Mentalizing and Perspective Taking in Autistic Adults – Probing Speech Intonation, Eye Gaze and Free Indirect Discourse

Inaugural Dissertation

zur

Erlangung des Doktorgrades
philosophiae doctor (PhD) in Health Sciences
der Medizinischen Fakultät
der Universität zu Köln

vorgelegt von
Juliane T. Zimmermann
aus Köln

Köln, 2025

Betreuer*in:	Prof. Dr. Dr. Kai Vogeley
	Prof. Dr. Martine Grice
Gutachter*in:	Prof. Dr. Elke Kalbe
	Prof. Dr. Peter Weiss-Blankenhorn
Datum der Mündlichen Prüfung:	07.05.2025

Index

List of Abbreviations 4
List of Studies 5
1 Introduction 6
1.1 Autism and Language 6
1.2 Perspective Taking 7
1.2.1 Conceptual Perspective Taking in Autism 8
1.2.1.1 ToM in Adolescents and Adults with Autism 9
1.2.2 Perspective Taking and Language in Autism 10
1.2.3 Perspective Taking in Written Text in Free Indirect Discourse 11
1.2.3.1 The Role of Prominence in Free Indirect Discourse Anchoring 12
1.2.4 Speech Perception (Prosody) 13
1.2.4.1 The Role of Prominence in Prosody 13
1.2.4.2 Prosody and Mentalizing 14
1.2.4.3 The Influence of Pitch on Observers' Attentional Focus and Memory 15
1.2.5 Gaze Perception 15
1.2.5.1 Gaze and Mentalizing 16
1.2.5.2 The Influence of Gaze on Observers' Attentional Focus and Memory 16
1.2.6 Multimodal Perception 17
1.3 Prominence and Salience 18
1.4 Aims of the Current Thesis 18
2 Study 1a 20
3 Study 1b 26
3.1 Background and Aim of the Study 27
3.2 Material and Methods 27
3.2.1 Participants 28
3.2.2 Eye-Tracking 28
3.2.3 Analysis 28
3.3 Results 30
3.3.1 Rating Behavior 30
3.3.1.1 Individuality 31
3.3.1.2 Exploratory Analysis 33
3.3.2 Gaze Fixation Durations 33
3.3.3 Effect of Memory on Object Recognition 35
3.4 Discussion 35
3.4.1 Rating Behavior 35
3.4.1.1 Individuality 36
3.4.1.2 Exploratory Analysis 38
3.4.2 Fixation Durations 39

3.4.3 Effect of Memory on Object Recognition 40
3.5 Further Limitations 42
3.6 Conclusion 42
4 Study 1c 43
5 Study 2 61
6 General Discussion 75
6.1 Inferring Mental States from Nonverbal Information 75
6.1.1 Summary 75
6.1.2 A Mentalizing Task amongst Others 76
6.1.3 Relative Impact of Gaze Duration and Intonation 77
6.1.3.1 Relative Impact of Gaze Duration and Intonation in Autism 77
6.1.3.2 Relative Impact of Gaze Duration and Intonation in Non-autistic People 78
6.1.4 Audiovisual Interaction 79
6.1.5 Individual Behavior 81
6.1.6 Gaze Fixation Durations 83
6.1.7 Attention (Re-)Orienting 85
6.1.7.1 Participants' Gaze Behavior 85
6.1.7.2 Effect of Memory on Object Recognition 86
6.2 Perspective Taking in Written Text 87
6.2.1 Summary 87
6.2.2 Free Indirect Discourse Perception 87
6.2.3 Open Research Questions 88
6.2.3.1 What if Free Indirect Discourse Involves Perspective Taking? 88
6.2.3.2 What if Free Indirect Discourse does not Involve Perspective Taking? 92
6.2.3.3 Beyond Naturalness Ratings: Measures and Material 93
6.2.4 Ageing in Non-Autistic Readers: Decline of Rating Differences 95
6.3 Clinical Implications 96
6.4 Conclusion and Outlook 96
7 Summary 98
8 Zusammenfassung 100
9 Appendix 103
9.1 Supplementary Material for Studies 1b and 1c 104
9.1.1 Perceptual Pretest of Auditory Stimuli 104
9.1.2 List of Object Names and their English Translations 104
9.2 Supplementary Material for Study 2 107
10 References 120
11 Acknowledgments 146
12 Eidesstattliche Erklärung 147
12.1 Veröffentlichte Publikationen und Eigenanteil 147

List of Abbreviations

AQ	Autism spectrum quotient
ASD	Autism spectrum disorder
BDI-II	Beck depression inventory-II
BF	Bayes factor
CI	Confidence interval
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, 5th edition
EQ	Empathy quotient
F_0	Fundamental frequency
FID	Free indirect discourse
H*	High tone pitch accent
ICD-10 / ICD-11	International Statistical Classification of Diseases and Related Health Problems, 10th revision / 11th revision
L + H*	High tone pitch accent, preceded by a low tone
LKJ	Lewandowski-Kurowicka-Joe (distribution)
SPQ	Sensory perception quotient
SQ	Systemizing quotient
ToM	Theory of Mind
WIE-III	Hamburg-Wechsler-Intelligenz-Test für Erwachsene III
WST	WST-Wortschatztest

List of Studies

This thesis includes the following studies:

- **Study 1a**

  Zimmermann, J. T., Wehrle, S., Cangemi, F., Grice, M., & Vogeley, K. (2020). Listeners and Lookers: Using pitch height and gaze duration for inferring mental states. In *Proceedings of the 10th International Conference on Speech Prosody* (pp. 290–294). https://doi.org/10.21437/SpeechProsody.2020-59

- **Study 1b**

  Laboratory study in non-autistic participants: *Listeners and Lookers investigated further – Their use of pitch height and gaze duration for inferring mental states is mirrored by gaze behavior.* [unpublished data]

- **Study 1c**

  Zimmermann, J. T., Ellison, T. M., Cangemi, F., Wehrle, S., Vogeley, K., & Grice, M. (2024). Lookers and Listeners on the autism spectrum: The roles of gaze duration and pitch height in inferring mental states. *Frontiers in Communication, 9,* Article 1483135. https://doi.org/10.3389/fcomm.2024.1483135

- **Study 2**

  Zimmermann, J. T., Meuser, S., Hinterwimmer, S., & Vogeley, K. (2021). Preserved perspective taking in free indirect discourse in autism spectrum disorder. *Frontiers in Psychology*, *12*, Article 675633. https://doi.org/10.3389/fpsyg.2021.675633

1 Introduction

Mentalizing and perspective taking have both been discussed as key difficulties in autism, contributing to problems in social interaction and social communication. Perception of language, amongst other faculties, requires an understanding of what is specifically highlighted by a speaker or writer to indicate what is in the current focus of attention. In other words, listeners or readers need to infer from speech or text which linguistic element (e.g. a particular word) the speaker or writer is marking out as important. It has not been understood comprehensively which aspects of mentalizing and perspective taking are challenging for autistic people when encoding or decoding the importance – also referred to as prominence – of entities in language. The aim of this thesis is to investigate mentalizing and perspective taking in the perception of language in autism – in particular in the perception of prominence – by (i) establishing a paradigm to investigate mentalizing drawing on perception of nonverbal information, namely gaze and intonation (study 1a and study 1b), (ii) investigating mentalizing abilities in autistic adults based on nonverbal information using the established paradigm (study 1c) and (iii) investigating perspective taking in autistic adults in a written verbal task (study 2). Study 1a, 1b and 1c investigate the perception of audio-visual stimuli in conversation-like videos. Study 2 makes first steps into examining free indirect discourse perception in autistic people while reading short stories.

The introduction will focus on mentalizing and perspective taking in autism. In this context, perception of language in autism will be introduced, especially the perception of speech intonation and deictic eye gaze. The main part of this thesis comprising studies 1a–c and study 2 is followed by a general discussion.

1.1 Autism and Language

Autism spectrum disorder[1] is defined as a neurodevelopmental disorder characterized by difficulties in social communication and interaction (American Psychiatric Association, 2013). It affects approximately 1 % of the population (Zeidan et al., 2022). The latest clinical classification subsumes autistic conditions of different degrees of severity under the diagnosis "autism spectrum disorder" (American Psychiatric Association, 2013; World Health Organization, 2019). At the time the studies comprising the current thesis were carried out, this diagnosis was not yet in practice. Participants with autism spectrum conditions taking part in

[1] The terms "autism spectrum disorder", "autism spectrum condition", "autism spectrum" and "autism" are used interchangeably in this thesis.

study 1c and study 2 were mostly diagnosed with Asperger's (ICD-10 identifier: F.84.5), few with childhood autism (ICD-10 identifier: F.84.0) and all had normal intelligence.

Difficulties in social communication and interaction in autism can manifest in various ways (American Psychiatric Association, 2013): Language deficits or abnormalities across the autism spectrum range from a complete lack of speech in more severe autism to language delays, poor comprehension of speech and stilted or overly literal language in less severe autism. In the absence of intellectual impairment and language delays, autistic adults may nevertheless show difficulties processing and responding to complex social cues such as when and how to join a conversation or what not to say in a certain situation. Language characteristics in autism further include impairments in normal back-and-forth conversation, poorly integrated verbal and nonverbal communication, problems understanding the different ways that language may be used to communicate (e.g. irony; sarcasm; metaphors; white lies) as well as stereotyped or repetitive use of speech (e.g. echolalia, i.e. repeating speech; repetitive questioning; idiosyncratic phrases; stereotyped use of words, phrases, or prosodic patterns). Additionally, communicative body language such as eye contact, gestures or speech intonation can be absent, reduced or atypical: For example, early signs of autism include an impaired initiation of joint attention and impaired following of another person's deictic gestures such as via eye gaze or hand gestures.

1.2 Perspective Taking

With researchers trying to pin down an underlying core impairment in autism that can comprehensively explain autistic characteristics, cognitive theories of autism have emerged in the 80's. One of these is the Theory of Mind (ToM) theory (Baron-Cohen et al., 1985), which has continued to influence studies on autism throughout the past decades. According to the ToM theory, perspective taking in autism is impaired or hampered – in particular the act of reading other people's thoughts and emotions. Although the ToM theory cannot account for the entirety of autistic characteristics, it continues to inspire research in autism, e.g. the current thesis.

Perspective taking has been described with different terminology – depending on the context and the investigated aspects of the process. It may be roughly divided into two forms of perspective taking: Perceptual and conceptual perspective taking (Marvin et al., 1976). Perceptional perspective taking such as visual perspective taking (Flavell, 1977) is required when trying to comprehend what another person is physically perceiving, e.g. when we try to

figure out what another person can see from their view point or what a certain object looks like from their perspective. On the other hand, conceptual perspective taking refers to the act of inferring inner mental states of others (or sometimes of oneself) such as thoughts, desires, attitudes and plans (Marvin et al., 1976). This ability is also referred to as having a "Theory of Mind" (ToM; Premack and Woodruff, 1978) or as "mentalizing" (Fonagy et al., 2004; Frith & Frith, 2006).

On another dimension, perspective taking can further be distinguished into implicit – or automatic – and explicit – or willful – perspective taking. This differentiation is based on the idea of two perspective taking processes: an implicit one allowing for fast automatic inferences and a slower, cognitively more effortful, explicit one (Apperly & Butterfill, 2009). For example, while a person can explicitly be instructed to take the perspective of another, implicit perspective taking happens spontaneously without an explicit instruction and sometimes without the requirement to take the other's perspective (e.g. Samson et al., 2010).

1.2.1 Conceptual Perspective Taking in Autism

Social difficulties of autistic people may in part be explained by impaired mentalizing abilities (Baron-Cohen et al., 1985; Frith et al., 1991). Such an impairment has often been demonstrated in language-based tasks with autistic children (Baron-Cohen, 1989; Baron-Cohen et al., 1985; Begeer et al., 2014; Hutchins et al., 2012; Leslie & Thaiss, 1992; Swettenham, 1996).

For investigating ToM abilities, explicit cognitive-linguistic tasks such as false belief tasks are often used. In these tasks, participants are required to answer questions referring to one or more protagonists' mental states. Depending on the level of ToM abilities, the observer is required to infer a character's believes (first-order ToM), or a character's believes about the believes of another character (second-order ToM). For example, a scenario in a false-belief task (Perner & Wimmer, 1985) probing first- and second-order ToM abilities may look as follows: John and Mary are in a park where an ice-cream vendor is selling ice-cream from his ice-cream van. When Mary goes home to fetch money to buy ice-cream, the ice-cream vendor tells John, who is staying in the park, that he will drive to the church as there are no customers left in the park. At this point in the story, a question probing first-order ToM may look as follows: Where is Mary going to go to buy ice-cream after she has fetched the money? The observer would be expected to answer "To the park" as Mary has not witnessed the ice-cream vendor change locations. However, the story continues, as we are also interested in probing the observer's second-order ToM. As the ice-cream vendor coincidentally passes Mary on his way to the church, he tells her that he will sell ice-cream at the church next. Mary therefore makes her way

to the church – not the park – to buy ice-cream. In the meantime, John has gone home to his house to have lunch and afterwards makes his way to Mary's house. Mary's mother tells him, Mary has gone out to buy ice-cream. The question probing second-order ToM would be the following: Where does John think Mary has gone? The observer is expected to answer "To the park", because John has not witnessed the ice-cream vendor tell Mary that he is driving to the church.

Tasks like these are based on verbal information (Gernsbacher & Yergeau, 2019) and participants need to make explicit use of cognitive reasoning to make sense of the given situation (Chung et al., 2014). In some of these tasks the material itself is presented verbally – such as in the Strange Stories task (e.g. Happé, 1994) and Faux Pas task (e.g. Baron-Cohen et al., 1999). In others, the material itself is visual, but explanations, instructions and the experimenter's questions are communicated via spoken language and require a verbal response – such as in the false belief task (e.g. Perner and Wimmer, 1985). In various types of tasks probing ToM abilities, participants' language abilities have a great impact on task performance in autistic people (Bennett et al., 2013; Loukusa et al., 2014; Norbury, 2005; Velloso et al., 2013) as well as in non-autistic people (Capage & Watson, 2001; Lawrence et al., 2004; Peterson et al., 2012; Shaked et al., 2006; Steele et al., 2003).

1.2.1.1 ToM in Adolescents and Adults with Autism

Adolescents or adults on the autism spectrum with normal intelligence – unlike children – usually pass these explicit cognitive-linguistic tasks successfully (Bowler, 1992; Gernsbacher & Yergeau, 2019; Happé, 1994; Ponnet et al., 2004; Scheeren et al., 2013), or with minor difficulties (Lever & Geurts, 2016).

Although autistic adults perform similarly to non-autistic participants when inferring a protagonist's mental state in stories presented verbally or in written form, they perform worse than non-autistic participants when providing reasons for their attributions (Bowler, 1992; Callenmark et al., 2014; Dziobek et al., 2006; Happé, 1994; Ozonoff et al., 1991). Despite explicit instructions, accessing the required information may exceed the capacity of a cognitive-linguistic approach in these types of tasks.

Likewise, tasks based on material that is even less accessible cognitive-linguistically reveal difficulties, e.g. when autistic participants are explicitly asked to infer a protagonist's mental state based not on primarily verbal information but on photo, image or video material depicting people (Baron-Cohen et al., 2001; Brewer et al., 2017; David et al., 2010; Dziobek et al., 2006;

Heavey et al., 2000; Murray et al., 2017; Ponnet et al., 2004; Wilson et al., 2014) or animated triangles (White et al., 2011; Wilson et al., 2014; Zwickel et al., 2011).

Similarly, tasks that require an implicit type of perspective taking, i.e. tasks in which participants are not explicitly asked to take the perspective of another, appear to be difficult or be processed differently by autistic adolescents or adults. This has for example been observed when eye movement was measured to assess overt attention in false belief tasks (Schneider et al., 2013; Schuwerk et al., 2015; Senju et al., 2009). Autistic observers in these studies did not exhibit the same anticipatory gaze towards the target location as non-autistic observers who tended to direct their gaze towards the location indicated by a character's false belief.

1.2.2 Perspective Taking and Language in Autism

In conversations, but really in every type of verbal or nonverbal language production and perception of language, perspective taking is an important skill: On the one side, speakers tend to adjust their speech to the listener, and stories are usually written with a reader in mind. On the other side, listeners and readers can better follow spoken speech or written stories if they are able to take into account the perspective of the speaker or of a protagonist in the story.

Effective communication is usually facilitated if communicators establish what is called "common ground" together – a pool of knowledge shared by all communicators involved (Stalnaker, 2002). The use of a pronoun, for instance, requires knowledge about the so-called referent, i.e. the entity the pronoun refers to: If person A told person B "I will paint him", Person B can know who will be painted only if they know who "him" refers to. If "he" is not part of the "common ground", person B will likely be confused.

The general population tends to adjust their choice of referential expressions to the listener or reader (i.e., depending on the context, they choose to substitute names with pronouns; Achim et al., 2017). While a study including autistic children and adolescents (Arnold et al., 2009) reported that only some of the younger autistic participants (but not the older ones) chose over-specific references more often than needed, a study with autistic adults (Colle et al., 2008) showed more pronounced group differences: When asked to tell a story based on an illustrated book, autistic participants used more full noun phrases when they could instead have used pronouns, which resulted in a pedantic narrating style. On the other side, they used more pronouns when full noun phrases would have been less ambiguous and would therefore have made it easier to understand the narration. The authors interpret this as a lack of ToM and thus the inability to account for common ground information (Colle et al., 2008).

1.2.3 Perspective Taking in Written Text in Free Indirect Discourse

Study 2 investigates the perception of written text. In written text, perspective taking is not only relevant for understanding what mental or emotional state a protagonist is in or why they behave a certain way. Perspective taking may further help processing so-called free indirect discourse (FID; Banfield, 1982), i.e. utterances or thoughts that are not directly anchored to one specific protagonist. Processing of FID has been proposed (Zeman, 2017) to share an important aspect with perspective taking involved in ToM as operationalized in many false belief tasks: the ability to identify and differentiate between different viewpoints at the same time.

