mixed visual group.

6.3 Limitations and future work

The work in this paper focuses on the mechanics of the initiation of JA between BVI children and their sighted peers. As this is a little-researched area, there are some limitations in the study, which show a need for further investigations. This piece of research was a case study of one blind child and her sighted peers, as we wanted to delve deeply into the social interaction dynamics of a group of mixed visual children in a mainstream school. We decided to keep the sample small, choosing depth over breadth, which has led to further questions for future research [141]. Further research with larger samples, both within and across groups would be valuable to develop our understanding of the initiation of JA and JA itself in mixed visual groups. Furthermore, we recognise that the children were recorded in a classroom with a facilitator, with tasks focused either on the technology or on JA-rich scenarios, thus giving plenty of opportunities for the children to initiate JA with each other. Future studies could focus on observing children in other, less controlled situations that don't have a facilitator or include strangers to see if this affects the initiation attempts. The study concluded with short semi-structured interviews to gauge the participants' opinions. We acknowledge that closing questionnaires could also add to the richness of establishing the perceptions of participants in future studies. We also consider the fact that the sample was imbalanced in that it had more sighted children than blind children in the group. Future work could explore the initiation of JA in peer groups with a more balanced dynamic between BVI and sighted children. We would also appreciate more work exploring JA experiences for BVI adults in various settings, including groups of friends, family and strangers. Further research would approach inter-coder reliability more robustly to include the coding process as it matures through the analysis process. An interview study to consider the opinions and experiences of BVI and sighted adults with regard to initiating JA in mixed visual groups is a valuable next step to place lived experiences at the forefront of the research. Additionally, participatory design workshops would greatly benefit our understanding of how to implement the design insights we proposed for the initiation of JA in mixed visual social situations. These future research proposals will be the next steps in providing a thorough analysis of the whole JA triad, including both BVI children and BVI adults, their sighted peers, friends, and strangers.

7 Conclusion

In this paper, we observed one blind child, four sighted children and two sighted adults initiating and engaging in JA-rich activities whilst using wearable technology designed to support these scenarios. We explored previously identified issues around initiating JA through interaction analysis. We looked at the initiation of JA from the sighted person and the blind person's point of view to be certain that the suggestions we make around the increasing difficulties of the initiation of JA are representative of mixed visual groups, with a specific focal point on the blind child. We focused on understanding the blind child's experience when initiating JA so that we can confidently identify what support BVI children and their sighted peers need to initiate JA episodes successfully in increasingly complex levels of social interactions they might encounter. We found that the more activity around the blind child, the less likely they were to successfully initiate JA with their sighted peers. We propose, therefore, that technology be designed to adapt to task-relevant changes in the environment so that it can support variations between interactions in mixed visual groups.

Acknowledgments

The authors would like to thank the participants for their engagement in this research. We would also like to thank the reviewers for their insightful and valuable feedback. Finally, we would like to extend special thanks to Matt Sutton, Mike Taverne and Rachel Keys for their valuable insights and contributions to the paper. This research was funded by a UKRI iCase Studentship and ERC-selected UKRI-funded project inclusiveXplay, Grant no.: EP/Y023676/1

References

[1] Drew H Abney, Linda B Smith, and Chen Yu. 2017. It's Time: Quantifying the Relevant Timescales for Joint Attention. In *Proceedings of the 39th Annual Meeting of the Cognitve Science Society*. Cognitive Science Society, London, UK, 1489–1494.

[2] Lauren B Adamson, Roger Bakeman, and Deborah F Deckner. 2004. The development of symbol-infused joint engagement. *Child development* 75, 4 (2004), 1171–1187.

[3] Lauren B Adamson, Roger Bakeman, Deborah F Deckner, and P Brooke Nelson. 2014. From interactions to conversations: The development of joint engagement during early childhood. *Child development* 85, 3 (2014), 941–955.

[4] Lauren B Adamson, Roger Bakeman, Deborah F Deckner, and MaryAnn Romski. 2009. Joint engagement and the emergence of language in children with autism and Down syndrome. *Journal of autism and developmental disorders* 39 (2009), 84–96.

[5] Lauren B Adamson, MaryAnn Romski, Roger Bakeman, and Rose A Sevcik. 2010. Augmented language intervention and the emergence of symbol-infused joint engagement. *Journal of Speech, Language, and Hearing Research* 53, 6 (2010), 1769–1773. doi:10.1044/1092-4388(2010/09-0208)

[6] Leila Aflatoony and Siniša Kolarić. 2022. One Size Doesn't Fit All: On the Adaptable Universal Design of Assistive Technologies. *Proceedings of the Design Society* 2 (2022), 1209–1220.

[7] Andrea Urqueta Alfaro, Valerie S Morash, Daisy Lei, and Deborah Orel-Bixler. 2018. Joint engagement in infants and its relationship to their visual impairment measurements. *Infant Behavior and Development* 50 (2018), 311–323. doi:10.1016/j.infbeh.2017.05.010

[8] Roger Bakeman and Lauren B. Adamson. 1984. Coordinating Attention to People and Objects in Mother-Infant and Peer-Infant Interaction. *Child Development* 55, 4 (1984), 1278–1289. http://www.jstor.org/stable/1129997

[9] Lucas Battich, Merle Fairhurst, and Ophelia Deroy. 2020. Coordinating attention requires coordinated senses. *Psychonomic bulletin & review* 27 (2020), 1126–1138.