Unlike direct speech, which is marked by apostrophes (e.g.: "'*I am a dancer*', Sally said."), or indirect speech, which is likewise unambiguously linked to a protagonist (e.g.: "Sally said *she needed money.*"), FID (also: free indirect speech) is not always linked to a protagonist unambiguously, i.e. the protagonist is not explicitly identifiable as the source of the utterance or thought. Consider the last sentence in FID in the following story:

> "Sally was a dancer. One day, she was booked to dance in a well-funded musical and was thrilled as she desperately needed the money. However, when she heard about the inadequate payment for the role, she got into a fight with her posh producer. *That greedy jerk*."

As the FID in the last sentence is not directly linked to any of the two protagonists in particular, it could be understood either as Sally's thoughts, as the producer's thoughts or as the narrator's thoughts. That the thought is most likely Sally's can only be understood within the context. When processing FID, readers are not explicitly instructed to take the perspective of a protagonist. Instead, the anchor for the utterance or thought expressed in FID needs to implicitly be identified.

In autism, FID perception has not been investigated so far. In other implicit perspective taking tasks, namely implicit false-belief tasks, it has been demonstrated that autistic participants do not, by default, exhibit anticipatory looking towards the target location, i.e. towards the location as indicated by a character's false belief (Schneider et al., 2013; Schuwerk et al., 2015; Senju et al., 2009). This has been interpreted as impaired implicit perspective taking in autism. FID perception may draw on implicit perspective taking and, similarly, be challenging or be processed differently in autistic people. Moreover, findings from studies on the production of referential expressions (Colle et al., 2008) and the perception of pronoun-induced perspective shift (Mizuno et al., 2011) suggest that autistic participants may have difficulties when shifting perspectival centers.

1.2.3.1 The Role of Prominence in Free Indirect Discourse Anchoring

Prominence plays an important role in identifying the perspectival center of a story, (Hinterwimmer, 2019). "Prominence" literally describes the property of jutting, standing or projecting out. Accordingly, it has been defined as "the property by which linguistic units are perceived as standing out from their environment" (Terken, 1991, p. 1768; see Himmelmann & Primus, 2015, for a cross-linguistic working definition of prominence).

Prominence is a relevant notion with respect to information structure (Halliday, 1967) – "the division of sentences into focus and background" (Roettger et al., 2019). Concepts linked to information structure – and prominence – are e.g. givenness (e.g. Halliday, 1967) and focus (Rooth, 1992). They connect to prominence by highlighting linguistic entities in their respective dimensions: A linguistic element may be prominent because it is new in a context of a conversation, or because it is already known, i.e. "given". It may also stand out because it is in focus. These attributes are typically manifested through syntactic and semantic properties as well as phonetic features (in speech).

Syntactic and semantic properties play an important role in the perception of prominence, such as subjecthood and animacy. They contribute to so-called discourse prominence, which can be described as a linguistic entity's likelihood of being referred to (Jasinskaja et al., 2015). For example, a person referred to in subject position is usually perceived more prominent than a person referred to in object position (Arnold, 2010). Likewise, an animate entity is usually perceived as more prominent than an inanimate entity (Lockwood & Macaulay, 2012). Concepts similar to discourse prominence are *givenness* (Chafe, 1976), *accessibility* (Ariel, 1990) or *activation* (Gundel et al., 1993). Prominent, given, accessible or activated entities all have in common, that they are the most likely candidates for further (linguistic) operations such as FID anchoring. Taken together, prominence encompasses a range of attentional processes that roughly cover two phenomena, namely (i) the (pre-)activation of an entity (before it is encountered), and (ii) the redirecting of attention towards an entity based on its features.

Prominence in the case of FID anchoring refers to discourse prominence and as such can be understood as the (pre-)activation of an entity. In the FID example above (see section 1.2.3: "*That greedy jerk."*), Sally is the most likely candidate for FID anchoring as she is more prominent than the producer both due to subjecthood (Himmelmann & Primus, 2015) and familiarity (Jasinskaja et al., 2015). Amongst others, grammatical function (e.g. subject or object) and the type of referential expression (e.g. "Sally", "her") (Hinterwimmer & Meuser, 2019) are aspects that contribute to prominence in the context of FID anchoring.

1.2.4 Speech Perception (Prosody)

Studies 1a–c investigate the perception of speech. Beyond the lexical content, meaning can nonverbally be conveyed and extracted in conversation, such as for example via prosody (O'Connor & Arnold, 1973). Opposed to lexicality, prosody does not refer to what is being said, but how it is said. Thus, prosodic measures include fundamental frequency (F_0, perceived as pitch), intensity (perceived as loudness), duration (perceived as length, e.g. of a vowel), and the distribution of audible frequencies (perceived as vowel quality and timbre).

Prosody plays an important role when inferring information on different levels. Grammatical prosody helps a listener understand syntactic aspects e.g. the correct partitioning of a sentence or whether an utterance is a statement or a question: The sentence "You told me to bring coffee, beans and milk?" needs to be realized by the speaker appropriately to avoid confusion, i.e. the intonation at the end of the sentence needs to rise to indicate a question as opposed to a statement, and the word "coffee" needs to be followed by a pause to avoid a mix-up with "coffee beans". In contrast, affective prosody can help a listener infer the feelings and emotional states of the speaker and can be interpreted accordingly. For instance, an utterance produced with joyful surprise is associated with higher mean and maximum pitch than an utterance produced with contempt (Hammerschmidt & Jürgens, 2007). Pragmatic prosody, on the other hand, helps the listener to infer the speaker's thoughts and intentions in the context of an utterance, e.g. which linguistic element they want to highlight to convey its importance: "I told you to bring coffee, beans and OAT milk!" If we as listeners correctly identify that "OAT" is prosodically highlighted in the speaker's utterance, we can infer that it was important to the speaker that the addressee would bring oat milk as opposed to other types of milk.

1.2.4.1 *The Role of Prominence in Prosody*

Resembling discourse prominence, which is relevant for FID anchoring (see above), prosodic prominence refers to the quality of standing out and can thus indicate that something is important, new or in the focus of attention. In German, this can be achieved by means of pitch accent placement and type, cued primarily by F_0, which is perceived as pitch height (Féry & Kügler, 2008; Grice & Baumann, 2007). Speakers usually take into account what the listener already knows and adjust intonation appropriately (Breen et al., 2010).

Correspondingly, raise of pitch is read out by listeners from the general population as prosodic prominence and importance (Arnold et al., 2013; Baumann and Winter, 2018). Amongst all prosodic means to communicate prominence, F_0 is a particularly important one for listeners (Arnold et al., 2013).

1.2.4.2 Prosody and Mentalizing

Inferring the speaker's focus of attention from their use of pitch accents allows the listener to mentalize (Kaland et al., 2014). In the general population, greater sensitivity to pitch accent types has been shown to be linked to better performance on the AQ communication subscale (Bishop, 2016; Bishop et al., 2020; Hurley & Bishop, 2016) which, as suggested by the authors (cf. Bishop et al., 2020, p. 3), may roughly indicate an individual's pragmatic skills, or more precisely in the context of prosody their "sensitivity to the relation between prosody and meaning-in-context".

Findings regarding prosody perception in autism do not paint a clear picture. A narrative review found that intuitive, less rule-based aspects of prosody such as affective and pragmatic prosody have more often been reported as impaired in autism than more formal, less flexible aspects of prosody such as grammatical prosody (Grice et al., 2023).

Regarding more formal aspects of prosody, performance in autistic and non-autistic listeners has been reported to be similar, for example when judging lexical tone placement in Cantonese (Cheng et al., 2017). Likewise, lexical stress perception (e.g. when perceiving the different stress patterns in the noun PREsent and the verb preSENT) has been reported to be similar (Paul et al., 2005), whereas another study reported reduced sensitivity to lexical stress placement in autistic listeners (Kargas et al., 2016). Because in the latter study, stress perception was correlated with speech abnormalities, reduced sensitivity cannot, however, be generalized to the entire autistic group. When resolving syntactic ambiguities with the help of prosodic boundaries (Paul et al., 2005), and when discriminating prosodically indicated questions and statements (Wang et al., 2022) autistic listeners have been reported to perform comparably to non-autistic listeners.

Prosodically expressed basic emotions are recognized well by autistic listeners (Ben-David et al., 2020; O'Connor, 2012; Stewart et al., 2013; Zhang et al., 2022), although performance has been reported to depend on the emotion at hand: While recognizing happiness in speech seems to be particularly difficult for autistic listeners, differences between the autistic and non-autistic group regarding the recognition of other basic emotions were not robustly demonstrated across studies (Zhang et al., 2022). Moreover, autistic listeners show impaired recognition of prosodically expressed emotions that are more complex (Golan et al., 2007; Kleinman et al., 2001; Rosenblau et al., 2017; Rutherford et al., 2002) or produced with low intensity (Globerson et al., 2015). Difficulties regarding the perception and interpretation of vocal pitch modulation during speech may be an underlying cause (Schelinski et al., 2017; Schelinski and von Kriegstein, 2019, see Grice et al., 2023 for a review).

With regard to pragmatic prosody, autistic adults – comparable to non-autistic adults – have been reported to be able to detect contrastive focus (Globerson et al., 2015), whereas another study reported that in most focus-detection or -discrimination tasks included in the study, autistic listeners performed worse than non-autistic listeners (Zaidenberg, 2015). In a different study, pitch accents have been reported to be taken into account to a lesser extent by autistic listeners compared to non-autistic listeners when judging whether a word refers to an object that is already known to the communicating partners or whether it is newly introduced into a conversation (Grice et al., 2016). The results with respect to the perception of prosodically prominent words are thus inconsistent.

1.2.4.3 The Influence of Pitch on Observers' Attentional Focus and Memory

Pitch accents not only guide attention towards linguistic elements (Kristensen et al., 2013), they can also direct our attention towards objects in the environment, as studied in participants' gaze behavior in visual world paradigms in the general population (Dahan et al., 2002; Ito & Speer, 2008; Kurumada et al., 2014; Roettger et al., 2020; Watson et al., 2008; Weber et al., 2006). In the aforementioned studies, audio recordings of another person uttering instructions or descriptions affected participants' proportions of gaze fixations towards a target object.

In a study in children, pitch accents also had an effect on reactions to later joint attention bids both in an autistic and a non-autistic group: they looked at the object longer if the respective utterance had previously received a pitch accent (Ito et al., 2022).

Not only do prominent linguistic elements stand out and affect the observer's gaze behavior, they are also better remembered in the general population: Words produced with greater prosodic prominence are more easily recognized later than words produced with less prosodic prominence (Fraundorf et al., 2010, 2012; Kember et al., 2021; Kushch et al., 2018; Morett & Fraundorf, 2019). This relationship is yet to be investigated in autism.

1.2.5 Gaze Perception

Studies 1a–c investigate the perception of gaze. Like other body movements, gaze behavior corresponds with speech production (Brône et al., 2017; Kendon, 1967; Kendrick et al., 2023; Spaniol et al., 2023). Moreover, eye gaze is closely linked to attention: people from the general population tend to look at objects (Buswell, 1935; DeAngelus & Pelz, 2009; Yarbus, 1967) or locations they pay attention to (Ferreira et al., 2008; Theeuwes et al., 2009). Faces or face-like compositions are particularly attention-grabbing stimuli for newborns (Johnson et al., 1991). They are, however, also of interest to the adult observer as indicated, for

example, by longer fixation durations for faces than for objects (Zhang et al., 2018). Particularly interesting for the current thesis, an object is fixated longer the more relevant it is (Klami, 2010; Klami et al., 2008) and the more it is preferred by the observer (Chuk et al., 2016; Shimojo et al., 2003).

1.2.5.1 Gaze and Mentalizing

Another person's gaze behavior can help us infer their intentions or their attentional state (Baron-Cohen et al., 1995; Einav & Hood, 2006; Freire et al., 2004; Jording, Engemann, et al., 2019; Jording, Hartz, et al., 2019; Lee et al., 1998). Especially in situations, in which other information is ambiguous, observers rely on gaze cues to make sense of a situation (Macdonald & Tatler, 2013). Considering the importance of eye gaze for understanding others' minds, it is not surprising that when a person – regardless whether they are autistic (Auyeung et al., 2015; Dalton et al., 2005; Fedor et al., 2018; Freeth et al., 2010; Hernandez et al., 2009) or not (Fedor et al., 2018; Frischen et al., 2007; Henderson et al., 2005; Itier & Batty, 2009) – observes another person's or a virtual character's face, they predominantly fixate their eye region.

Especially in free-viewing tasks, however, autistic participants exhibit shorter fixation durations for the eye region compared to non-autistic participants (Setien-Ramos et al., 2022). Moreover, autistic participants are less accurate than non-autistic participants in judging gaze directions of others (Forgeot d'Arc et al., 2017; Pantelis & Kennedy, 2017) and perform worse when asked to infer feelings, thoughts and intentions based on eye gaze (Baron-Cohen et al., 1997; Baron-Cohen et al., 2001; Hobson et al., 1988).

1.2.5.2 The Influence of Gaze on Observers' Attentional Focus and Memory

Eye gaze can influence a perceiver's focus of attention, e.g. towards an indicated object: In studies investigating the gaze cueing effect on an implicit level, participants from the general population usually respond faster in reaction to objects that appear in a position previously gazed at by a virtual character (Driver et al., 1999; Friesen & Kingstone, 1998; Mazzarella et al., 2012), suggesting that their attention has shifted towards the indicated objects' position. Additionally, participants are more likely to look at the object that is being looked at by the other person (Ricciardelli et al., 2002). Correspondingly, gaze-cueing can enhance item memory (Dodd et al., 2012; Frischen & Tipper, 2006; Gregory & Jackson, 2017; Gregory & Kessler, 2022). As opposed to other directional cues such as arrows, eye gaze has a special impact on attention allocation which has been demonstrated in gaze-cueing paradigms in the general population (Cañadas & Lupiáñez, 2012; Friesen et al., 2004; Quadflieg et al., 2004).

In adolescents and adults on the autism spectrum, this gaze cueing effect has also been demonstrated, both in settings in which participants know that the eyes carry important information (Ristic et al., 2005) as well as in settings in which attending to the gaze cue can be understood to be more intuitive, i.e. in which participants have no information about gaze cue informativeness (Kuhn et al., 2010) or know that gaze cues are not informative (Vlamings et al., 2005). In the latter study, the authors further reported effects specific to the non-autistic comparison group, most importantly longer visual orienting in response to eye gaze – as opposed to arrow cues – suggesting that deictic eye gaze is processed differently compared to other deictic cues. In the autistic group, arrows and eye gaze produced comparable results suggesting they were both processed as deictic cues of similar quality (Vlamings et al., 2005).

Not only can eye gaze increase the likelihood and speed of orienting towards the indicated direction, but it also influences sustained attention towards that direction: Observers tend to look longer towards an object another person is looking at (Adil et al., 2018; Castelhano et al., 2007; Hutton & Nolte, 2011; Theuring et al., 2007). Moreover, the longer an observer perceives another person looking at an object, the longer they tend to look at the object themselves (Freeth et al., 2010). Correspondingly, observing another person directing their gaze towards an item can increase item memory (Adil et al., 2018; Sajjacholapunt & Ball, 2014).

Compared to non-autistic participants, autistic participants less often look at objects gazed at by another person (Wang et al., 2015) and tend to spend less time fixating those objects (Fletcher-Watson et al., 2009; Freeth et al., 2010).

1.2.6 Multimodal Perception

Verbal and nonverbal language are closely related. Spoken language such as in conversation is not only enhanced by prosody, but usually accompanied by movements of the eyes and by other body movements (see e.g. Brône et al., 2017; Cantalini & Moneglia, 2020; Kendon, 1967, 1972; Wagner et al., 2014). Moreover, nonverbal aspects of language correspond with one another: For example, production of pitch accents has been shown to be linked to gestures of the hands (Krahmer & Swerts, 2007; Rohrer et al., 2023), eyebrows (Ambrazaitis & House, 2017; Krahmer & Swerts, 2007) and the head (Alexanderson et al., 2013; Ambrazaitis & House, 2017; Esteve-Gibert et al., 2017; Krahmer & Swerts, 2007) as well as deictic gestures such as pointing with the index finger (Esteve-Gibert & Prieto, 2013).

In what way and under which circumstances visual and auditory information is integrated during the perception of prosodic prominence has not been elucidated comprehensively. In

some situations, visual and auditory information can add to a common percept: For instance, a speaker's movements in the mouth region (Scarborough et al., 2009) as well as a speaker's eyebrow and head movements can facilitate prominence processing (Biau & Soto-Faraco, 2013; Wang & Chu, 2013) and can add to the listeners' perception of prosodic prominence as conveyed via intonation (Ambrazaitis et al., 2020; House et al., 2001; Krahmer et al., 2002a; Krahmer & Swerts, 2007; Mixdorff et al., 2013; Prieto et al., 2015).

1.3 Prominence and Salience

In this thesis, prominence and salience are considered conceptually equivalent phenomena. Irrespective of terminology, both share similar attentional processes. In psychological terms, prominence can most suitably be compared or equated to the concept of salience, which can – analogous to prominence – roughly be divided into two categories: (i) bottom-up salience, and (ii) top-down salience (Zarcone, van Schijndel, Vogels, & Demberg, 2016). Bottom-up salience is associated with high surprisal and low predictability. Top-down salience is associated with low surprisal and high predictability – and is thus similar to discourse prominence.

It is important to note that the concept of salience is used in linguistic contexts as well, such as in the case of top-down salience contributing to the accessibility of linguistic entities (Ariel, 1990). Moreover, in the scope of linguistic studies, the concept of prominence has been extended to the visual domain (Al Moubayed et al., 2009; Swerts & Krahmer, 2010); Likewise, extra-linguistic factors may feed into an entity's prominence: in a conversation, an object's prominence can be increased, if it is nearby or if it is pointed at (Lewis, 1970). In which cases both concepts refer to the same phenomenon is not always clear as the concepts of prominence and salience have not systematically been compared in-depth.

The comparability of prominence and salience is mentioned here because of the lack of a common terminology across and within research fields and the resulting need to terminologically situate this thesis' studies' objectives. To adhere to disciplinary conventions, the terms "prominence" and "salience" are used in this thesis to refer to linguistic prominence and visual salience, respectively.

1.4 Aims of the Current Thesis

This thesis investigates conceptual perspective taking in the perception of language in autism – in particular regarding the perception of prominence (here: the perception of elements highlighted by speech intonation, by gaze behavior, or by referential expressions in stories).

Against the backdrop of difficulties reported for autistic people in mentalizing tasks involving nonverbal stimuli, the first line of studies (study 1a–c) aims to investigate explicit ToM abilities in autism using audio-visual stimuli in conversation-like videos depicting a virtual character. In study 1a and study 1b, a paradigm to investigate mentalizing drawing on perception of nonverbal information, namely gaze and intonation, was established. These studies were carried out with non-autistic participants. Study 1c aimed to investigate mentalizing abilities in autistic adults based on nonverbal information using the established paradigm. The manipulation in studies 1a–c includes an auditory stimulus, i.e. F_0 variation – to influence what linguistically is most commonly referred to as prosodic prominence perception –, and a visual stimulus, i.e. gaze duration – as a means to influence what psychologically is most commonly referred to as visual salience. Thus, in the context of study 1a–c, cues arguably manipulate bottom-up salience since they are presented without a broader context that would allow for the anticipation of a more prominent auditory or a more salient visual cue.

Study 2 investigates implicit perspective taking, similar to ToM, in autistic adults reading short stories. Implicit perspective taking in these stories is operationalized as FID anchoring. FID perception in autism has not been investigated so far. However, aberrant use of pronouns and difficulties in implicit perspective taking suggest that FID may be processed differently by autistic readers compared to non-autistic readers. Potential targets for FID anchoring vary by prominence status in study 2, more precisely: discourse-prominence – or top-down salience.

Taken together, these studies aim to enhance the understanding of perspective-taking in autistic adults, as well as the role of prominence in language perception and general communication.

2 Study 1a

Zimmermann, J. T., Wehrle, S., Cangemi, F., Grice, M., & Vogeley, K. (2020). Listeners and Lookers: Using pitch height and gaze duration for inferring mental states. In *Proceedings of the 10th International Conference on Speech Prosody* (pp. 290–294). https://doi.org/10.21437/SpeechProsody.2020-59

Listeners and Lookers: Using Pitch Height and Gaze Duration for Inferring Mental States

Juliane T. Zimmermann[1], *Simon Wehrle*[2], *Francesco Cangemi*[2], *Martine Grice*[2], *Kai Vogeley*[1,3]

[1]Department of Psychiatry, Faculty of Medicine and University Hospital Cologne, University of Cologne, Germany; [2]IfL – Phonetics, University of Cologne, Germany; [3]Institute of Neuroscience and Medicine, Cognitive Neuroscience (INM-3), Research Centre Juelich, Germany

juliane.zimmermann@uk-koeln.de

Abstract

Pitch height and gaze duration are used to infer other people's mental states, e.g. their attentional focus, attitudes or emotions. To shed light on the interplay of these two cues we varied pitch height in German utterances and gaze duration in a paradigm including a virtual character and different objects. At a group level, greater pitch height and longer gaze duration on a given object similarly increased participants' ratings of the perceived importance of that object to the virtual character. At the individual level, most participants showed a tendency to be influenced predominantly by only one of the two channels (pitch or gaze). The data suggest a high interindividual variability in the employment of the different, potentially competing nonverbal cues used in estimating the thoughts and judgment of another person.

Index Terms: pitch height, gaze duration, mental state, individual behavior, prominence

1. Introduction

Interaction substantially relies on complex and multimodal nonverbal communication [1]. Meaning can be conveyed and extracted in speech material beyond lexical content on the basis of prosody [2]. In the visual domain, gaze behavior plays a crucial part in conveying and inferring information in social communication [3]. Both of these types of nonverbal cues, prosody and eye gaze, provide us with important information in their respective domain. We can make use of them to infer others' mental states, e.g. their attentional focus, attitudes or emotions. In complex social encounters, most often a combination of more than one channel is involved.

1.1. Intonation as a key to inferring mental states

Prosody can be used to indicate that something is important, new or in the focus of attention, be it an aspect of conversation, or an object in the environment. In German this is achieved through pitch accent placement and type, cued primarily by fundamental frequency, perceived as pitch [4]. For successful communication, speakers take into account what the listener already knows ('givenness') and then apply prosody appropriately [5]. For listeners, pitch is especially relevant to the perception of prosodic prominence [6], indicating how far something is marked as important [7]. Even without speakers intentionally conveying this information, they may inadvertently transmit inferable prosodic cues to what is important in the current situation or to the speakers themselves [8]. Thus, not only 'face-value-importance' is communicated prosodically. Rather, by successfully decoding prosodic information, listeners can infer a speaker's intentions, thoughts and feelings; emotional states can also be encoded and decoded prosodically [9], [10].

1.2. Gaze as a key to inferring mental states

Gaze behavior is a very strong signal by which we express our inner experience. We direct our eyes towards objects we pay attention to and inform others about whether these objects are of general interest or importance in a specific situation [11], [12]. From an early age, our gaze is drawn towards new objects [13], which are likely to be more interesting and informative. We are also able to interpret observed gaze behavior. Another person's directed gaze can lead the observer to attend to the same direction [14]. Gaze directed towards objects can help us understand which object might be especially important in a given situation [15]. Moreover, gaze direction and duration are indicative of preferences [16], [17]: we tend to look longer and more often at preferred stimuli compared to non-preferred stimuli. Crucially, observers are able to interpret gaze duration [18] and direction [19], [20] towards preferred or desired objects.

1.3. Integrating visual and auditory cues

In real life, we are forced to make sense of complex stimuli from many different sources of information and to integrate them into a coherent representation of what is being communicated. A combination of auditory and visual information can be helpful in the interpretation of a message if the incoming information is difficult to understand (e.g. due to noise [21]), but can also be detrimental if both channels provide conflicting information [22]. Likewise, acoustic and visual information are integrated to infer how important a particular object might be for another person. Visual information, such as head nods or eyebrow raises, can increase prosodic prominence perception if it is already present, thus indicating an additive effect of visual and auditory information for prominence ratings [23]. When asked to identify prominent elements of spoken sentences presented in video sequences, the upper half of the head including the eye region is particularly informative [24]. However, it is unclear how exactly prosodic prominence and gaze are used to infer another person's mental states, e.g. importance ratings.

1.4. Study Design and Hypotheses

In the current study, we systematically compare the effect of the two information channels, prosodic prominence and gaze behavior, on the perceived attitude of an agent, i.e. a virtual character towards objects in her environment (importance judgement). More precisely, we manipulate pitch height and gaze duration, both allegedly attributable to the virtual character. Participants are asked to rate how important the object present

in the current situation is to the 'person' represented by the virtual character. We expected participants' perception of the virtual character's mental state to be affected by both a higher pitch excursion on the word referring to an object (suggesting a more prominent pitch accent type) and longer gaze duration of the virtual character towards that object. Specifically, we expected to find an increase of participants' ratings of importance of the object for the virtual character if the object was presented with a more prominent accent type and/or longer gaze duration compared to a less prominent accent type and shorter gaze duration. As it has been reported that less frequent words elicit greater prominence perception [25], [26], we also expected word frequency to have a general influence on ratings.

2. Material and Methods

We tested both the individual and combined influence of pitch height and gaze duration on participants' ratings of the importance of objects for a virtual character. We presented 106 different video sequences of a virtual character's face positioned above an object with a duration of 6.6 s. Depicted objects were different in each trial and each object was presented only once. One female virtual character was presented, corresponding to recordings from one female speaker. The movements performed by the agent were limited to the eyes. The agent's attention towards the object suggesting high importance was operationalized as an auditorily presented utterance with a higher pitch excursion and a longer gaze duration directed towards the object.

2.1. Experimental design

We systematically varied the factors 'pitch height' and 'gaze duration towards the object' on two levels. Pitch height on the accented syllable was either comparatively low or high. Gaze duration towards the object was either comparatively short (0.6 s) or long (1.8 s). Thus, we effectively created four conditions establishing a 2 x 2 experimental design: low pitch and long gaze, low pitch and short gaze, high pitch and short gaze, high pitch and long gaze.

2.2. Selection of objects

We selected 106 different images of objects from a pre-established and well-characterized set of images [27] based on their referential expressions. To reduce any possible influence of the number of syllables on the perception of word prominence, we only selected words with two syllables and penultimate stress. These were most frequent in the set and allowed us to avoid any interference effects due to word boundary effects. Additionally, we partly excluded well-known and often used homonyms.

2.3. Auditory stimulus material

Auditory stimuli comprising the German two-syllable words denoting the 106 different objects including the definite article (e.g. "der Toaster": "the toaster") were created from an H*-accented rendition of each of the 106 target phrases produced by a trained female speaker. An analysis of H* and L+H* on a subset of target words by this speaker indicated that she mainly modulated F0-peak height in differentiating between these two categories. Recordings took place in a soundproof booth, using an AKG C420L headset microphone connected to a computer running Adobe Audition via a USB audio interface (PreSonus AudioBox 22VSL). Stimuli were recorded with a sampling rate of 44100 Hz, 16 bit. We subsequently edited F0-peak height on the target words, so as to obtain a lower and a higher pitch peak (henceforth low and high), with a difference of 45 Hz. As other parts of the pitch contour were unchanged, higher peaks led to greater pitch excursions. Stimuli were tested for 'naturalness' and accent type by six trained phoneticians. Stimuli produced for the 'low' and 'high' condition were rated as sounding natural in 92.14 % and 74.37 % of cases, respectively, and were rated as H* and L+H*, respectively, in 83.65 % and 78.46 % of cases. The resulting speech stimuli were normalized to equal loudness. F0 was edited using smoothing [28], stylisation and resynthesis [29]. Examples are provided in the online multimedia files.

2.4. Visual stimulus material

Video sequences were created by arranging a picture of the female virtual character and an image of one of the 106 different objects in a vertical fashion (Figure 1). At the beginning and the end of the video, the agent exhibited idle gaze behavior, i.e. she performed gaze movements directed towards random locations in the environment. The agent fixated neither the object nor the participant during these phases.

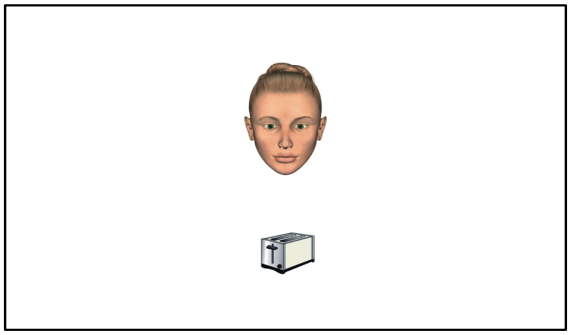

Figure 1: *Still of an example video.*

After 2.0 s, the virtual character looked at the participant for 1.0 s. This gaze was included to induce the experience of mutual interaction between virtual character and participant [30]. Subsequently, the virtual character directed her gaze towards the object (Fig. 1). This fixation lasted either 0.6 s (short gaze conditions) or 1.8 s (long gaze conditions). These durations are based on findings from human-robot interaction regarding different gaze durations and their perception [31]. Afterwards, the agent looked at the participant again. This gaze sequence (participant, object, participant) was preceded and followed by a blink, i.e. the agent's eyes closing for 0.1 s to simulate naturalistic blinking behavior. In addition, some video animations included an additional blink during one or both of the first and second idle, non-communicative phases. After the described gaze sequence, the virtual character continued gazing at random locations until the end of the video, lasting for either 2.0 or 0.8 s depending on short or long gaze conditions, so as to keep the total presentation duration of the object image of 6.6 s constant across videos. All images of the agent's face were taken from a study investigating the perception of gaze direction [32].

Video creation and integration of auditory stimuli was carried out using Python [33] and the FFmpeg module [34]. We created a total of 424 videos (106 per condition). Example videos are provided in the online multimedia files.

2.5. Participants and procedure

We recruited 64 monolingual native German speakers aged between 18 and 65 via an online platform (www.prolific.ac). They were reimbursed with 3.25 Euro for their participation. The study was performed in SoSci Survey [35]. Participants were instructed to imagine that the utterances they perceived were produced by the character on screen. They were informed that the character can convey the importance of the object. Participants were then instructed to answer the same question after each trial: "How important is the object to the virtual character?" (German: "Wie wichtig findet die Figur das abgebildete Objekt?"). Participants were presented with half of the stimuli to keep the task short. Each trial consisted of a video and its subsequent rating. Items were presented in randomized fashion. Each video sequence was followed by a screen asking for the rating on a scale from 1 to 4: 1="not important at all", 2="rather unimportant", 3="rather important", 4="very important".

2.6. Analysis

Data was analysed with R [36] in RStudio [37]. A Bayesian ordinal model (r package 'brms' [38]) was fitted to the data. Fixed effects for participants' ratings were 'gaze duration', 'pitch height', their statistical interaction and the logarithmized and z-transformed values for word frequency of the objects in German [39]. As random effects, we included random intercepts and slopes for the 'subject' effect, and random intercepts for the 'object' effect. A weakly informative prior was used (intercept prior: normal distribution, $M = 2.5$, $SD = 1.5$; slope priors: normal distribution, $M = 0$, $SD = 2$; SD prior: normal distribution, $M = 0$, $SD = 2$). The model ran with four sampling chains of 12,000 iterations each and a warm-up period of 2,000 iterations.

3. Results

The condition characterized by low pitch height and short gaze duration yielded the lowest mean ratings. The condition with both high pitch and long gaze duration yielded the highest mean ratings. The conditions with either increased pitch height or longer gaze duration yielded mean ratings in a middle range between the two aforementioned conditions. Mean ratings within the four conditions corroborated the initial hypotheses (Fig. 2).

Overall, there is strong evidence for our model as opposed to the model not including the factors 'pitch height' and 'gaze duration' ($BF_{10} > 1000$). Higher pitch increased the ratings by 0.56 standard deviations (SD) on the latent rating scale, 95% CI = [0.33, 0.79]. Likewise, longer gaze duration also increased the ratings ($\hat{\beta} = 0.65$, 95% CI = [0.39, 0.91]. In this study, both effects had comparable effect sizes. Their statistical interaction did not affect ratings ($\hat{\beta} = 0.02$, 95% CI = [-0.14, 0.18]). Higher word frequency increased the ratings ($\hat{\beta} = 0.06$, 95% CI = [0.01, 0.12]). The random subject effects were considerable in the model (random intercepts: $\hat{\beta} = 0.57$, 95% CI = [0.47, 0.70]; random effect of pitch height: $\hat{\beta} = 0.87$, 95% CI = [0.71, 1.06]; random effect of gaze duration: $\hat{\beta} = 1.02$, 95% CI = [0.83, 1.24]), except for the random interaction effect ($\hat{\beta} = 0.12$, 95% CI = [0.00, 0.35]), which was not statistically robust. The random object effect, however, was statistically robust ($\hat{\beta} = 0.18$, 95% CI = [0.11, 0.24]).

3.1. Explorative analysis of individual behavior

At the individual level, the effects of pitch height and gaze duration accounted for a change of 0.87 and 1.02 SD on our rating scale, respectively. Therefore, we further investigated to what extent the factors predicted the ratings for each individual participant. Figure 3 shows the individual slope coefficients for the two factors for each subject. Participants' ratings tended to be influenced by either 'pitch height' or 'gaze duration' rather than by both factors in combination. This was mirrored by a negative correlation ($\hat{\beta} = -0.48$, 95% CI = [-0.68, -0.23] of the factors 'pitch height' and 'gaze duration' within the random subject effect.

To identify possible subgroups based on cue 'preference', we applied a hierarchical cluster analysis [40] using Euclidian distance and Ward's method. The resulting classification suggested a two- or three-cluster solution. The clustering is included in Fig. 3, showing the three distinct groups. Due to the degree to which participants took into account pitch height and gaze duration for their ratings, we labelled them 'Listeners', 'Lookers' and 'Neither'. In the two-cluster solution, 'Listeners' and 'Neither' clustered together.