[10] Andrew P Bayliss, Emily Murphy, Claire K Naughtin, Ada Kritikos, Leonhard Schilbach, and Stefanie I Becker. 2013. "Gaze leading": Initiating simulated joint attention influences eye movements and choice behavior. *Journal of Experimental Psychology: General* 142, 1 (2013), 76.

[11] Cynthia L Bennett, Erin Brady, and Stacy M Branham. 2018. Interdependence as a frame for assistive technology research and design. In *Proceedings of the 20th international acm sigaccess conference on computers and accessibility*. 161–173.

[12] Cynthia L Bennett, Daniela K Rosner, and Alex S Taylor. 2020. The care work of access. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–15.

[13] Laura M Bennett Murphy, Cynthia Laurie-Rose, Tara M Brinkman, and Kelly A McNamara. 2007. Sustained attention and social competence in typically developing preschool-aged children. *Early Child Development and Care* 177, 2 (2007), 133–149.

[14] Thore J Bergman and Jacinta C Beehner. 2015. Measuring social complexity. *Animal behaviour* 103 (2015), 203–209.

[15] Alexy Bhowmick and Shyamanta M. Hazarika. 2017. An insight into assistive technology for the visually impaired and blind people: state-of-the-art and future trends. *Journal on Multimodal User Interfaces* 11, 2 (2017), 149–172. doi:10.1007/s12193-016-0235-6

[16] Alexy Bhowmick and Shyamanta M Hazarika. 2017. An insight into assistive technology for the visually impaired and blind people: state-of-the-art and future trends. *Journal on Multimodal User Interfaces* 11 (2017), 149–172.

[17] Ann E Bigelow. 2003. The development of joint attention in blind infants. *Development and psychopathology* 15, 2 (2003), 259–275.

[18] Ann E Bigelow, Kim MacLean, and Jane Proctor. 2004. The role of joint attention in the development of infants' play with objects.

[19] Anne Böckler, Günther Knoblich, and Natalie Sebanz. 2011. Giving a helping hand: effects of joint attention on mental rotation of body parts. *Experimental brain research* 211 (2011), 531–545.

[20] Monique Botha, Jacqueline Hanlon, and Gemma Louise Williams. 2023. Does language matter? Identity-first versus person-first language use in autism research: A response to Vivanti. *Journal of autism and developmental disorders* 53, 2 (2023), 870–878.

[21] Stacy M. Branham and Shaun K. Kane. 2015. Collaborative Accessibility: How Blind and Sighted Companions Co-Create Accessible Home Spaces. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 2373–2382. doi:10.1145/2702123.2702511

[22] Cassi A Breaux and Kristin Smith. 2023. Assent in applied behaviour analysis and positive behaviour support: ethical considerations and practical recommendations. *International Journal of Developmental Disabilities* 69, 1 (2023), 111–121.

[23] Rachel Brown, R Peter Hobson, Anthony Lee, and Jim Stevenson. 1997. Are there "autistic-like" features in congenitally blind children? *Journal of child psychology and psychiatry* 38, 6 (1997), 693–703.

[24] Donald Bungum. 2013. Joint attention, union with God, and the dark night of the soul. *European journal for philosophy of religion* 5, 4 (2013), 187–210.

[25] George Butterworth and Nicholas Jarrett. 1991. What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy. *British journal of developmental psychology* 9, 1 (1991), 55–72.

[26] Héctor Caltenco, Charlotte Magnusson, Bitte Rydeman, Sara Finocchietti, Giulia Cappagli, Elena Cocchi, Lope Ben Porquis, Gabriel Baud-Bovy, and Monica Gori. 2016. Co-designing wearable technology together with visually impaired children. *International Journal of Mobile Human Computer Interaction (IJMHCI)* 8, 4 (2016), 68–84.

[27] Giulia Cappagli, Sara Finocchietti, Elena Cocchi, Giuseppina Giammari, Roberta Zumiani, Anna Vera Cuppone, Gabriel Baud-Bovy, and Monica Gori. 2019. Audio motor training improves mobility and spatial cognition in visually impaired children. *Scientific reports* 9, 1 (2019), 3303.

[28] Cecily Morrison, Katherine Jones, Martin Grayson, Ed Cutrell. 2022. PeopleLens: Using AI to support social interaction between children who are blind and their peers. https://www.microsoft.com/en-us/research/blog/peoplelens-using-ai-to-support-social-interaction-between-children-who-are-blind-and-their-peers/ [Online; accessed 13-August-2024].

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

[30] Lucia B Chisari, Akvilė Mockevičiūtė, Sterre K Ruitenburg, Lian van Vemde, Ellen M Kok, and Tamara van Gog. 2020. Effects of prior knowledge and joint attention on learning from eye movement modelling examples. *Journal of Computer Assisted Learning* 36, 4 (2020), 569–579.