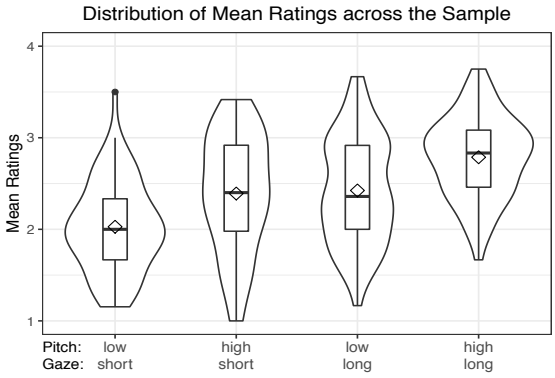

Figure 2: *Distribution of participants' mean ratings of stimuli. The range of the y-axis equals the total rating scale (1-4). Diamonds indicate means across subjects.*

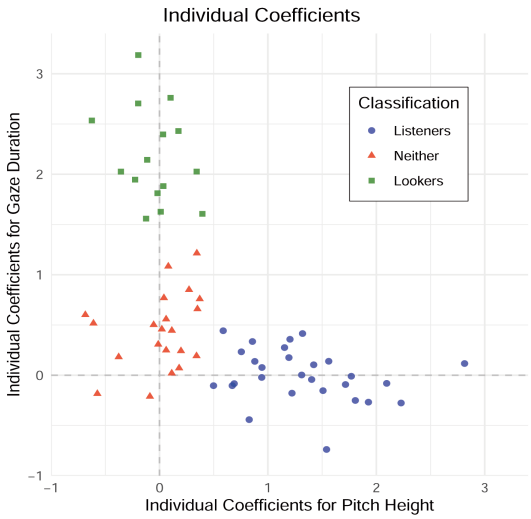

Figure 3: *Individual slope coefficients for pitch height (x-axis) and gaze duration (y-axis) by participant.*

4. Discussion

We investigated how participants' ratings of the ascription of a 'person's' attitude, here the perception of object importance, were affected by pitch height (of utterances referring to target objects) and gaze duration (directed towards target objects by the virtual agent). According to our initial hypothesis, both factors substantially affect the ratings of participants.

The condition characterized by low pitch height and short gaze duration towards the object resulted in the lowest mean ratings of importance. Highest mean ratings were registered for the condition characterized by high pitch and long gaze duration. The mean ratings for the two 'mixed' conditions in which only one of the two signals indicates importance were in the medium range between these two extremes. The effects of pitch height and gaze duration on participants' ratings presented as similar and statistically robust in our analysis. Our findings corroborate previous findings showing that acoustic and visual cues are relevant sources from which the mental states of others can be inferred [8], [9], [18], [19]. Interestingly, most participants were influenced by only one of the two cues, dividing the whole group into either 'Listeners', 'Lookers' or 'Neither'.

Our data do not provide evidence for interactional effects of pitch height and gaze duration in the current task. This is in line with previous studies investigating the interplay of prosody and visual body cues and showing no interaction of the two: Only an additive effect of eyebrow raises and pitch accents was shown for the perception of word prominence [23] whilst no interactional effect of general visual facial information and prosody on prosodic prominence ratings was observed [24]. In our study, participants tended to not take into account both at the same time, so that we cannot conclude that effects in our study add up to contribute to the perception of object importance to the virtual agent.

Higher word frequency was associated with higher importance ratings. At first glance, this seems to contradict the observation that infrequent words elicit higher prominence perception [25], [26]. However, in our dataset, word frequency was intertwined with other properties that also affect the perception of importance: the five most frequent words in our dataset were the German words for 'car', 'key', 'eye', 'plane' and 'finger'. The five least frequent words were the German words for 'spinning wheel', 'doorknob', 'seal', 'spinning top' and 'roller skate'. We assume that relevance for everyday life affected the ratings, so that word frequency and general importance were correlated in our study. Taking a look at the five 'most important'-rated items ('key', 'traffic light', 'spoon', 'brush', 'sun') and the five 'least important'-rated items ('desk', 'peanut', 'church', 'sandwich', 'seal') corroborates this notion.

Variance introduced by individual participants was substantial. Great individual variability has been reported for influence of pitch on the perception of prosodic prominence, i.e. the degree to which words are perceived as highlighted or important [7], [25]. Moreover, the perception of prosodic prominence is not only influenced by speaker and listener characteristics, but also by a combination of both [44]. Interpretation of directional gaze cues also depends on individuals (e.g biological sex [41]).

We found that participants' ratings tended to be influenced by either pitch height or gaze duration or by neither of the two cues, but never by both at the same time. This led us to the identification of groups of 'Listeners', 'Lookers' and 'Neither'. We reject the two-cluster solution because it is theoretically not convincing to cluster participants making use of pitch height with participants making use of neither cue. The existence of a 'neither'-group in our study does simply allow for the conclusion that these people did not take into account either cue. Other possible explanations are suggested in the following paragraph. As for the differentiation of participants into 'listeners' and 'lookers', other studies have provided similar findings. In a production study, it was shown that participants increase speaking efforts and change their gaze behavior to improve communication [42]. However, the authors did not find a strong correlation of both measures. This supports the idea that the majority of people focus more on one channel than the other. Another study reports that people tend to produce pitch accent categories either by altering the shape of the F0-contour peak or its timing [43]. In perception studies, similar results have been reported: Persons rating prosodic prominence tend to concentrate on either prosodic cues or visual facial information [24]. Similarly, studies concerned with the perception of prosodic prominence as well as its reproduction report a division into one group of subjects that relies mainly on pitch and another that relies more on other aspects (such as word frequency) [7], [25], [44].

There are some limitations to the study. First of all, the findings of this reductionistic design cannot be easily transferred to any kind of complex social situation. We created a situation devoid of variation of other cues usually present in a comparable real-life situation. While the voice stimuli were derived from natural speech, the virtual character was not seen to move her mouth along with the presentation of the utterance. Moreover, neither the virtual agent nor the objects were photo-realistic depictions. Second, people were informed that the virtual character is able to convey the importance of the object. This information might have led participants to actively search for a cue to make sense of the otherwise uninformative setting and stop searching once one valid cue (out of two possible cues) was identified as a reliable source of information. This might have led participants to not make use of both cues, which would explain why we did not find an interaction of pitch height and gaze duration. Third, participants were required to indicate how important the character finds the depicted object. Even with the information that the agent can indeed convey importance, the task still is rather vague and relies on subjects' perceptual and mentalizing skills. This could lead to participants having trouble integrating the cues as meaningful in this rather unnatural setting or to them being reluctant to assign mental capacities to a virtual agent in the first place. We did not collect participants' ratings of general object importance.

5. Conclusion

Pitch excursion and gaze duration can be used to infer the mental state of a virtual character. Persons differ in terms of the degree to which they make use of these cues to infer the importance of an object to the virtual character.

The study's limitations might be overcome by using a more life-like experimental environment, including variation of other cues, along with a more engaging task. Future studies could benefit from further investigation into the individual factors accounting for the substantial amount of variability found in the present study.

6. Funding

The study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 281511265 – SFB 1252.

7. References

[1] J. K. Burgoon, L. K. Guerrero, and V. Manusov, "Nonverbal signals," in *The SAGE Handbook of Interpersonal Communication*, 4th ed., M. L. Knapp and J. A. Daly, Eds. Thousand Oaks: SAGE Publications, Inc., 2011.

[2] J. D. O'Connor and G. F. Arnold, *Intonation of Colloquial English*, 2nd ed. London: Longman, 1973.

[3] M. Argyle, R. Ingham, F. Alkema, and M. McCallin, "The Different Functions of Gaze," *Semiotica*, vol. 7, no. 1, pp. 19–32, 1973.

[4] C. Féry and F. Kügler, "Pitch accent scaling on given, new and focused constituents in German," *Journal of Phonetics*, vol. 36, no. 4, pp. 680–703, Oct. 2008.

[5] M. Breen, E. Fedorenko, M. Wagner, and E. Gibson, "Acoustic correlates of information structure," *Language and Cognitive Processes*, vol. 25, no. 7–9, pp. 1044–1098, Sep. 2010.

[6] D. Arnold, P. Wagner, and H. Baayen, "Using generalized additive models and random forests to model prosodic prominence in German," 2013.

[7] S. Baumann and B. Winter, "What makes a word prominent? Predicting untrained German listeners' perceptual judgments," *Journal of Phonetics*, vol. 70, pp. 20–38, Sep. 2018.

[8] C. Kaland, E. Krahmer, and M. Swerts, "White Bear Effects in Language Production: Evidence from the Prosodic Realization of Adjectives," *Lang Speech*, vol. 57, no. 4, pp. 470–486, Dec. 2014,

[9] W. Thompson and L.-L. Balkwill, "Decoding speech prosody in five languages," *Semiotica*, vol. 2006, pp. 407–424, Jan. 2006.

[10] K. R. Scherer, "Vocal affect expression: a review and a model for future research," *Psychol Bull*, vol. 99, no. 2, pp. 143–165, Mar. 1986.

[11] A. L. Yarbus, "Eye Movements During Perception of Complex Objects," in *Eye Movements and Vision*, A. L. Yarbus, Ed. Boston, MA: Springer US, 1967, pp. 171–211.

[12] A. Klami, "Inferring task-relevant image regions from gaze data," in *2010 IEEE International Workshop on Machine Learning for Signal Processing*, 2010, pp. 101–106.

[13] R. L. Fantz, "Visual experience in infants: Decreased attention to familiar patterns relative to novel ones," *Science*, vol. 146, no. 3644, pp. 668–670, Oct. 1964.

[14] J. Driver, G. Davis, P. Kidd, E. Maxwell, P. Ricciardelli, and S. Baron-Cohen, "Gaze Perception Triggers Reflexive Visuospatial Orienting," *Visual Cognition*, vol. 6, no. 5, pp. 509–540, Oct. 1999.

[15] A. Freire, M. Eskritt, and K. Lee, "Are Eyes Windows to a Deceiver's Soul? Children's Use of Another's Eye Gaze Cues in a Deceptive Situation," *Dev Psychol*, vol. 40, no. 6, pp. 1093–1104, Nov. 2004.

[16] T. Chuk, A. B. Chan, S. Shimojo, and J. H. Hsiao, "Mind reading: Discovering individual preferences from eye movements using switching hidden Markov models," in *Proceedings of the 38th Annual Conference of the Cognitive Science Society, CogSci 2016*, 2016, pp. 182–187.

[17] S. Shimojo, C. Simion, E. Shimojo, and C. Scheier, "Gaze bias both reflects and influences preference," *Nat Neurosci*, vol. 6, no. 12, pp. 1317–1322, Dec. 2003.

[18] S. Einav and B. M. Hood, "Children's use of the temporal dimension of gaze for inferring preference," *Dev Psychol*, vol. 42, no. 1, pp. 142–152, Jan. 2006.

[19] K. Lee, M. Eskritt, L. A. Symons, and D. Muir, "Children's use of triadic eye gaze information for 'mind reading,'" *Dev Psychol*, vol. 34, no. 3, pp. 525–539, May 1998.

[20] S. Baron- Cohen, R. Campbell, A. Karmiloff- Smith, J. Grant, and J. Walker, "Are children with autism blind to the mentalistic significance of the eyes?," *British Journal of Developmental Psychology*, vol. 13, no. 4, pp. 379–398, 1995.

[21] W. H. Sumby and I. Pollack, "Visual Contribution to Speech Intelligibility in Noise," *The Journal of the Acoustical Society of America*, vol. 26, no. 2, p. 212, Jun. 2005.

[22] H. McGurk and J. Macdonald, "Hearing lips and seeing voices," *Nature*, vol. 264, no. 5588, pp. 746–748, Dec. 1976.

[23] E. Krahmer, Z. Ruttkay, M. Swerts, and W. Wesselink, "Perceptual evaluation of audiovisual cues for prominence," in *INTERSPEECH*, 2002.

[24] M. Swerts and E. Krahmer, "Facial expression and prosodic prominence: Effects of modality and facial area," *Journal of Phonetics*, vol. 36, no. 2, pp. 219–238, Apr. 2008.

[25] J. Roy, J. Cole, and T. Mahrt, "Individual differences and patterns of convergence in prosody perception," *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, vol. 8, no. 1, p. 22, Sep. 2017.

[26] J. Cole, Y. Mo, and M. Hasegawa-Johnson, "Signal-based and expectation-based factors in the perception of prosodic prominence," *JLP*, vol. 110, pp. 425–452, Jan. 2010.

[27] B. Rossion and G. Pourtois, "Revisiting Snodgrass and Vanderwart's object pictorial set: the role of surface detail in basic-level object recognition," *Perception*, vol. 33, no. 2, pp. 217–236, 2004.

[28] F. Cangemi, *mausmooth [Praat script]*. 2015. Retrieved from http://ifl.phil-fak.uni-koeln.de/sites/linguistik/Phonetik/mitarbeiterdateien/fcangemi/mausmooth.praat

[29] M. Winn, *Fade in, Fade out [Praat script]*. 2014. Retrieved from www.mattwinn.com/praat/RampOnsetAndOrOffset.txt

[30] N. Emery, "The Eyes Have It: The Neuroethology, Function and Evolution of Social Gaze," *Neuroscience and biobehavioral reviews*, vol. 24, pp. 581–604, Sep. 2000.

[31] N. Pfeiffer-Lessmann, T. Pfeiffer, and I. Wachsmuth, "An Operational Model of Joint Attention - Timing of Gaze Patterns in Interactions between Humans and a Virtual Human," in *Proceedings of the 34th annual conference of the Cognitive Science Society*, 2012, pp. 851–856.

[32] H. Eckert, "Erzeugung von Blickreizen virtueller Charaktere mit ambiger kommunikativer Absicht mittels systematischer Variierung zweier Faktoren einer Blickbewegung - Anfangsblick und Blickziel," University of Cologne, 2017.

[33] G. Van Rossum, *Python tutorial. Technical Report CS-R9526*. Amsterdam: Centrum voor Wiskunde en Informatica, 1995.

[34] FFmpeg Developers, *FFmpeg Tool [Software]*. 2018.

[35] D. J. Leiner, *SoSci Survey (Version 3.2.01-i) [Computer Software]*. 2018.

[36] R Core Team, *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2019.

[37] RStudio Team, *RStudio: Integrated Development for R*. Boston, MA: RStudio, Inc., 2016.

[38] P.-C. Bürkner, "brms: An R Package for Bayesian Multilevel Models Using Stan," *Journal of Statistical Software*, vol. 80, no. 1, pp. 1–28, Aug. 2017.

[39] M. Brysbaert, M. Buchmeier, M. Conrad, A. M. Jacobs, J. Bölte, and A. Böhl, "The Word Frequency Effect," *Experimental Psychology*, vol. 58, no. 5, pp. 412–424, Jan. 2011.

[40] D. Müllner, "fastcluster: Fast Hierarchical, Agglomerative Clustering Routines for R and Python," *Journal of Statistical Software*, vol. 53, no. 1, pp. 1–18, May 2013.

[41] A. Bayliss, G. Di Pellegrino, and S. Tipper, "Sex differences in eye gaze and symbolic cuing of attention," *The Quarterly journal of experimental psychology. A, Human experimental psychology*, vol. 58, pp. 631–50, Jun. 2005.

[42] V. Hazan, O. Tuomainen, J. Kim, and C. Davis, "The effect of visual cues on speech characteristics of older and younger adults in an interactive task," in *Proceedings of the 19th International Congress of the Phonetic Sciences*, 2019, pp. 815–819.

[43] O. Niebuhr, M. D'Imperio, B. Gili Fivela, and F. Cangemi, "Are There 'Shapers' and 'Aligners'? Individual Differences in Signalling Pitch Accent Category," 2011, pp. 17–21.

[44] P. Wagner, A. Ćwiek, and B. Samlowski, "Exploiting the speech-gesture link to capture fine-grained prosodic prominence impressions and listening strategies," *Journal of Phonetics*, Jul. 2019.

3 Study 1b

Listeners and Lookers Investigated Further – Their Use of Pitch Height and Gaze Duration for Inferring Mental States is Mirrored by Gaze Behavior

Supplementary material for study 1b can be found in appendix 9.1.

3.1 Background and Aim of the Study

In preparation for study 1c which includes autistic participants, study 1b was carried out to establish the paradigm presented in study 1a (Zimmermann et al., 2020) in a controlled laboratory setting. Participants of study 1b include those that entered the comparison group in study 1c. A further aim of study 1b was to assess attention during the rating task allowing for further characterization of the different subgroups' behavior by means of eye-tracking. Subsequently, memory performance was assessed to detect potential traces of different attentional stances during encoding.

Study 1b was expected to replicate the main results of study 1a regarding participants' rating behavior (Zimmermann et al., 2020), i.e. both higher pitch (as opposed to lower pitch) of an utterance referring to the object and longer gaze duration of the virtual character (as opposed to shorter gaze duration) were expected to lead to higher ratings of importance of the object for the virtual character. Moreover, participants were expected to cluster into the three subgroups identified in study 1a: (i) "Listeners" based their ratings exclusively on pitch height, (ii) "Lookers" based their ratings exclusively on gaze duration, and (iii) a group of "Neithers" did not base their ratings on either cue. Regarding eye-tracking, "Lookers" were expected to look at the virtual character's eyes for longer and fixate the object for a shorter duration when compared to "Listeners" and "Neithers".

3.2 Material and Methods

The same paradigm reported on in study 1a (Zimmermann et al., 2020) was used: The individual and combined influence of pitch height and gaze duration on participants' ratings of the importance of objects for a virtual character were tested. To this end, 92 different video sequences of a virtual character's face positioned above an object were presented. Depicted objects were different in each trial. The movements performed by the virtual character were limited to the eyes. The character's attention towards the object suggesting high importance was operationalized as an auditorily presented utterance with a higher pitch excursion and a longer gaze duration directed towards the object. The factors pitch height and gaze duration towards the object' were varied on two levels. Pitch height on the accented syllable was either comparatively low or high. Gaze duration towards the object was either comparatively short (0.6 s) or long (1.8 s). Thus, four conditions were created establishing a 2 x 2 experimental design: low pitch and long gaze, low pitch and short gaze, high pitch and short gaze, high pitch and long gaze. Materials and procedures were adjusted to adhere to the laboratory setting and

are identical to those reported in study 1c (Zimmermann et al., 2024). Likewise, inclusion criteria for participants were identical to those described for the comparison group in study 1c (Zimmermann et al., 2024). Analytic methods differ slightly, since no participants with an ASD diagnosis were included in the current study.