[31] Mariam Chohan and Emily A Jones. 2019. Initiating joint attention with a smile: Intervention for children with autism. *Behavioral Development* 24, 1 (2019), 29.

[32] Luigi F Cuturi, Elena Aggius-Vella, Claudio Campus, Alberto Parmiggiani, and Monica Gori. 2016. From science to technology: Orientation and mobility in blind children and adults. *Neuroscience & Biobehavioral Reviews* 71 (2016), 240–251. doi:10.1016/j.neubiorev.2016.08.019

[33] Tana D'Allura. 2002. Enhancing the social interaction skills of preschoolers with visual impairments. *Journal of Visual Impairment & Blindness* 96, 8 (2002), 576–584.

[34] Maitraye Das, Thomas Barlow McHugh, Anne Marie Piper, and Darren Gergle. 2022. Co11ab: Augmenting accessibility in synchronous collaborative writing for people with vision impairments. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–18.

[35] Andrew R. Dattel, Jason E. Vogt, Chelsea C. Sheehan, Kristen Madjic, Matthew C. Stefonetti, Marissa C. Miller, Daniel P. Dever, Jesse Brodsky, Sierra Bradley, and Jessica K. Fratzola. 2012. The Effects of Pointing out Failures of Inattentional Blindness on Performance and Situation Awareness. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 56, 1 (2012), 1094–1098. doi:10.1177/1071181312561238

[36] Mark Davies and Joseph L Fleiss. 1982. Measuring agreement for multinomial data. *Biometrics* 38, 4 (1982), 1047–1051.

[37] Johan De Smet and Helen De Cruz. 2013. Delighting in natural beauty: Joint attention and the phenomenology of nature aesthetics. *European Journal for Philosophy of Religion* 5, 4 (2013), 167–186.

[38] Ziedune Degutyte and Arlene Astell. 2021. The role of eye gaze in regulating turn taking in conversations: a systematized review of methods and findings. *Frontiers in Psychology* 12 (2021), 616471.

[39] Nada Dimcovic. 1992. Why is the change in the task difficult for blind children? *European journal of psychology of education* 7 (1992), 231–244.

[40] Brian L Due. 2023. Guide dog versus robot dog: Assembling visually impaired people with non-human agents and achieving assisted mobility through distributed co-constructed perception. *Mobilities* 18, 1 (2023), 148–166.

[41] Dana S Dunn and Erin E Andrews. 2015. Person-first and identity-first language: Developing psychologists' cultural competence using disability language. *American Psychologist* 70, 3 (2015), 255.

[42] William Easley, Michele A Williams, Ali Abdolrahmani, Caroline Galbraith, Stacy M Branham, Amy Hurst, and Shaun K Kane. 2016. Let's get lost: Exploring social norms in predominately blind environments. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 2034–2040.

[43] Mica R. Endsley. 2021. *Situation awareness.* John Wiley & Sons, Ltd, Chapter 17, 434–455. doi:10.1002/9781119636113.ch17 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781119636113.ch17

[44] Mica R Endsley, Daniel J Garland, et al. 2000. Theoretical underpinnings of situation awareness: A critical review. *Situation awareness analysis and measurement* 1, 1 (2000), 3–21.

[45] Be My Eyes. 2020. Be my eyes. *Retrieved May* 29 (2020).

[46] J Faria, S Lopes, H Fernandes, P Martins, and J Barroso. 2010. Electronic white cane for blind people navigation assistance. In *2010 World Automation Congress*. 1–7.

[47] Anika Fiebich and Shaun Gallagher. 2013. Joint attention in joint action. *Philosophical Psychology* 26, 4 (2013), 571–587. doi:10.1080/09515089.2012.690176

[48] Laurent Filliettaz, Stéphanie Garcia, and Marianne Zogmal. 2022. Video-based interaction analysis: A research and training method to understand workplace learning and professional development. In *Methods for researching professional learning and development: Challenges, applications and empirical illustrations*. Springer, 419–440.

[49] Sara Finocchietti, Giulia Cappagli, Gabriel Baud-Bovy, Charlotte Magnusson, Hector Caltenco, Graham Wilson, Stephen Brewster, Ann-Kathrin Rogge, Brigitte Roeder, Elena Cocchi, Elisabetta Capris, Paola Campana, Carla Gilio, and Monica Gori. 2015. ABBI, a New Technology for Sensory-motor Rehabilitation of Visual Impaired People. In *International Conference on Enabling Access for Persons with Visual Impairment*. University of Athens, Athens, Greece. https://access.uoa.gr/ICEAPVI-2015/proceedingsFiles/82-86.pdf

[50] Patrick Flink. 2021. Person-first & identity-first language: Supporting students with disabilities on campus. *Community College Journal of Research and Practice* 45, 2 (2021), 79–85.

[51] Fabia Franco, Paola Perucchini, and Barbara March. 2009. Is infant initiation of joint attention by pointing affected by type of interaction? *Social Development* 18, 1 (2009), 51–76.

[52] Euan Freeman, Graham Wilson, Stephen Brewster, Gabriel Baud-Bovy, Charlotte Magnusson, and Hector Caltenco. 2017. Audible Beacons and Wearables in Schools: Helping Young Visually Impaired Children Play and Move Independently. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 4146–4157. doi:10.1145/3025453.3025518

[53] Devon N Gangi, Lisa V Ibañez, and Daniel S Messinger. 2014. Joint attention initiation with and without positive affect: Risk group differences and associations with ASD symptoms. *Journal of autism and developmental disorders* 44 (2014), 1414–1424.

[54] MC Garzón-Rodríguez, LS Reyes-Figueredo, LÁ Velandia-Rodríguez, OD Méndez-Ruiz, MA Gómez-Rodríguez, LT Esguerra-Ochoa, and D García-Lozada. 2023. Causes of low vision in children: A systematic review. *Archivos de la Sociedad Española de Oftalmología (English Edition)* 98, 2 (2023), 83–97.

[55] Sandra R Gomes, Sharon A Reeve, Kevin J Brothers, Kenneth F Reeve, and Tina M Sidener. 2020. Establishing a generalized repertoire of initiating bids for joint attention in children with autism. *Behavior Modification* 44, 3 (2020), 394–428.

[56] João Guerreiro and Daniel Gonçalves. 2013. Blind people interacting with mobile social applications: open challenges. In *Mobile Accessibility Workshop at CHI*, Vol. 13.

[57] João Guerreiro, Daisuke Sato, Saki Asakawa, Huixu Dong, Kris M. Kitani, and Chieko Asakawa. 2019. CaBot: Designing and Evaluating an Autonomous Navigation Robot for Blind People. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility* (Pittsburgh, PA, USA) *(ASSETS '19)*. Association for Computing Machinery, New York, NY, USA, 68–82. doi:10.1145/3308561.3353771

[58] Judith A Hall, Terrence G Horgan, and Nora A Murphy. 2019. Nonverbal communication. *Annual review of psychology* 70, 1 (2019), 271–294.

[59] Kevin A Hallgren. 2012. Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in quantitative methods for psychology* 8, 1 (2012), 23.

[60] Sharon A Haymes, Alan W Johnston, and Anthony D Heyes. 2002. Relationship between vision impairment and ability to perform activities of daily living. *Ophthalmic and Physiological Optics* 22, 2 (2002), 79–91.

[61] Marion A Hersh and Michael A Johnson. 2008. On modelling assistive technology systems–Part I: Modelling framework. *Technology and disability* 20, 3 (2008), 193–215.

[62] Anna E Holt and Gedeon Deák. 2015. Children's task-switching efficiency: missing our cue? *Journal of Cognition and Development* 16, 2 (2015), 261–285.

[63] Amy Hurst and Jasmine Tobias. 2011. Empowering individuals with do-it-yourself assistive technology. In *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility*. 11–18.

[64] Ian Hutchby and Robin Wooffitt. 2008. *Conversation analysis*. Polity.

[65] IBM Corp. IBM SPSS Statistics for Windows, Version 29.0.2.0 Armonk, NY: IBM Corp. 2023. *SPSS*. https://www.ibm.com/spss [Online; accessed 26-August-2024].

[66] Rui Jin, Matthew A Petoe, Chris D McCarthy, Jaime R Serra, Scott Starkey, Jennifer McGinley, and Lauren N Ayton. 2024. Functional performance comparison of long cane and secondary electronic travel aids for mobility enhancement. *British Journal of Visual Impairment* 0, 0 (2024), 02646196241285098. doi:10.1177/02646196241285098

[67] Kristiina Jokinen, Hirohisa Furukawa, Masafumi Nishida, and Seiichi Yamamoto. 2013. Gaze and turn-taking behavior in casual conversational interactions. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 3, 2 (2013), 1–30.

[68] Katherine Mary Jones, Ute Leonards, and Oussama Metatla. 2024. "I Don't Really Get Involved In That Way": Investigating Blind and Visually Impaired Individuals' Experiences of Joint Attention with Sighted People. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 52, 16 pages. doi:10.1145/3613904.3642940

[69] Brigitte Jordan and Austin Henderson. 1995. Interaction analysis: Foundations and practice. *The journal of the learning sciences* 4, 1 (1995), 39–103.

[70] Lorcan Kenny, Caroline Hattersley, Bonnie Molins, Carole Buckley, Carol Povey, and Elizabeth Pellicano. 2016. Which terms should be used to describe autism? Perspectives from the UK autism community. *Autism* 20, 4 (2016), 442–462.

[71] Joann Keyton. 2018. Interaction Analysis: An Introduction. In *The Cambridge Handbook of Group Interaction Analysis*, Elisabeth Brauner, Margarete Boos, and Michaela Kolbe (Eds.). Cambridge University Press, New York, NY, US, 3–19. doi:10.1017/9781316286302.002

[72] Sulaiman Khan, Shah Nazir, and Habib Ullah Khan. 2021. Analysis of navigation assistants for blind and visually impaired people: A systematic review. *IEEE access* 9 (2021), 26712–26734.