3.2.1 Participants

42 monolingual German native speakers within an age range of 18 and 65 that had normal or corrected-to-normal vision as well as hearing were recruited. The study was conducted in accordance with the Declaration of Helsinki (World Medical Association, 2013) and approved by the Ethics Committee of the University Clinic of Cologne. To ensure that results were not influenced by lower cognitive performance, only participants with verbal and total intelligence scores of at least 85, as measured with the *WIE-III* (Aster et al., 2006), with attentional scores greater 80, as measured with the *D2* (Brickenkamp, 2002), and with maximally moderate depressive symptoms as measured with the *Beck Depression Inventory* (*BDI-II*; Beck et al., 1996) (i.e. with *BDI-II* scores < 18) were included. Sample characteristics are provided in table 1.

Sex	**Age**	***WIE* IQ verbal**	***WIE* IQ performance**	***WIE* IQ total**	***D2* total error corrected**	***BDI-II***
21 men 21 women	19–62 years *M* = 38.4 (*SD* = 15.1)	*M* = 112.0 (*SD* = 13.0)	*M* = 106.0 (*SD* = 13.5)	*M* = 110.0 (*SD* = 12.4)	*M* = 102.0 (*SD* = 9.81)	*M* = 4.81 (*SD* = 4.42)

Table 1: Sample characteristics (*N* = 42). WIE IQ = Hamburg-Wechsler-Intelligenz-Test für Erwachsene III (intelligence test for adults); D2 = d2 Aufmerksamkeits-Belastungs-Test (attention load test); BDI-II = Beck Depression Inventory, 2nd Version (questionnaire on depressive symptom severity).

3.2.2 Eye-Tracking

Eye-tracking data of four participants had to be discarded due to technical problems and did not enter the relevant analyses, i.e. the analysis of fixation durations and the Bayesian models for object recognition rates.

3.2.3 Analysis

The data was analyzed using *R* (R Core Team, 2019) in *RStudio* (RStudio Team, 2016). When reporting significance of t-tests and correlations, a 95%-confidence interval was assumed. Bayesian models (package *brms*; Bürkner, 2017; Bürkner & Vuorre, 2019) were fitted to the data of the rating task (i.e. ratings of importance), the corresponding eye-tracking data, and the

recognition task (i.e. the correctness of the responses). If not otherwise stated, dichotomous factors were deviation-coded, and continuous factors were z-transformed. In each model, random intercepts and slopes for *subject* as well as random intercepts for *object* were included. Estimated parameters are reported in terms of posterior means and 95% credibility intervals. The *emmeans* package (Lenth et al., 2021) was used to extract contrast coefficients. To investigate the evidence for or against the investigated effects, models were compared by calculating Bayes factors applying the *bayesfactor_models* function from the *bayestestR* package (Makowski et al., 2019) which uses bridge sampling (Gronau et al., 2020). Respective Bayes factors of model comparisons are reported. Interpretation of Bayes factors adheres to Lee and Wagenmakers (2014). All models ran with four sampling chains of 12,000 iterations each including a warm-up period of 2,000 iterations.

Fixed effects used in the Bayesian ordinal model for participants' ratings were *pitch height, gaze duration* and the logarithmized values for *word frequency*. A weakly informative prior was used to fit the described model as well as the models used for comparison (intercept prior: normal distribution, $M = 2.5$, $SD = 1$; slope priors for pitch height and gaze duration: normal distribution, $M = 0.5$, $SD = 1$; slope prior for their interaction: normal distribution, $M = 0$, $SD = 1$; slope prior for word frequency: normal distribution, $M = 0.05$, $SD = 0.2$; SD prior: normal distribution, $M = 0$, $SD = 2$; LKJ prior: 1).

In study 1a, three subgroups were identified by hierarchical cluster analysis (Müllner, 2013) using Euclidian distance and Ward's method on the individual slope coefficients for *pitch height* and *gaze duration* from the Bayesian ordinal model for participants' ratings. This analysis was repeated with the current dataset, this time clustering the sample into three subgroups deliberately.

For the analysis of the influence of *cluster* and *gaze duration* on the duration of fixations within the three regions of interest, eye-tracking data starting at the onset of the gaze cue (= onset of the auditory stimulus) was included. A proportional value for fixation duration, namely fixation duration directed towards the region of interest divided by the video duration starting at cue onset, was modelled separately for each region. Bayesian linear zero-inflated beta models (r package *brms*; Bürkner, 2017; Bürkner & Vuorre, 2019) were fitted to the data. Fixed effects were *cluster* and *gaze duration.* Weakly informative priors were used (intercept prior: normal distribution, $M = 0.5$, $SD = 0.5$; slope priors: normal distribution, $M = 0$, $SD = 0.5$; SD priors: normal distribution, $M = 0$, $SD = 0.5$; phi priors: normal distribution, $M = 0.5$, $SD = 0.5$; zi prior: $M = 0.2$, $SD = 0.5$; LKJ prior: 1).

In the Bayesian logistic binomial regression model for object recognition in the memory task, the following fixed effects were included: untransformed proportional values for *participants' gaze duration* towards the object region during the rating task; the logarithmized values of *word frequency;* the *number of trials that had passed since object presentation*. This model was compared to models that additionally included *pitch height* and the virtual character's *gaze duration* towards the object. Weakly informative priors were used (intercept prior: normal distribution, $M = 0$, $SD = 0.5$; slope priors: normal distribution, $M = 0$, $SD = 0.5$; SD priors: normal distribution, $M = 0$, $SD = 0.5$; LKJ prior: 1). Results are reported on the log-odds scale.

3.3 Results

3.3.1 Rating Behavior

The condition characterized by low pitch height and short gaze duration yielded the lowest mean ratings ($M = 2.12$, $SD = 0.36$). The condition with both high pitch and long gaze yielded the highest mean ratings ($M = 2.76$, $SD = 0.44$). The conditions with either increased pitch height ($M = 2.56$, $SD = 0.56$) or longer gaze duration ($M = 2.38$, $SD = 0.47$) yielded mean ratings between the two aforementioned conditions. Mean ratings (see figure 1) therefore replicate the general pattern reported in the web-based study 1a (Zimmermann et al., 2020).

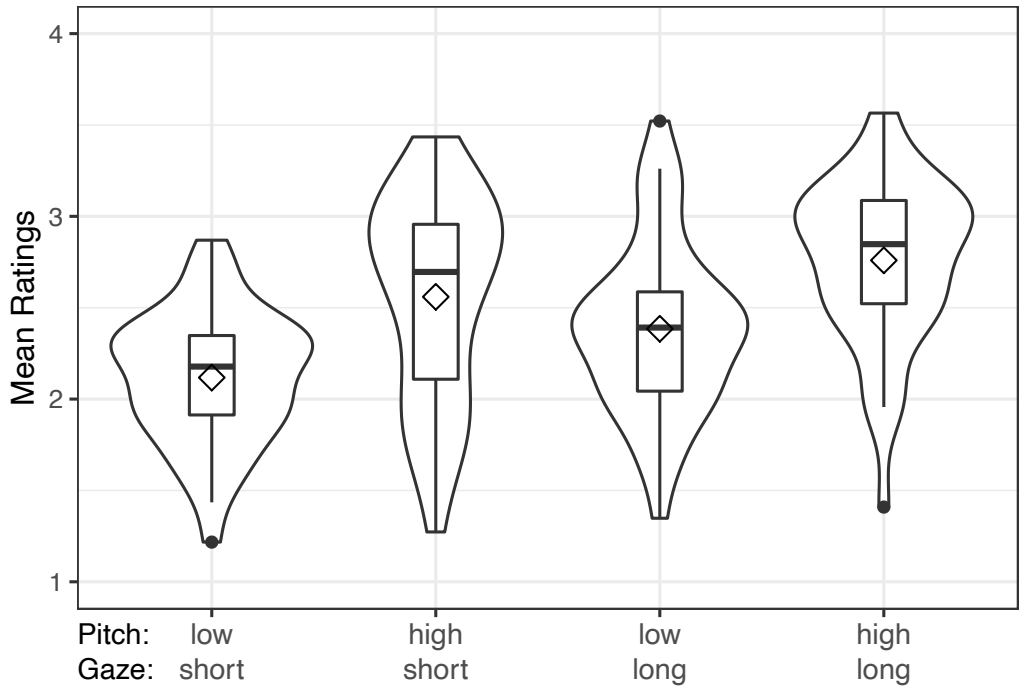

Figure 1: Distribution of participants' mean ratings of stimuli. The range of the y-axis equals the total rating scale (1–4). Diamonds indicate means across subjects. Horizontal lines indicate medians.

Model comparisons indicated extreme evidence (BF > 1000) for the influence of both *pitch height* and *gaze duration*: Higher pitch increased the ratings by 0.65 standard deviations (*SD*) on the latent rating scale, 95% CI = [0.39, 0.90]. Likewise, longer gaze duration increased the ratings (b = 0.39, 95% CI = [0.08, 0.70]. As expected, higher word frequency also increased the ratings (b = 0.08, 95% CI = [0.02, 0.14], BF = 54.89). Subjects differed with regard to rating baselines and the influence the factors had on their ratings (random intercepts: b = 0.44, 95% CI = [0.34, 0.56]; random effect of *pitch height:* b = 0.80, 95% CI = [0.62, 1.02]; random effect of *gaze duration:* b = 1.00, 95% CI = [0.79, 1.27]). Additionally, the object had an effect on the ratings (b = 0.18, 95% CI = [0.13, 0.24]). Comparing the model with one including the interaction of *pitch height* and *gaze duration* revealed strong evidence against an interaction effect (BF = 0.04). Taken together, the main results from the web-based study 1a could be replicated in a controlled laboratory setting.

3.3.1.1 Individuality

Similar to the findings in study 1a, there was substantial inter-individual variability. Figure 2 shows the individual slope coefficients for *pitch height* and *gaze duration* for each subject. Participants' ratings tended to be influenced by either one or the other factor rather than by both factors in combination. This was mirrored by a negative correlation (b = –0.39, 95% CI = [–0.64, –0.09]) of the factors *pitch height* and *gaze duration* within the random subject effect. In study 1a, three subgroups were identified based on clustering according to cue "preference". These subgroups were labelled "Listeners", "Lookers" and "Neithers", based on the degree to which they took into account pitch height and gaze duration for their ratings. Clustering the current sample into the three respective subgroups resulted in 26 % of participants being identified as "Lookers", 31 % as "Listeners" and 43 % as "Neithers" (see figure 2).

Some participants reported having noticed both the change in intonation as well as in gaze duration, e.g. two participants in the "Lookers" cluster that are located more towards the center of the graph (see figure 2). One of these participants reported to have decided at one point during the rating task to focus on gaze duration rather than intonation.

To shed some light on the rating strategies of the "Neithers" in this study, an investigation of the subgroup's feedback provided in the questionnaire after the experiment was conducted. In fact, some participants of the "Neithers" subgroup correctly identified that both the intonation as well as the virtual character's gaze duration towards the object changed throughout the experiment. However, they did not make use of the cues in the expected fashion, i.e. they did not increase their ratings for objects presented with high pitch or long gaze duration. This may

in part be explained by participants additionally using other (non-informative) cues. Five participants (three of these in the "Neithers" subgroup) reported object-related rating behavior: They considered the object's characteristics (its animacy, entertaining quality, potential benefit or danger, as well as the participant's own experience with the object) and the virtual character's age, sex and appearance. Some of these "Neithers" thought about the virtual character's hobbies, preferences or desires. Participants in the "Neithers" subgroup that did not consider object- or character-specific aspects concentrated on parameters of the virtual character's gaze behavior which were not systematically varied such as mutual gaze, blinking, and directionality of idle gaze.

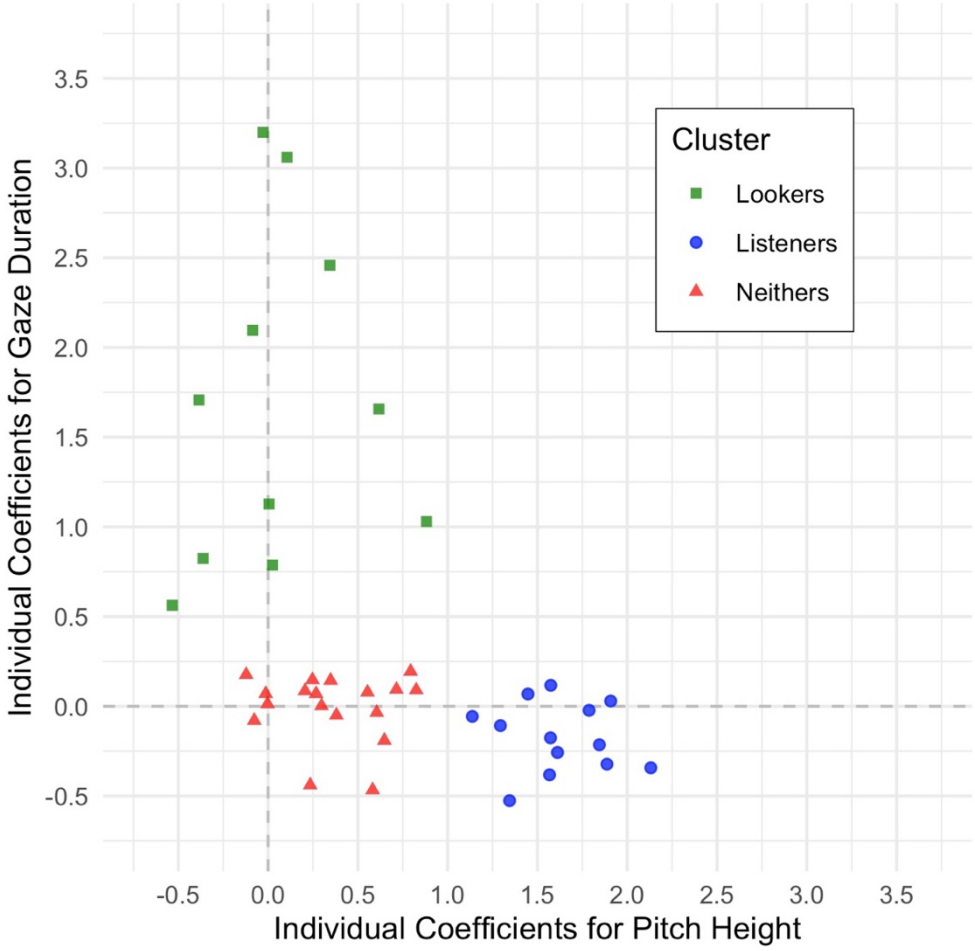

Figure 2: Individual slope coefficients (from Bayesian ordinal modelling used to model participants' ratings) for pitch height (x-axis) and gaze duration (y-axis) by participant. The coefficients are combined to one coordinate for each participant. Cluster labelling is based on the individual's cue-use: The data of all participants entered a hierarchical cluster analysis in which individuals were deliberately clustered into three groups based on their slope coefficients, using Euclidian distance and Ward's method.

3.3.1.2 Exploratory Analysis

Exploratory correlation analyses were carried out for the model coefficients for pitch height and gaze duration in combination with the *AQ*, *EQ*, *SQ* as well as with the *SPQ visual* and *auditory* scores. In sum, no statistically reliable correlations were found for the *AQ* and *SQ*, but two statistically noteworthy relationships for the *EQ* and *SPQ* regarding *pitch height* coefficients and *gaze duration* coefficients: correlations between *EQ* scores and *pitch height* coefficients (r_S = .368, p = .016) suggested a tendency for more empathic traits to be associated with taking into account pitch height to a greater extent. This relationship was not found for *gaze duration* coefficients (r_S = –.180, p = .255). Higher *SPQ visual* scores (indicating lower visual sensitivity) were linked to taking pitch height into account to a greater extent (r_S = .312, p = .044), and taking gaze duration into account to a lesser extent (r_S = –.340, p = .028). Such a relationship was not found between *SPQ auditory* scores and the coefficients for *pitch height* (r_S = .043, p = .789) and *gaze duration* (r_S = .019, p = .906). Correlating *AQ* scores with *pitch height* coefficients (r_S = –.297, p = .056) revealed a tendency for autistic-like traits to be linked to taking into account pitch height less. Such a link was not observed for *AQ* scores and *gaze duration* coefficients (r_S = .024, p = .881). *SQ* scores were not linked to *pitch height* coefficients (r_S = –.029, p = .854). A tendency for *SQ* scores to be linked to *gaze duration* coefficients (r_S = –.272, p = .081) was observed, suggesting a tendency for higher systemizing needs to be associated with taking into account gaze duration to a lesser degree.

3.3.2 Gaze Fixation Durations

Across the sample, participants spent more time looking at the eye region (M = 3.88 s, SD = 1.38) than at the object (M = 1.24 s, SD = 1.05) and head region (M = 1.10 s, SD = 0.61) (see figure 3).

Within the three clusters, rating behavior was reflected by fixation durations within the three regions: Compared to the other groups, the group of "Lookers" tended to look longer at the eye region ("Lookers": M = 4.70 s, SD = 1.08; "Listeners": M = 3.41 s, SD = 1.43; "Neithers": M = 3.63 s, SD = 1.36), but spent less time fixating the object region ("Lookers": M = 0.57 s, SD = 0.26; "Listeners": M = 1.66 s, SD = 1.34; "Neithers": M = 1.41 s, SD = 1.00). Fixation durations within the head region (not including the eye region) were similar between clusters ("Lookers": M = 1.07 s, SD = 0.74; "Listeners": M = 1.11 s, SD = 0.69; "Neithers": M = 1.12 s, SD = 0.50).

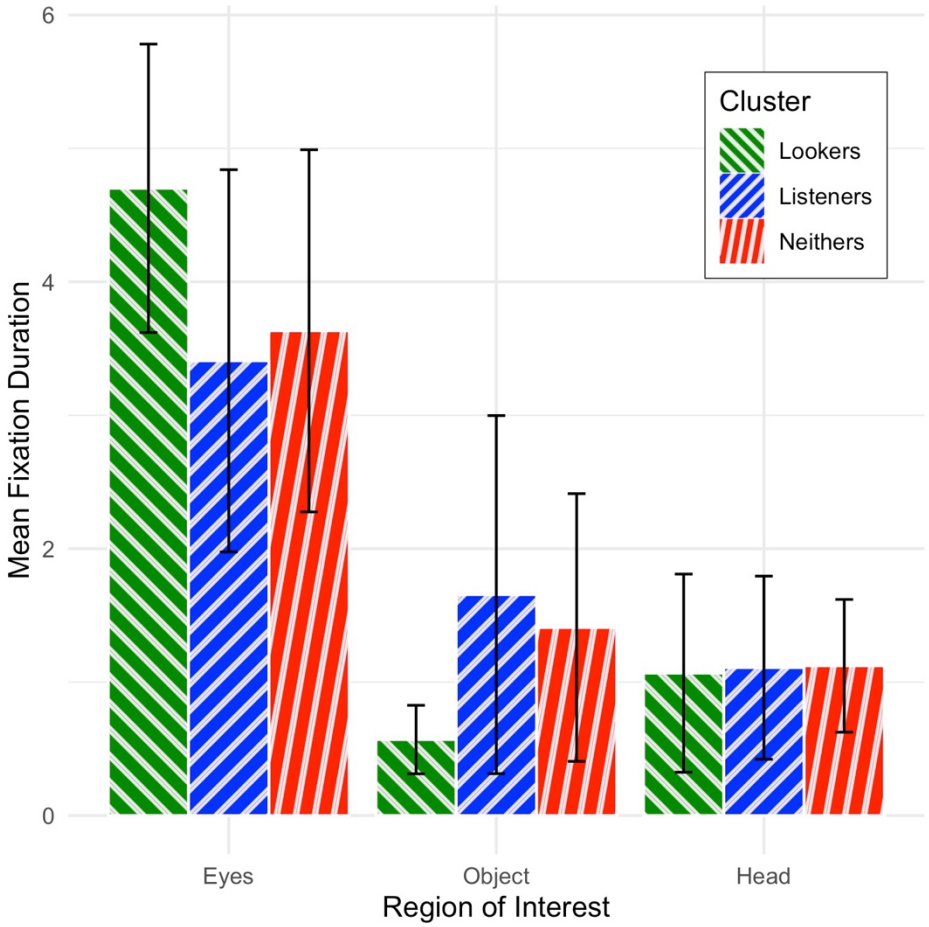

Figure 3: Mean fixation durations and standard deviations for the three clusters within the three regions of interest. The total possible per-trial fixation duration is 6.6 s.

The influence of *cluster* on fixation duration beginning at gaze cue onset (which coincides with the onset of the auditory stimulus) was analyzed separately for each region of interest. In support of the above reported cluster-dependent gaze patterns, extreme evidence for an effect of *cluster* was found within the eye region (BF > 100) and the object region (BF > 100). Very strong evidence for an effect of *cluster* (BF = 49.95) was also found in the head region, which was mainly driven by a tendency of "Neithers" to fixate the head region for a longer duration after stimulus onset than the other two clusters. Additionally including the factor *gaze duration* of the virtual character towards the object did not improve model fit: strong evidence against an effect of *gaze duration* was found in the eyes (BF = 0.09) and object region (BF = 0.03), moderate evidence against an effect of *gaze duration* was found in the head region (BF = 0.28). Moreover, strong evidence against an interaction effect of *cluster* and *gaze duration* was found in all three regions (eye region: BF = 0.04; object region: BF = 0.05; head region: BF = 0.10).

3.3.3 Effect of Memory on Object Recognition

Recognition rates of target words were similar for all four conditions (see table 2).

	Low pitch	High pitch
Short gaze duration	*M* = 64.2 % (*SD* = 20.0)	*M* = 65.2 % (*SD* = 21.6)
Long gaze duration	*M* = 64.7 % (*SD* = 19.0)	*M* = 63.7 % (*SD* = 20.6)

Table 2: Object recognition rates

Extreme evidence for an effect of the *participant's fixation* duration towards the object on their memory performance was found, with longer fixation of an object increasing recognition (b = 0.36; 95% CI = [–0.15, 0.86], BF > 1000). The *number of trials that had passed since object presentation* had a statistically robust effect on memory performance: The fewer trials passed since object presentation, the greater the likelihood the respective word was recognized correctly in the memory task (b = –0.18, 95% CI = [–0.28, –0.08]; BF > 1000). Additionally, strong evidence was found for an effect of *word frequency*: more frequent words tended to lead to better recognition (b = 0.06, 95% CI = [–0.10, 0.23]; BF > 100). The factors *cluster* (BF = 0.47), *pitch height* (BF = 0.16) and *gaze duration* (BF = 0.05) did not improve model fit.

3.4 Discussion

3.4.1 Rating Behavior

The importance of the object to the virtual character was rated higher when gaze duration was long (as opposed to short) or the pitch accent was high (as opposed to low). Study 1b thus replicates the main finding of web-based study 1a in a controlled laboratory setting.

In study 1a, mean ratings for the "mixed" conditions were similar. In study 1b, however, higher mean ratings for the "mixed" condition characterized by high pitch and short gaze in comparison to the other "mixed" condition characterized by low pitch and long gaze were found. In accordance with this, the effect of pitch height on participants' ratings was slightly bigger than the one of gaze duration. This suggests that the manipulation of the pitch accent had a greater impact on participants' importance ratings than the manipulation of the virtual character's gaze duration. One explanation for this may be the different experimental setting. As study 1a was a web-based study, confounding factors may have affected the outcome, such as different screen sizes and different levels of loudness. The laboratory setting in study 1b allowed for controlled visual and acoustic conditions. Further controlled studies are necessary to support the notion of a greater effect of pitch height as opposed to gaze duration. In

accordance with the current findings, studies investigating the perception of prosodic prominence, i.e. of prosodically highlighted elements in speech, have reported greater effects of pitch accents as opposed to eye-brow movements (Krahmer et al., 2002b; Mixdorff et al., 2013) and other facial movements (Swerts & Krahmer, 2008). However, these studies did not systematically examine different cue intensities. A study investigating effects on memory (Morett & Fraundorf, 2019) demonstrated that the effect of pitch accents on memory can be eradicated if salient visual cues – in this case beat gestures – are presented alongside the auditory stimulus to highlight certain speech elements. Weighting of cue effects may therefore be better understood if different alterations of the current paradigm were tested, e.g. by varying the strength of the manipulation for each cue separately or by manipulating only one factor.

Comparable to study 1a, higher word frequency increased importance ratings, which seems to contradict the notion that infrequent words elicit higher prominence perception (Cole et al., 2010; Roy et al., 2017). However, as stated in study 1a and 1c, word frequency in the stimulus set is possibly confounded with other object properties such as everyday-life importance or general object importance: The five most frequent words were the German words for “car”, “airplane”, “window”, “sun” and “church” (i.e. “Auto”, “Flugzeug”, “Fenster”, “Sonne” and “Kirche”). The five least frequent words were the German words for “spinning wheel”, “doorknob”, “spinning top”, “chisel” and “roller skate” (i.e. “Spinnrad”, “Türgriff”, “Kreisel”, “Meißel” and “Rollschuh”). The five objects receiving the highest importance ratings were “bicycle”, “traffic light”, “coat”, “bird” and “sun” (i.e. “Fahrrad”, “Ampel”, “Mantel”, “Vogel” and “Sonne”). The five objects with the lowest importance ratings were “zebra”, “pipe”, “hammer”, “ladder” and “caterpillar” (i.e. “Zebra”, “Pfeife”, “Hammer”, “Leiter” and “Raupe”). See appendix 9.1.2 for a complete list of object names used in studies 1b and 1c and their English translations.

3.4.1.1 Individuality

The three behaviorally different clusters “Listeners”, “Lookers” and “Neithers” identified in study 1a were replicated in study 1b. People perceiving prosodic prominence differ with regard to their use of visual facial information and prosodic information (Swerts & Krahmer, 2008), as well as different prosodic and non-prosodic factors such as word frequency (Baumann & Winter, 2018; Roy et al., 2017). As was the case in study 1a, no interaction effect and no additive effect of the factors pitch height and gaze duration was observed in study 1b suggesting that individuals did not integrate auditory information from speech and

visual information from the virtual character's gaze but rather made their judgements based on unimodal cues.

Considering that cognitive load can hinder audio-visual integration (Ren et al., 2023) and impair performance in tasks that require simultaneously keeping track of auditory and visual information (Fougnie et al., 2018), one possible reason for a lack of interactional and additive effects in the rating task of the current paradigm is that cognitive load was quite high. In experimental settings that require participants to merely report auditory perception, audio-visual integration can occur (Krahmer et al., 2002a) and happen automatically (Dohen & Lœvenbruck, 2009; McGurk & Macdonald, 1976; Thézé et al., 2020). However, if the task is more demanding but highly structured, additional audio- or visual information has been shown to only be used if necessary (Macdonald & Tatler, 2013). Likewise, in the current rating paradigm, auditory and visual information may be used strategically in isolation by most participants to solve the task as efficiently as possible, resulting in a lack of additive and interactional effects. The behavior of some participants that noticed the variation in both the visual and the auditory cue but made a conscious decision to focus exclusively on one of these cues, may serve to support this notion. Since one cue is sufficient to perform the task, participants where likely not motivated to actively pay attention to both channels.

Another participant, despite having noticed the difference of the virtual character's gaze duration towards the object, reported deliberately not having taken it into account for their ratings, because it did not affect their perception of object importance for the virtual character. While this may be a perception specific to that participant, it is also possible that the use of a virtual character who moved only the eyes when an utterance was presented auditorily hampered audio-visual integration. The degree to which a virtual character is perceived as a sentient character with thoughts and agency can influence the tendency to mentalize: not only gaze direction can influence an observer's interpretation, but also what the gazer is allegedly perceiving themselves (Mayrand et al., 2024) and whether the virtual character is perceived as an agent or not (Terrizzi & Beier, 2016).

A different explanation may be that the perceptual link between pitch height and gaze duration is not as strong as the perceptual link between other actions of speech and their visual by-products. As argued by Prieto and colleagues (Prieto et al., 2015), in cases in which articulatory gestures are unavoidably part of production (such as during the production of syllables in which the movements of the lips are unavoidable), audio-visual integration during perception may be more likely to occur (McGurk & Macdonald, 1976; Thézé et al., 2020). On

the other hand, the production of pitch accents is not necessarily accompanied by articulatory gestures. However, while articulation required to produce prosodic prominence does not unavoidably elicit a visual by-product, pitch accents are nevertheless often accompanied by perceivably greater jaw and lip opening (Cvejic et al., 2010; Dohen et al., 2006; Scarborough et al., 2009) as well as by movements that are not directly linked to pitch accent production such as eyebrow or head movement (Ambrazaitis & House, 2017; Cvejic et al., 2010). The weaker perceptual link may lead to less automatic integration of these visual by-products with the auditory information. But they may equip the listener/observer with an additional source of information they may use if necessary.

The feedback of participants in the "Neithers" group suggests that some did, in fact, notice variation in the virtual character's gaze duration towards the object and in the tone of voice of the presented utterance. The reason this did not show in their ratings may be that they weighed other cues more strongly, such as object properties and characteristics of the virtual character as well as participants' own experience with the object. Some participants took into account the virtual character's gaze behavior as well as the voice stimuli but focused on information other than gaze duration towards the object or intonation – such as mutual gaze and blinking. With regard to the auditory stimulus, participants may have focused on information other than pitch height on the accented syllable: Because the auditory stimuli were derived from natural speech, they introduced a source of variance. Although voice-associated noise was kept to a minimum, stimuli naturally differed with regard to the length of the utterance itself as well as to the accented syllable and other prosodic parameters that influence the perception of prominence or preference.

3.4.1.2 Exploratory Analysis

The exploratory analyses in part corroborate participants' behavioral patterns: The more visually sensitive a participant, (i) the more they made use of the gaze cue, and (ii) the less they made use of the pitch cue. Importantly, participants filled in the information on sensory perception before entering the rating task. It remains unclear, if participants were "Listeners", "Lookers" and "Neithers" before entering the experiment. However, these relationships support the idea that participants brought a prior disposition to the laboratory instead of developing a "Listeners"-, "Lookers"- or "Neithers"-behavior throughout the experiment. This notion of default or preferred individual decoding types is supported by a study investigating perception of pitch accents in speech: Participants' cue preference was shown to coincide with their general perceptual abilities, i.e. participants with worse pitch perception abilities prioritized pitch less compared to other cues such as duration (Jasmin et al., 2019).

In the current study, the pitch cue was used less by participants with lower empathic abilities and tended to be used less by participants with higher autistic traits. Additionally, the gaze cue tended to be used less by participants with higher systemizing needs. Since participants on the autism spectrum tend to report lower empathic abilities, higher autistic traits and greater systemizing tendencies (Baron-Cohen et al., 2001, 2003; Baron-Cohen & Wheelwright, 2004), this pattern could indicate different cue use behavior in an autistic population.

3.4.2 Fixation Durations

Cluster membership was linked to fixation durations within the eyes region and the object region. In particular, compared to the cluster of "Listeners", the cluster of "Lookers" spent more time fixating the eye region, less time fixating the object region and a similar amount of time fixating the head region. Participants' individual behavior in the rating task was thus mirrored by their gaze behavior. These results along with qualitative feedback participants made after the experiment support the idea that participants were actively monitoring their preferred input modality.

The virtual character's eye gaze duration towards the object did not affect fixation durations in any of the three regions of interest, i.e. the eyes region, the object region and the head region (excluding the eyes region). Studies investigating joint attention have demonstrated that participants are more likely to look at an object that is being looked at by another person (Ricciardelli et al., 2002) and a person's eye gaze directed towards an object tends to increase the observers' fixation duration towards the respective object (Adil et al., 2018; Castelhano et al., 2007; Hutton & Nolte, 2011; Theuring et al., 2007). Moreover, the longer a person is observed looking at an object, the longer the observer's gaze duration towards the respective object (Freeth et al., 2010). However, these studies did not require participants to make responses regarding the gazing person's mental states, which was the main focus of the current study. The instructions of the current paradigm were not primarily created to direct attention towards the objects presented alongside the virtual character and are likely not optimally suited to do so. However, in a more natural setting with multiple cues that do not allow for such a strategic approach as subjects in the current paradigm could come up with, gaze duration may have a greater impact.

Taken together, the analysis of the eye-tracking data highlights the importance of top-down attentional processes. Other studies have shown that participants' fixation durations can be manipulated by instilling different strategies via task instructions (Buswell, 1935; Klami, 2010; Klami et al., 2008). In the current paradigm, task instructions were open with regard to which

rating strategy was to be applied. Therefore, participants had to come up with their own strategy and adjust their gaze behavior accordingly.

3.4.3 Effect of Memory on Object Recognition

Neither pitch height of the utterances referring to the object nor gaze duration of the virtual character's gaze towards the object had an influence on memory as assessed by object recognition. Previous studies have reported that attention directed towards an object guided by prosodic prominence or gaze during encoding can have an impact on object memory (Adil et al., 2018; Fraundorf et al., 2010, 2012; Gregory & Jackson, 2017; Gregory & Kessler, 2022; Kember et al., 2021; Kushch et al., 2018; Morett & Fraundorf, 2019; Sajjacholapunt & Ball, 2014; Wahl et al., 2019). None of these studies, however, is directly comparable to the current study, amongst other reasons because manipulation of prosodic prominence and gaze in these studies is not specifically and exclusively operationalized as pitch excursion or gaze duration.

One explanation for a lack of an effect of pitch height and gaze duration on object recognition in this study could be the instructions for the rating task, which required top-down control of attention: Participants were asked to attribute mental states to the virtual character, thus withdrawing attention from the target object. Interfering processes such as higher task demands have been shown to eradicate or attenuate gaze-cueing effects (Bobak & Langton, 2015; Chen et al., 2021). Moreover, in contrast to the current study, participants in some of the studies mentioned above (Fraundorf et al., 2010, 2012; Kember et al., 2021; Kushch et al., 2018) were aware of the later memory test and could thus successfully direct attentional resources towards memorizing target words.

Another explanation for the lack of an effect on object recognition may be the time lag after stimulus presentation. Effects that may have ensued from pitch height or gaze duration in the first task may have been too transient to transpire into the subsequent recognition task. Studies investigating the influence of gaze-cueing on item recognition (Gregory & Jackson, 2017; Gregory & Kessler, 2022) and item recall (Dodd et al., 2012) showed positive effects shortly after cue presentation: 1.0 s and approximately 1 min, respectively. Although long-term (3 min) gaze-cueing effects have been reported by Frischen and Tipper (Frischen & Tipper, 2006), in their study these arose only under certain conditions which do not apply to the current study, namely when the face was that of a famous person and when gaze cues were directed towards the left side as opposed to towards the right side. In the current study, presentation of items alone took at least 10.2 min (time spent making the ratings not included). Possible effects of gaze duration on memory may not have survived this long.

Moreover, analyzing response correctness in the memory task may not be an optimal task for detecting small cueing effects on object recognition as this measure may be too coarse. Instead of assessing response correctness within a relatively long time window (up to five seconds), effects may better be detected by for example assessing response times in a recognition task in which an old target item is simultaneously presented alongside a new item or a lexical decision task in which participants have to identify words and non-words as quickly as possible. More accessible words are usually identified faster in these tasks. By analyzing these response time differences, subtle effects may be better detectable.

The impact of other factors relevant to object recognition previously reported in the literature were however also found in the current study, thus affirming general object recognition task validity. The more time participants spent fixating an object, the better they could later recognize that object as already seen before. Participants could also better recognize objects that had more previously been presented in comparison to objects presented earlier during the rating task. These finding are in line with previous studies which have shown that the longer participants look at an object, scene or face, the better they can later remember it (Droll & Eckstein, 2009; Martini & Maljkovic, 2009; Melcher, 2001, 2006) and with the literature on serial position effects in recognition tasks reporting that participants can better recognize objects they have seen more recently (Brady et al., 2008; Konkle et al., 2010).