[73] Susanne Klauke, Chloe Sondocie, and Ione Fine. 2023. The impact of low vision on social function: The potential importance of lost visual social cues. *Journal of Optometry* 16, 1 (2023), 3–11.

[74] Bineeth Kuriakose, Raju Shrestha, and Frode Eika Sandnes. 2022. Tools and technologies for blind and visually impaired navigation support: a review. *IETE Technical Review* 39, 1 (2022), 3–18.

[75] Imme Lammertink, Daan Hermans, Angela Stevens, Hedwig Van Bakel, Harry Knoors, Constance Vissers, and Evelien Dirks. 2022. Joint attention in the context of hearing loss: A meta-analysis and narrative synthesis. *Journal of deaf studies and deaf education* 27, 1 (2022), 1–15.

[76] Sooyeon Lee, Madison Reddie, and John M Carroll. 2021. Designing for Independence for people with visual impairments. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–19.

[77] Amy M Lieberman, Marla Hatrak, and Rachel I Mayberry. 2014. Learning to look for language: Development of joint attention in young deaf children. *Language Learning and Development* 10, 1 (2014), 19–35.

[78] Richard J Light. 1971. Measures of response agreement for qualitative data: some generalizations and alternatives. *Psychological bulletin* 76, 5 (1971), 365.

[79] Berthold Lowenfeld. 1949. A psychological approach to blindness. *Journal of exceptional children* 16, 1 (1949), 1–19.

[80] Margaret Loy, Elise Frank Masur, and Janet Olson. 2018. Developmental changes in infants' and mothers' pathways to achieving joint attention episodes. *Infant Behavior and Development* 50 (2018), 264–273.

[81] Casimir JH Ludwig and David R Evens. 2017. Information foraging for perceptual decisions. *Journal of Experimental Psychology: Human Perception and Performance* 43, 2 (2017), 245.

[82] Lumivero. 2024. NVIVO. https://techcenter.qsrinternational.com/Content/welcome/toc_welcome.htm [Online; accessed 26-August-2024].

[83] Emiliano Macaluso. 2010. Orienting of spatial attention and the interplay between the senses. *Cortex* 46, 3 (2010), 282–297.

[84] Arien Mack. 2003. Inattentional blindness: Looking without seeing. *Current directions in psychological science* 12, 5 (2003), 180–184.

[85] Anna-Karin Magnusson. 2006. Nonverbal conversation-regulating signals of the blind adult. *Communication Studies* 57, 4 (2006), 421–433.

[86] Jessica Markus, Peter Mundy, Michael Morales, Christine EF Delgado, and Marygrace Yale. 2000. Individual differences in infant skills as predictors of child-caregiver joint attention and language. *Social Development* 9, 3 (2000), 302–315.

[87] Chiara Martolini, Anna V Cuppone, Giulia Cappagli, Sara Finocchietti, Antonio Maviglia, and Monica Gori. 2018. ABBI-K: a novel tool for evaluating spatial and motor abilities in visually impaired children. In *2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. IEEE, 1–6.

[88] David Matsumoto and Bob Willingham. 2009. Spontaneous facial expressions of emotion of congenitally and noncongenitally blind individuals. *Journal of personality and social psychology* 96, 1 (2009), 1.

[89] Julie F McClelland, Una O'Connor, Christine Shannon, Kathryn J Saunders, and Julie-Anne Little. 2023. Exploring the role and experience of classroom assistants supporting pupils with visual impairment. *International Journal of Inclusive Education* (2023), 1–15.

[90] Michele C McDonnall, Anne Steverson, Rachael Sessler Trinkowsky, and Katerina Sergi. 2023. Assistive technology use in the workplace by people with blindness and low vision: Perceived skill level, satisfaction, and challenges.

Assistive Technology (2023), 1–8.

[91] Sarah C Mead Jasperse, Michelle P Kelly, Shannon N Ward, Jonathan K Fernand, P Raymond Joslyn, and Wilhelmina van Dijk. 2023. Consent and assent practices in behavior analytic research. *Behavior Analysis in Practice* (2023), 1–16.

[92] P.B.L. Meijer. 1992. An experimental system for auditory image representations. *IEEE Transactions on Biomedical Engineering* 39, 2 (1992), 112–121. doi:10.1109/10.121642

[93] Peter B.L. Meijer. 2024. *"The vOICe - New Frontiers in Artificial Vision"*. Retrieved 7 December 2024 from https://www.seeingwithsound.com/

[94] James N Meindl and Helen I Cannella-Malone. 2011. Initiating and responding to joint attention bids in children with autism: A review of the literature. *Research in developmental disabilities* 32, 5 (2011), 1441–1454.

[95] Oussama Metatla, Sandra Bardot, Clare Cullen, Marcos Serrano, and Christophe Jouffrais. 2020. Robots for Inclusive Play: Co-designing an Educational Game With Visually Impaired and sighted Children. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3313831.3376270

[96] Oussama Metatla, Nick Bryan-Kinns, and Tony Stockman. 2018. "I Hear You": Understanding Awareness Information Exchange in an Audio-only Workspace. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3173574.3174120

[97] Oussama Metatla and Clare Cullen. 2018. "Bursting the Assistance Bubble": Designing Inclusive Technology with Children with Mixed Visual Abilities. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3173574.3173920

[98] Oussama Metatla, Alison Oldfield, Taimur Ahmed, Antonis Vafeas, and Sunny Miglani. 2019. Voice User Interfaces in Schools: Co-designing for Inclusion with Visually-Impaired and Sighted Pupils. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–15. doi:10.1145/3290605.3300608

[99] Jennie Middleton and Hari Byles. 2019. Interdependent temporalities and the everyday mobilities of visually impaired young people. *Geoforum* 102 (2019), 76–85.

[100] Cecily Morrison, Edward Cutrell, Martin Grayson, Elisabeth RB Becker, Vasiliki Kladouchou, Linda Pring, Katherine Jones, Rita Faia Marques, Camilla Longden, and Abigail Sellen. 2021. Enabling meaningful use of AI-infused educational technologies for children with blindness: Learnings from the development and piloting of the PeopleLens curriculum. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–13.

[101] Cecily Morrison, Edward Cutrell, Martin Grayson, Anja Thieme, Alex Taylor, Geert Roumen, Camilla Longden, Sebastian Tschiatschek, Rita Faia Marques, and Abigail Sellen. 2021. Social Sensemaking with AI: Designing an Open-ended AI Experience with a Blind Child. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>Yokohama</city>, <country>Japan</country>, </conf-loc>) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 396, 14 pages. doi:10.1145/3411764.3445290

[102] Peter Mundy. 2018. A review of joint attention and social-cognitive brain systems in typical development and autism spectrum disorder. *European Journal of Neuroscience* 47, 6 (2018), 497–514.

[103] Peter Mundy and Antoinette Gomes. 1998. Individual differences in joint attention skill development in the second year. *Infant behavior and development* 21, 3 (1998), 469–482.

[104] Peter Mundy and William Jarrold. 2010. Infant joint attention, neural networks and social cognition. *Neural Networks* 23, 8-9 (2010), 985–997.

[105] Peter Mundy and Lisa Newell. 2007. Attention, joint attention, and social cognition. *Current directions in psychological science* 16, 5 (2007), 269–274.

[106] Peter Mundy and Jennifer Willoughby. 2014. Nonverbal communication, joint attention, and early socioemotional development. In *Emotional development in atypical children*. Psychology Press, 65–87.

[107] Arslan Munir, Alexander Aved, and Erik Blasch. 2022. Situational awareness: techniques, challenges, and prospects. *AI* 3, 1 (2022), 55–77.

[108] Lauren Murray, Philip Hands, Ross Goucher, and Juan Ye. 2016. Capturing social cues with imaging glasses. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*. 968–972.

[109] Tabah Aris Nurjaman. 2018. Exploring the interdependence between visually impaired and sighted people in the early phase of friendship. *Indonesian Journal of Disability Studies* 5, 1 (2018), 115–126.

[110] Aude Oliva. 2005. Gist of the scene. In *Neurobiology of attention*. Elsevier, 251–256.

[111] Aude Oliva and Antonio Torralba. 2006. Building the gist of a scene: The role of global image features in recognition. *Progress in brain research* 155 (2006), 23–36.

[112] Cliodhna O'Connor and Helene Joffe. 2020. Intercoder reliability in qualitative research: debates and practical guidelines. *International journal of qualitative methods* 19 (2020), 1609406919899220.

[113] Sethuraman Panchanathan, Shayok Chakraborty, and Troy McDaniel. 2016. Social interaction assistant: a person-centered approach to enrich social interactions for individuals with visual impairments. *IEEE Journal of Selected Topics in Signal Processing* 10, 5 (2016), 942–951.

[114] Michael A Peters, E Jayne White, Tina Besley, Kirsten Locke, Bridgette Redder, Rene Novak, Andrew Gibbons, John O'Neill, Marek Tesar, and Sean Sturm. 2021. Video ethics in educational research involving children: Literature review and critical discussion. *The Methodology and Philosophy of Collective Writing* (2021), 292–313.

[115] Madeline Phillips and Michael J Proulx. 2018. Social interaction without vision: an assessment of assistive technology for the visually impaired. *Technology & Innovation* 20, 1-2 (2018), 85–93.

[116] Caroline Pigeon, Tong Li, Fabien Moreau, Gilbert Pradel, and Claude Marin-Lamellet. 2019. Cognitive load of walking in people who are blind: Subjective and objective measures for assessment. *Gait & posture* 67 (2019), 43–49.

[117] Gunilla M Preisler. 1991. Early patterns of interaction between blind infants and their sighted mothers. *Child: care, health and development* 17, 2 (1991), 65–90.

[118] Gunilla M Preisler. 1993. A descriptive study of blind children in nurseries with sighted children. *Child: care, health and development* 19, 5 (1993), 295–315.