In the current study, higher word frequency was linked to better object recognition. Early studies investigating word recognition have shown that subjects are better at recognizing infrequent words as opposed to high-frequent words (Glanzer & Adams, 1985; Schulman, 1967; Shepard, 1967). However, this effect is susceptible to several factors such as delay until test phase, serial position, number of items, response time limitation (Joordens & Hockley, 2000), task instructions (Criss & Shiffrin, 2004; Hirshman & Arndt, 1997) and item presentation duration (Criss & Shiffrin, 2004; Malmberg & Nelson, 2003). The absence of a positive effect of infrequency on object recognition may thus be explained by the design of the rating paradigm. It was not aimed at solely investigating object recognition, but at probing mentalizing processes. The manipulation introduces different kinds of influence on memory performance that could easily hamper the impact of a positive effect of infrequent words. For example, the presence of other auditory and visual stimuli as well as the instruction to rate the importance of the object for the virtual character draw on attentional resources that may otherwise be available for word frequency processing. Additionally, the presence of a negative effect of infrequency on word recognition in the current study may be explained by word

frequency being confounded with other object properties in the set of selected words (Zimmermann et al., 2020) such as general object importance. As mentioned above, this may have led to a high-frequent word also being considered important and thus potentially remembered better.

3.5 Further Limitations

Apart from the limitations discussed above, further limitations regarding the rating paradigm apply to this study: (i) The findings obtained in this reductionistic paradigm including a virtual character have limited external validity, (ii) the task instructions may have fostered the systematic search for a rating strategy and hindered audio-visual integration, (iii) because the task instructions leave room for interpretation with regard to how to solve the task, participants' ratings are affected by their individual personality, perception and behavior, thus their ratings are affected by multiple factors that introduce noise which could in future studies be investigated further.

3.6 Conclusion

As demonstrated in study 1a (Zimmermann et al., 2020), pitch excursion and gaze duration are suitable nonverbal channels to infer the mental state of a virtual character. People differ in terms of the degree to which they make use of these cues to infer the importance of an object to the virtual character. People's cue use is mirrored by their gaze behavior.

Under which circumstances object memory can be affected by these cues needs to be further investigated in future studies. These would benefit from including a rating task for general object importance and a different memory task which is optimized for the analysis of response times instead of response correctness. Furthermore, it may be interesting to investigate the factors that contribute to individual differences regarding rating behavior and eye gaze such as sensory perception or autistic traits.

4 Study 1c

Zimmermann, J. T., Ellison, T. M., Cangemi, F., Wehrle, S., Vogeley, K., & Grice, M. (2024). Lookers and Listeners on the autism spectrum: The roles of gaze duration and pitch height in inferring mental states. *Frontiers in Communication, 9,* Article 1483135. https://doi.org/10.3389/fcomm.2024.1483135

Supplementary material for study 1c can be found in appendix 9.1.

frontiers | Frontiers in Communication

TYPE Original Research
PUBLISHED 08 November 2024
DOI 10.3389/fcomm.2024.1483135

OPEN ACCESS

EDITED BY
Martina Micai,
National Institute of Health (ISS), Italy

REVIEWED BY
Helene Kreysa,
Friedrich Schiller University Jena, Germany
Catherine Caldwell-Harris,
Boston University, United States

*CORRESPONDENCE
Martine Grice
✉ martine.grice@uni-koeln.de

†PRESENT ADDRESS
Francesco Cangemi,
International Education Division, Center for Global Education, The University of Tokyo, Bunkyō, Japan

‡These authors have contributed equally to this work and share senior authorship

RECEIVED 19 August 2024
ACCEPTED 22 October 2024
PUBLISHED 08 November 2024

CITATION
Zimmermann JT, Ellison TM, Cangemi F, Wehrle S, Vogeley K and Grice M (2024) Lookers and listeners on the autism spectrum: the roles of gaze duration and pitch height in inferring mental states. *Front. Commun.* 9:1483135. doi: 10.3389/fcomm.2024.1483135

Lookers and listeners on the autism spectrum: the roles of gaze duration and pitch height in inferring mental states

Juliane T. Zimmermann[1], T. Mark Ellison[2], Francesco Cangemi[2†], Simon Wehrle[2], Kai Vogeley[1,3‡] and Martine Grice[2*‡]

[1]Department of Psychiatry, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany, [2]IfL – Phonetics, University of Cologne, Cologne, Germany, [3]Institute of Neuroscience and Medicine, Division of Cognitive Neuroscience (INM-3), Research Centre Juelich, Juelich, Germany

Although mentalizing abilities in autistic adults without intelligence deficits are similar to those of control participants in tasks relying on verbal information, they are dissimilar in tasks relying on non-verbal information. The current study aims to investigate mentalizing behavior in autism in a paradigm involving two important nonverbal means to communicate mental states: eye gaze and speech intonation. In an eye-tracking experiment, participants with ASD and a control group watched videos showing a virtual character gazing at objects while an utterance was presented auditorily. We varied the virtual character's gaze duration toward the object (600 or 1800 ms) and the height of the pitch peak on the accented syllable of the word denoting the object. Pitch height on the accented syllable was varied by 45 Hz, leading to high or low prosodic emphasis. Participants were asked to rate the importance of the given object for the virtual character. At the end of the experiment, we assessed how well participants recognized the objects they were presented with in a recognition task. Both longer gaze duration and higher pitch height increased the importance ratings of the object for the virtual character overall. Compared to the control group, ratings of the autistic group were lower for short gaze, but higher when gaze was long but pitch was low. Regardless of an ASD diagnosis, participants clustered into three behaviorally different subgroups, representing individuals whose ratings were influenced (1) predominantly by gaze duration, (2) predominantly by pitch height, or (3) by neither, accordingly labelled "Lookers," "Listeners" and "Neithers" in our study. "Lookers" spent more time fixating the virtual character's eye region than "Listeners," while both "Listeners" and "Neithers" spent more time fixating the object than "Lookers." Object recognition was independent of the virtual character's gaze duration towards the object and pitch height. It was also independent of an ASD diagnosis. Our results show that gaze duration and intonation are effectively used by autistic persons for inferring the importance of an object for a virtual character. Notably, compared to the control group, autistic participants were influenced more strongly by gaze duration than by pitch height.

KEYWORDS

autism spectrum disorder, theory of mind, pitch height, eye gaze duration, intonation, perception and gaze behavior, mentalizing, adult

1 Introduction

Autism spectrum disorder (ASD) is characterized by difficulties in social communication and interaction (American Psychiatric Association, 2013). These difficulties might in part be explained by impaired perspective-taking or mentalizing skills (Baron-Cohen et al., 1985; Frith et al., 1991; Frith and Frith, 2006). However, adults with autism without intelligence deficits perform similarly to control participants in mentalizing tasks – inferring mental states of others – that strongly rely on verbal abilities (Bowler, 1992; Happé, 1994; Scheeren et al., 2013; Gernsbacher and Yergeau, 2019), whereas they show difficulties in non-verbal mentalizing tasks (*cf.* Baron-Cohen et al., 2001a; Baron-Cohen et al., 2001b; Ponnet et al., 2004; Dziobek et al., 2006; White et al., 2011), for example, when inferring mental states of people depicted in videos of social interactions (Ponnet et al., 2004; Dziobek et al., 2006). Accordingly, autistic adults tend to rely on verbal information (e.g., the words spoken) more than on non-verbal information (e.g., the body language accompanying the words and the way they are spoken) (Kuzmanovic et al., 2011; Stewart et al., 2013). However, the interplay between non-verbal modalities has not been studied systematically in this context. For the current study, we will focus on the interplay of two powerful means to communicate nonverbally in face-to-face interactions: eye gaze and intonation.

In human communication as well as in the communication between humans and virtual characters, eye gaze can be very informative, as it is closely linked to attention: people tend to look at objects (Buswell, 1935; Yarbus, 1967; DeAngelus and Pelz, 2009) or locations they attend to (Ferreira et al., 2008; Theeuwes et al., 2009). The relevance (Klami et al., 2008; Klami, 2010) of and the preference for an object (Shimojo et al., 2003; Chuk et al., 2016) is indicated by the time one spends looking at the object. This implies that another person's gaze behavior is key to inferring their intentions or attentional state (Baron-Cohen et al., 1995; Lee et al., 1998; Freire et al., 2004; Einav and Hood, 2006; Jording et al., 2019a, 2019b). Observing another person gazing towards an object in their environment can re-direct the observer's attention and increase the duration the observer spends looking at the respective object themselves (Freeth et al., 2010). However, adults with autism tend to have difficulties inferring emotions and mental states based on another person's eye region (Hobson et al., 1988; Baron-Cohen et al., 1997; Baron-Cohen et al., 2001a). They look at gaze-indicated objects less often (Wang et al., 2015) and tend to spend less time fixating those objects (Fletcher-Watson et al., 2009; Freeth et al., 2010). One explanation for this could be reduced attention towards gaze cues in individuals on the autism spectrum (Itier et al., 2007). Certainly, overt attention towards social stimuli in general is reduced in persons with autism (Chita-Tegmark, 2016), who tend to fixate the eye region for a shorter amount of time than control participants (Setien-Ramos et al., 2022), while differences for other parts of the face are less pronounced (Klin et al., 2002; Pelphrey et al., 2002; Dalton et al., 2005; Nakano et al., 2010; Auyeung et al., 2015). Irrespective of an ASD diagnosis, the time spent fixating a person's eyes is linked to the observer's mentalizing abilities (Müller et al., 2016). In autism, a decreased fixation duration on the eye region is associated with impaired social functioning and increased autism symptom severity (Riddiford et al., 2022). However, attention towards social stimuli in autism is dependent on the stimulus at hand (Guillon et al., 2014; Chita-Tegmark, 2016), and eye gaze behavior is influenced by the experimental task and task instructions (Del Bianco et al., 2018; Setien-Ramos et al., 2022). In classical false-belief tasks, which test the ability of an observer to understand that other people can believe things which the observer knows to be untrue (most famously the "Sally-Anne" test), eye gaze behavior can indicate impaired mentalizing in autism (Senju et al., 2009; Schneider et al., 2013; Schuwerk et al., 2015). By including eye-tracking in our study, we aimed to investigate the influence of an ASD diagnosis in combination with behavioral differences on participants' gaze behavior.

Prosody—referring to the non-verbal aspects of speech—is an important aspect of spoken language, as it adds an additional layer of information to the verbal content of an utterance, and can significantly change the meaning, and consequently the interpretation, of what is being said. This is important, for example, when deciphering emotions. Most prosodically expressed basic emotions, such as fear or sadness, can be recognized well by persons with autism (O'Connor, 2012; Stewart et al., 2013; Ben-David et al., 2020; Zhang et al., 2022). However, the identification of prosodically expressed emotions that are complex, such as curiosity or concern (Kleinman et al., 2001; Rutherford et al., 2002; Golan et al., 2007; Hesling et al., 2010; Rosenblau et al., 2017), or low-intensity (Globerson et al., 2015) has been reported to be impaired in autistic adults, possibly due to difficulties with the perception and interpretation of vocal pitch modulation (how the speech melody is changed) during speech (Schelinski et al., 2017; Schelinski and von Kriegstein, 2019; see Grice et al., 2023 for a review). Moreover, the imitation of vocal pitch can also be impaired in autistic adults (Wang et al., 2021).

Aspects of conversation that are important, new, or in focus are often highlighted prosodically by the speaker. In German, this can be achieved through pitch accent placement and type, cued *inter alia* by fundamental frequency, which is perceived as pitch height (Grice and Baumann, 2007; Féry and Kügler, 2008). The raising of pitch conveys prosodic prominence and importance for the listener (Arnold et al., 2013; Baumann and Winter, 2018). Autistic listeners have been reported to take pitch accents into account to a lesser extent than control persons when judging the givenness of a word, i.e., judging whether the object it denotes is known to the interlocutors in a given context or has not been previously introduced (Grice et al., 2016). Findings from the general population show that an attenuated sensitivity to pitch accent types is associated with poor pragmatic skills, i.e., the appropriate use of language in social situations (Bishop, 2016; Hurley and Bishop, 2016; Bishop et al., 2020).

Analogously to gaze, prosody (and pitch accents in particular) can function as a deictic cue (referring or "pointing" to an entity) and orient a listener's attention (Dahan et al., 2002; Weber et al., 2006; Ito and Speer, 2008; Watson et al., 2008). Studying overt attention in children with autism, Ito et al. (2022) found that, although the autistic group responded relatively slowly and weakly to a target word denoting an object, both the control group and the autistic group looked at the respective object longer if the referring utterance received an emphatic pitch accent (i.e., it was produced with longer duration and higher pitch). This demonstrates that autistic children can shift overt attention towards an important object in their environment. No comparable study has been performed with autistic adults to date.

In a previous web-based study (Zimmermann et al., 2020), we showed that both gaze duration and pitch height are used as cues by the general population when interpreting how a virtual character

conveys the importance of an object being referred to. In that study, participants rated objects as having greater importance for the virtual character both when the character looked at the object for a longer period of time (as opposed to a shorter period of time) as well as when she produced the word referring to it with higher vocal pitch (as opposed to lower pitch). Based on the tendency of participants to take into account only one of the two cues, we subdivided the sample into three behavioral clusters: (i) "Lookers," who based their ratings primarily on gaze duration, (ii) "Listeners," who based their ratings primarily on pitch height, and (iii) a group of "Neithers," who did not predominantly base their ratings on either cue.

Continuing this line of work on the influence of gaze duration and pitch height, the present study is a lab-based experiment investigating not only participants' responses but also their eye gaze fixation durations using a desk-mounted eye-tracker. We carried out a comparative analysis of participants with and without a diagnosis of ASD. In particular, we investigated whether similar behavioral patterns can be found in both groups. We hypothesized that the autistic group would rely on the gaze and pitch cues to a lesser extent, based on reports of difficulties in autism with using social gaze and intonation as cues for mentalizing (as summarized above). We also expected this to be reflected in the participants' own gaze behavior. Additionally, we examined how participants' gaze behavior, the character's gaze and pitch cues, as well as the presence of an ASD diagnosis affected performance in a memory task involving recognition of the objects used as visual stimuli (i.e., participants had to indicate whether an object had been or had not been present in the previous part of the experiment).

2 Materials and methods

2.1 Participants

For the autistic group, we recruited 24 monolingual German native speakers within an age range from 18 to 55 who had been diagnosed with Asperger syndrome (ICD-10 identifier: F.84.5) or with childhood autism (ICD-10 identifier: F.84.0) by the outpatient clinic for autism in adulthood or by the pediatric outpatient clinic for autism of the University Hospital Cologne. For the control group, we recruited 24 age-matched (within a range of 5 years) native German speakers. All participants of both groups had normal or corrected-to-normal vision as well as hearing. The study was conducted in accordance with the Declaration of Helsinki (World Medical Association, 2013) and approved by the Ethics Committee of the University Hospital Cologne.

To ensure that results were not influenced by lower cognitive performance, we only included participants with verbal and total intelligence scores of at least 85, as measured with the *WIE-III*, (Aster et al., 2006), with attentional scores greater 80, as measured with the *D2* (Brickenkamp, 2002), and for participants in the control group with maximally moderate depressive symptoms as measured with the *Beck Depression Inventory* (*BDI-II*, Beck et al., 1996), i.e., with *BDI-II* scores <18. Sample characteristics are provided in Table 1.

Verbal intelligence scores as measured with the *WIE* indicated average or above-average verbal intelligence for all participants (Table 1). Diagnostic groups did not differ significantly regarding the *WIE* verbal scores [two-samples *t*-test, $t(46) = -1.50$, $p = 0.140$] or the *WIE* performance scores [two-samples *t*-test, $t(46) = -1.73$, $p = 0.091$]. Scores indicating depression or depressive tendencies as measured with the *BDI* were significantly higher in participants with autism compared to the control group [Welch two-samples *t*-test, $t(32.03) = -4.25$, $p < 0.001$]. Attention scores measured with the *D2* tended to be somewhat higher in the autistic sample [two-samples *t*-test, $t(46) = -1.88$ $p = 0.066$].

2.2 Experimental design

We used a paradigm established in the previous web-based study referred to above (Zimmermann et al., 2020). The material and procedures were adjusted for the laboratory setting.

We tested the individual and combined influence of *gaze duration* of a virtual character towards an object and *pitch height* of an utterance on the rating of how important the object appeared to the virtual character. In addition, we obtained object memory scores by assessing recognition rates for all objects in a subsequent recognition task. To create a socially "plausible" and at the same time standardized situation, we presented videos of a virtual character's face positioned above an object. The object was different in each trial, and each object was only shown once. The movements performed by the virtual character were limited to the eyes. The character's attention towards the object, suggesting greater importance, was operationalized as longer *gaze duration* directed towards the object alongside an auditorily presented utterance characterized by a pitch accent with a fundamental frequency peak located on the stressed syllable of the target word. We systematically varied the factors *gaze duration* and *pitch height* on two levels. *Gaze duration* towards the object was either comparatively short (600 ms) or long (1800 ms). *Pitch height* on the accented syllable was either low or high, characterized by f0 peak

TABLE 1 Sample characteristics, general.