[119] Shi Qiu, Siti Aisyah Anas, Hirotaka Osawa, Matthias Rauterberg, and Jun Hu. 2016. E-Gaze Glasses: Simulating Natural Gazes for Blind People. In *Proceedings of the TEI '16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction* (Eindhoven, Netherlands) *(TEI '16)*. Association for Computing Machinery, New York, NY, USA, 563–569. doi:10.1145/2839462.2856518

[120] Shi Qiu, Jun Hu, Ting Han, Hirotaka Osawa, and Matthias Rauterberg. 2020. Social glasses: simulating interactive gaze for visually impaired people in face-to-face communication. *International Journal of Human–Computer Interaction* 36, 9 (2020), 839–855.

[121] Shi Qiu, Matthias Rauterberg, and Jun Hu. 2016. Tactile band: Accessing gaze signals from the sighted in face-to-face communication. In *Proceedings of the TEI'16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction*. 556–562.

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

[123] Joao Roe. 2008. Social inclusion: meeting the socio-emotional needs of children with vision needs. *British Journal of Visual Impairment* 26, 2 (2008), 147–158.

[124] Janelle C Rosip and Judith A Hall. 2004. Knowledge of nonverbal cues, gender, and nonverbal decoding accuracy. *Journal of Nonverbal Behavior* 28 (2004), 267–286.

[125] Royal National Institute of Blind People. 2021. *Key information and statistics on sight loss in the UK*. Retrieved April 2023 from https://www.rnib.org.uk/professionals/health-social-care-education-professionals/knowledge-and-research-hub/key-information-and-statistics-on-sight-loss-in-the-uk

[126] Gennaro Ruggiero, Francesco Ruotolo, and Tina Iachini. 2018. Congenital blindness limits allocentric to egocentric switching ability. *Experimental Brain Research* 236 (2018), 813–820.

[127] Carolina Sacramento and Simone Bacellar Leal Ferreira. 2022. Accessibility on social media: exploring congenital blind people's interaction with visual content. In *Proceedings of the 21st Brazilian Symposium on Human Factors in Computing Systems*. 1–7.

[128] Nuzhah Gooda Sahib, Tony Stockman, Anastasios Tombros, and Oussama Metatla. 2013. Participatory design with blind users: a scenario-based approach. In *Human-Computer Interaction–INTERACT 2013: 14th IFIP TC 13 International Conference, Cape Town, South Africa, September 2-6, 2013, Proceedings, Part I 14*. Springer, 685–701.

[129] Michael Scaife and Jerome S Bruner. 1975. The capacity for joint visual attention in the infant. *Nature* 253, 5489 (1975), 265–266.

[130] Catarina Schmidt. 2019. Classroom observations through video recordings - Methodological approaches and ethical considerations. In *Classroom research : Methodology, categories and coding*. Vol. s. 16-20. Linnaeus University Press, Växjö, 16–21. http://lnu.diva-portal.org/smash/get/diva2:1361576/FULLTEXT01#page=17

[131] Axel Seemann. 2024. Joint attention as the base of common knowledge and collective intentionality. *Topoi* (2024), 1–12.

[132] Ashley Shew. 2023. *Against technoableism: rethinking who needs improvement*. WW Norton & Company.

[133] Minsun Shin. 2012. The role of joint attention in social communication and play among infants. *Journal of Early Childhood Research* 10, 3 (2012), 309–317.

[134] Daniel J Simons and Christopher F Chabris. 1999. Gorillas in our midst: Sustained inattentional blindness for dynamic events. *perception* 28, 9 (1999), 1059–1074.

[135] Kiley Sobel, Maitraye Das, Sara Behbakht, and Julie A Kientz. 2024. Incloodle-Classroom: Technology for Inclusive Joint Media Engagement in a Neurodiverse Kindergarten Classroom. *ACM Transactions on Computer-Human Interaction* 31,

3 (2024), 1–45.
[136] Abigale Stangl, Nitin Verma, Kenneth R Fleischmann, Meredith Ringel Morris, and Danna Gurari. 2021. Going beyond one-size-fits-all image descriptions to satisfy the information wants of people who are blind or have low vision. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–15.
[137] Melisa Stevanovic and Anssi Peräkylä. 2015. Experience sharing, emotional reciprocity, and turn-taking. *Frontiers in psychology* 6 (2015), 450.
[138] Tricia Striano, Vincent M Reid, and Stefanie Hoehl. 2006. Neural mechanisms of joint attention in infancy. *European Journal of Neuroscience* 23, 10 (2006), 2819–2823.
[139] K. Subapriya. 2009. The Importance of Non-Verbal Cues. *ICFAI Journal of Soft Skills* 3, 2 (Jun. 2009), 37 – 42. https://www.iupindia.in/609/IJSS_Non-Verbal%20Cues_37.html
[140] Valerie Tadić, Linda Pring, and Naomi Dale. 2009. Attentional processes in young children with congenital visual impairment. *British Journal of Developmental Psychology* 27, 2 (2009), 311–330.
[141] Helen Tager-Flusberg. 2014. *Constraints on language acquisition: Studies of atypical children*. Psychology Press.
[142] Lucinda J Teoh, Ameenat Lola Solebo, Jugnoo S Rahi, Joe Abbott, Wajda Abdullah, Gill Adams, Louise Allen, Christopher Anderson, Karen Ansell, Samira Anwar, et al. 2021. Visual impairment, severe visual impairment, and blindness in children in Britain (BCVIS2): a national observational study. *The Lancet Child & Adolescent Health* 5, 3 (2021), 190–200.
[143] The International Agency for the Prevention of Blindness. 2024. *Child eye health*. IAPB. Retrieved 3 September 2024 from https://www.iapb.org/learn/vision-atlas/magnitude-and-projections/child-eye-health/
[144] Catherine Thinus-Blanc and Florence Gaunet. 1997. Representation of space in blind persons: vision as a spatial sense? *Psychological bulletin* 121, 1 (1997), 20.
[145] Michael Tomasello. 1995. Joint attention: Its origins and role in development. *(No Title)* (1995), 103.
[146] Michael Tomasello. 1999. *The cultural origins of human cognition*. Harvard university press.
[147] Michael Tomasello and Malinda Carpenter. 2007. Shared intentionality. *Developmental science* 10, 1 (2007), 121–125.
[148] Michael Tomasello and Michael Jeffrey Farrar. 1986. Joint Attention and Early Language. *Child Development* 57, 6 (1986), 1454–1463. doi:10.2307/1130423
[149] UNGeneva 2024. *Disability Inclusive Language Guidelines*. UNGeneva. Retrieved 12 February 2024 from https://www.ungeneva.org/en/about/accessibility/disability-inclusive-language
[150] Lotte van der Jagt. 2023. *Social Navigation: Assisting people that are blind or have low vision in finding and navigating towards familiar individuals in close proximity*. Master's thesis. Delft University of Technology. https://resolver.tudelft.nl/uuid:6bf0e43a-093c-4628-8b8e-f0065eb26ea9
[151] Maarten Van Mechelen, Bieke Zaman, Ann Laenen, and Vero Vanden Abeele. 2015. Challenging group dynamics in participatory design with children: Lessons from social interdependence theory. In *Proceedings of the 14th international conference on interaction design and children*. 219–228.
[152] Amy Vaughan Van Hecke, Peter C Mundy, C Françoise Acra, Jessica J Block, Christine EF Delgado, Meaghan V Parlade, Jessica A Meyer, A Rebecca Neal, and Yuly B Pomares. 2007. Infant joint attention, temperament, and social competence in preschool children. *Child development* 78, 1 (2007), 53–69.
[153] Suzanne H Verver, Mathijs PJ Vervloed, and Bert Steenbergen. 2020. Facilitating play and social interaction between children with visual impairments and sighted peers by means of augmented toys. *Journal of Developmental and Physical Disabilities* 32, 1 (2020), 93–111.
[154] Beatrice Vincenzi, Alex S Taylor, and Simone Stumpf. 2021. Interdependence in action: people with visual impairments and their guides co-constituting common spaces. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–33.
[155] Ethan Waisberg, Joshua Ong, Mouayad Masalkhi, Nasif Zaman, Prithul Sarker, Andrew G Lee, and Alireza Tavakkoli. 2024. Meta smart glasses—large language models and the future for assistive glasses for individuals with vision impairments. *Eye* 38, 6 (2024), 1036–1038.
[156] LeAnn M Weih, Jennifer B Hassell, and Jill Keeffe. 2002. Assessment of the impact of vision impairment. *Investigative ophthalmology & visual science* 43, 4 (2002), 927–935.
[157] J Dennis White and Donal E Carlston. 1983. Consequences of schemata for attention, impressions, and recall in complex social interactions. *Journal of Personality and Social Psychology* 45, 3 (1983), 538.
[158] Michele A Williams, Amy Hurst, and Shaun K Kane. 2013. " Pray before you step out" describing personal and situational blind navigation behaviors. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–8.
[159] Bo Xiao, Viktor Rozgic, Athanasios Katsamanis, Brian R Baucom, Panayiotis G Georgiou, and Shrikanth S Narayanan. 2011. Acoustic and Visual Cues of Turn-Taking Dynamics in Dyadic Interactions.. In *INTERSPEECH*. 2441–2444.
[160] Chi Xu. 2021. Nreal: Ready-to-wear mixed reality glasses. In *Spie avr21 Industry Talks II*, Vol. 11764. SPIE, 1176409.
[161] Reda Yaagoubi and Geoffrey Edwards. 2008. Cognitive design in action: developing assistive technology for situational awareness for persons who are blind. *Disability and Rehabilitation: Assistive Technology* 3, 5 (2008), 241–252.
[162] Rayoung Yang, Sangmi Park, Sonali R Mishra, Zhenan Hong, Clint Newsom, Hyeon Joo, Erik Hofer, and Mark W Newman. 2011. Supporting spatial awareness and independent wayfinding for pedestrians with visual impairments. In *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility*. 27–34.
[163] Anon Ymous, Katta Spiel, Os Keyes, Rua M Williams, Judith Good, Eva Hornecker, and Cynthia L Bennett. 2020. " I am just terrified of my future"—Epistemic Violence in Disability Related Technology Research. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–16.
[164] Aziz N Zeidieh. 2024. " Seven Stitches Later": A Technologically Interdependent Travel Experience From The Perspective Of A Visually Impaired Individual. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–6.