	Sex	Age	*WIE* IQ verbal	*WIE* IQ performance	*WIE* IQ total	*D2* total error corrected	*BDI-II*
ASD ($N = 24$)	13 men 10 women 1 not indicated	18–55 years $M = 39.4$ ($SD = 11.7$)	$M = 115.9$ ($SD = 14.1$)	$M = 110.7$ ($SD = 16.2$)	$M = 114.9$ ($SD = 14.6$)	$M = 105.6$ ($SD = 9.7$)	$M = 15.5$ ($SD = 10.8$)
Control ($N = 24$)	14 men 10 women	21–58 years $M = 38.9$ ($SD = 11.9$)	$M = 110.1$ ($SD = 12.5$)	$M = 103.2$ ($SD = 13.7$)	$M = 107.4$ ($SD = 12.1$)	$M = 100.5$ ($SD = 8.8$)	$M = 5.3$ ($SD = 4.9$)

WIE IQ, Hamburg-Wechsler-Intelligenz-Test für Erwachsene III (intelligence test for adults); D2, d2 Aufmerksamkeits-Belastungs-Test (attention load test); BDI-II, Beck Depression Inventory, 2nd Version (questionnaire on depressive symptom severity).

10.3389/fcomm.2024.1483135

height, which was raised by 45 Hz in the respective high-pitch-height condition. Thus, we effectively created four conditions, establishing a 2 × 2 experimental design (Table 2).

2.3 Video material

Videos were created by combining images and sound material using *Python* and the *FFmpeg* module (FFmpeg Developers, 2018). The videos used in the rating task showed a female character's face positioned above the center of the screen (screen dimensions: 1,920 × 1,200 px). The face and its position were always the same during the entire experiment. One object was presented below the center of the screen (see Figure 1 for image positions and the time course of a single trial). The background color was white. At the beginning and the end of the video, the virtual character exhibited idle gaze behavior, i.e., she performed gaze movements in the direction of random locations in the environment, but neither fixated the object nor the participant during this phase. All images of the virtual character's face were taken from previous studies investigating the perception of gaze direction (Eckert, 2017; Jording et al., 2019a). The virtual character's face was created using Poser R (Poser 8, Smith Micro Software, Inc., Columbia, USA) using Python 2.4. For the idle gaze phases, we chose eight images of gaze directions that were diverted horizontally to the left or to the right as well as diverted slightly to the bottom. The choice of the female character was based on the decision to use a female speaker after pretesting for production of the auditory stimuli.

After 2.0 s of idle gaze, the virtual character made three fixations establishing a social situation: (1) looking at the participant for 1.0 s, (2) looking towards the object for either 0.6 (short gaze) or 1.8 s (long gaze), and (3) looking at the participant again for 1.0 s. The onset of the virtual character looking towards the object was also the onset of the auditory utterance. The durations of 0.6 and 1.8 s of the virtual character's gaze toward the object were chosen based on a previous study of human–robot interaction (Pfeiffer-Lessmann et al., 2012), where the durations of 0.6 s and 1.8 s were associated with different perceptions of the robot's intention to make the participant follow their gaze. Importantly, in that study, 1.8 s was the participants' own preferred gaze duration towards an object with the intention of making the robot follow their gaze.

This set of three gazing actions (looking at the participant, looking towards the object, and looking at the participant again) conveying communicative intent was both preceded and followed by a blink simulated by presenting an image of the virtual character's face with their eyes closed for 0.1 s to simulate naturalistic interblink-interval durations (Doughty, 2001). However, to make the character's blinking behavior appear less mechanical, the videos were created by randomly

TABLE 2 2 × 2 experimental design.

		Gaze duration	
		Short	Long
Pitch height	Low	Low pitch height and short gaze	Low pitch height and long gaze
	High	High pitch height and short gaze	High pitch height and long gaze

Variation of gaze duration and pitch height resulted in four conditions.

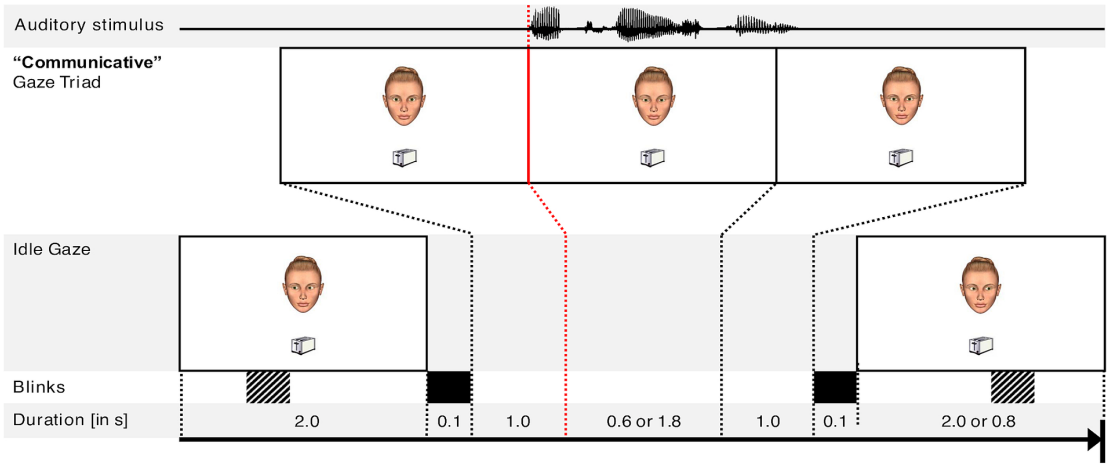

FIGURE 1
Schematic time course of a video depicting the different phases of an example trial. The object in this example is a toaster, the auditory stimulus is the utterance "der Toaster" (English: the toaster). Total duration of each video was 6.6 s. Blinks are demarcated as either fixed (solid black) or appearing at random (striped). The red line indicates the simultaneous onset of the auditory stimulus and the virtual character's gaze towards the object.

including either no blink or only one additional blink during the first and second idle (i.e., "non-communicative") phases. Following the "communicative" gaze triad, the virtual character continued gazing at random locations until the end of the video, i.e., for 2.0 s (short gaze conditions) or for 0.8 s (long gaze conditions) in order to keep the total presentation duration of the object constant in all videos.

On the basis of our experience with the web-based study (Zimmermann et al., 2020), we excluded 14 problematic items from the previous stimulus set. These exclusions resulted in a final set of 92 test items, with each participant observing 23 items per condition. Four of the discarded stimuli were used for practice trials in the current study, but did not enter analysis.

2.4 Object images

Object images used for video creation were selected from the set described in Rossion and Pourtois (2004). Images were selected based on the phonology of their referential German expressions (Genzel et al., 1995). To reduce any possible influence of the number of syllables on the perception of word prominence, only words with two syllables and initial stress were included in the subset, such as "Toaster," "Hammer," "Meißel," "Sofa" (respectively *toaster, hammer, chisel, sofa*). The full list of object names and their English translations can be found in the Supplementary material. Additionally, we partly excluded homonyms if the homonym-partner was present in the image set or if one homonym-partner was semantically clearly more salient (e.g., the German homonym "Mutter" is semantically more salient when referring to "mother" than to "nut" as the counterpart of a screw).

2.5 Auditory stimulus material

The auditory stimuli were produced by a trained female speaker, who uttered each of the 92 target phrases including the definite article (e.g., "der Toaster": *the toaster*) with an H*-accented rendition [following the categorization of German accent types by Grice et al., 2005]. The H* accent type has been found to be generic, and can be used for different focus types in German, namely broad focus, narrow focus and contrastive focus (Grice et al., 2017). Recordings took place in a soundproof booth, using an *AKG C420L* headset microphone connected to a computer running *Adobe Audition* via a USB audio interface (PreSonus AudioBox 22VSL). Stimuli were recorded with a sampling rate of 44,100 Hz, 16 bit. The resulting speech stimuli were normalized to equal loudness using *Myriad* (Aurchitect Audio Software, LLC, 2018). The editing was performed using *Praat* (Boersma and Weenink, 2018). Sound was faded in and out (Winn, 2014) to avoid any salient on- and offset of noise. Fundamental frequency (f0) contours were extracted, manually corrected, and smoothed according to an established procedure (Cangemi, 2015). The resulting pitch contours were stylized to a resolution of one semitone. These stylized versions were used directly as the audio stimuli for the low-pitch-height condition. The pitch contours of utterances for the high-pitch-height condition were resynthesized: *pitch height* maxima on the accented vowels were raised by 45 Hz. This difference between *pitch height* maxima for the different conditions was based on the individual production characteristics of the speaker for a subset of 15 words. These words were selected due to their ability to bear pitch (i.e., the amount of periodic energy, typically high in vowels and low in, e.g., fricatives and stops such as /f/ and /d/ respectively).

The speaker was asked to produce all utterances in two versions: (1) applying an H*-accent and (2) applying an L+H*-accent, the latter of the two resulting in a perceptually more strongly accented utterance which expresses greater prominence. We extracted the following measures for characterization of speaker-specific production parameters for utterances bearing an H*-accent and those bearing an L+H*-accent: onset and duration of the accented syllable, height of f0 contour peak on the accented vowel as well as the associated timepoint, height of f0 contour trough within the timeframe starting at voice onset and ending at f0 peak on the accented vowel as well as the associated timepoint. Timepoints for the onset of the accented syllable were set manually; for f0 peaks and troughs, they were set automatically and corrected manually. Subsequently, f0 peak alignment was calculated as the percentage of duration from accented syllable onset until the timepoint of the f0 contour peak in relation to the total duration of the accented syllable. F0 trough alignment was calculated analogously to f0 peak alignment. F0 onglide was calculated as the difference between f0 trough height and f0 peak height (Figure 2). We plotted the following parameters to examine to what degree they contributed to distinguishing between utterances bearing an H*-accent and those bearing an L+H*-accent in our speaker: (a) f0 peak and f0 peak alignment (Figure 3A), (b) f0 trough and f0 trough alignment (Figure 3B), (c) f0 onglide and the duration of the accented syllable (Figure 3C).

Visual inspection of production parameters showed that *pitch height* most reliably separated the two stimulus conditions for the selected speaker (Figure 3). *Pitch height* was on average 205.4 Hz ($SD = 0.65$) for the utterances produced with an H*-accent, and 251.1 Hz ($SD = 1.08$) for the utterances produced with an L+H*-accent. To mirror this difference, a positive adjustment of 45 Hz was chosen to simulate an otherwise comparably accented L+H*-like version of our stylized H*-accented utterances.

The resulting auditory stimuli were submitted to a perceptual pretest: The original H*-accented utterances and their stylized versions were rated by six trained phoneticians for "similarity." All stylized stimuli were rated for "naturalness" and accent type. Details on the pretest's methods and results as well as auditory and video stimuli can be found in the Supplementary material.

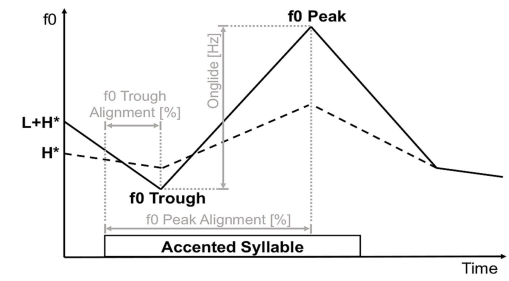

FIGURE 2
Schematic illustration of acoustic measures of an utterance. Extracted measures were onset and offset of the accented syllable, f0 contour peak on the accented vowel and f0 contour trough. Calculated measures are depicted in gray.

10.3389/fcomm.2024.1483135

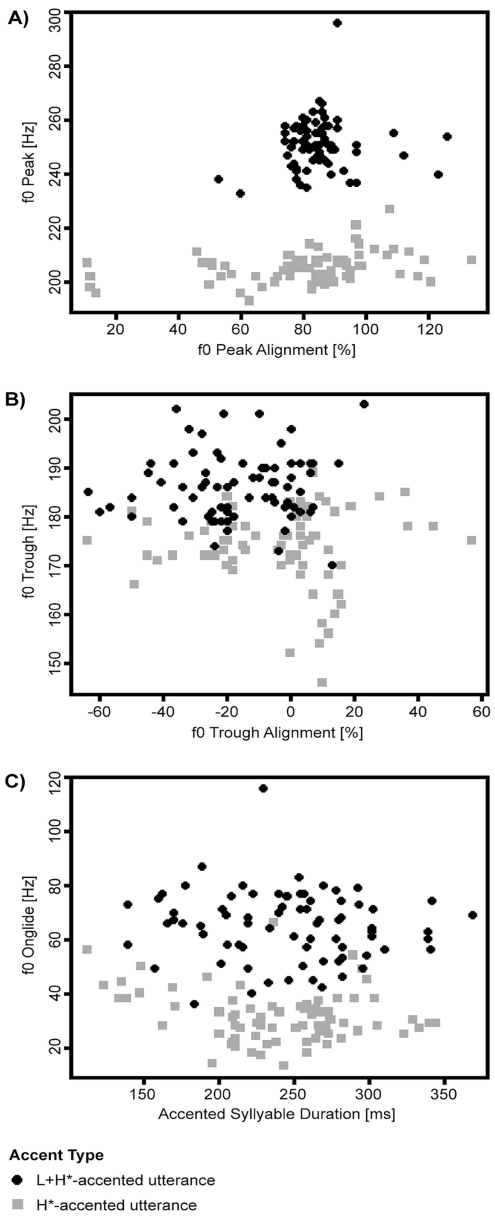

FIGURE 3
Acoustic parameters of the speaker's utterances: **(A)** F0 peak and f0 peak alignment, **(B)** f0 trough and f0 trough alignment, **(C)** f0 onglide and the duration of the accented syllable. Gray squares and black dots distinguish between the utterance produced with lesser (H*) and greater (L + H*) prominence.

2.6 Selection of distractor words for the recognition task

The 92 distractor words presented alongside the 92 target words in the recognition task were selected by identifying words of similar word frequency compared to the words we used in the rating task (Brysbaert et al., 2011). Since animacy has been reported to lead to better recognition (Leding, 2020), we included an equal number of animals in the list of distractor words and target words.

2.7 Psychological tests

To infer mentalizing abilities, we employed the *"Reading the Mind in the Eyes" test* (Baron-Cohen et al., 2001a; Baron-Cohen et al., 2001b), henceforth referred to as *Eyes-test*. For a proxy of sensory perception we included a German translation of the *Sensory Perception Quotient* (*SPQ*, Bierlich et al., 2024). As indicators for autistic traits, we included the *Autism Quotient* (*AQ*, Baron-Cohen et al., 2001b), the *Empathy Quotient* (*EQ*, Baron-Cohen and Wheelwright, 2004) and the *Systemizing Quotient* (*SQ*, Baron-Cohen et al., 2003).

2.8 Procedure

The study was conducted at the Department for Psychiatry of the University Hospital of Cologne. Participants provided informed consent and filled in the *AQ, EQ, SQ, SPQ* and a questionnaire on demographic data as well as information on (their history of) visual, auditory, psychological or speech impairments. Afterwards, they filled in the *BDI-II* and were tested with the *Eyes-test* and the *D2* (as described above). For the duration of the rating task, participants were seated in front of a desk-mounted eye-tracker. Head movement was minimized with the use of a fixed chin rest. They were instructed to imagine that the utterances they heard were produced by the character on screen and were informed that the character could convey the importance of the object. Participants were then instructed to answer the same question after each trial: "How important does the character find the depicted object?" (original German instruction: "Wie wichtig findet die Figur das abgebildete Objekt?"). Each trial of the rating task consisted of a video and its subsequent rating. To ensure that each of the 92 videos was viewed by the same number of participants, they were randomly assigned to one of four experimental groups. Items were presented in randomized order. Before and after the video presentation, a fixation cross was presented in the center of the screen for a random duration in the range 500–1,000 ms. Each video sequence was followed by a screen asking for ratings on a scale from 1 to 4 (through keyboard presses): 1 = "not important at all" (German: "unwichtig"); 2 = "rather unimportant" (German: "eher unwichtig"); 3 = "rather important" (German: "eher wichtig"); 4 = "very important" ("sehr wichtig"). Four items were used as practice trials.

The rating study was followed by a recognition task. Here the words from the rating task and the same number of distractor words were presented on screen alongside their respective definite articles in the nominative case. Participants were instructed to indicate whether the respective object had been presented during the rating task or not (through keyboard presses). Thus, this task was designed to test whether they recognized the objects used in the rating task. After the recognition task, participants filled in a questionnaire regarding their experience with the tasks and stimuli as well as possible rating strategies. Finally, participants were debriefed and reimbursed for their participation.

2.9 Eye-tracking

Eye-tracking was carried out using an *SR Research Eyelink 1,000 plus* configured for desktop mount. The distance from the chin rest was 55 cm to the eye-tracker and 90 cm to the screen. The sampling rate was 1,000 Hz. Calibration and validation were performed before the rating task with a 9-point calibration procedure. During the rating task, we additionally included a drift check after every tenth trial to improve the quality of the eye-tracking data. Blinks were excluded from the analysis. Eye-tracking data of 3 participants (2 controls, 1 autistic) had to be discarded due to technical problems and did not enter the relevant analyses, i.e., the analysis of fixation durations and the Bayesian models for object recognition rates.

2.10 Analysis

The permutation software was implemented in *R* (R Core Team, 2023). Other analyses were implemented in *R* (R Core Team, 2019) and *RStudio* (RStudio Team, 2016). When reporting significance of t-tests, we assumed a 95% confidence interval.

For the analysis of participants' ratings, we performed non-parametric permutation tests (Odén and Wedel, 1975; Pesarin and Salmaso, 2010; Berry et al., 2011; Good, 2013) to determine likelihoods of the effects of conditioning arising by chance. These tests explored the effect of the virtual character's gaze duration and pitch height on the participant's rating as to how important an object was considered to be for the character. The dependent variable predicted in these tests was the raw rating data. Corresponding to the four experimental groups, participants' data sets were grouped into four sets of equal size, with the same number of participants with an ASD diagnosis and control participants. Within each experimental group, participants were arranged into pairs, each containing one person of each diagnostic group, with the pairs aligned for maximum age similarity. Thus, experimental group, age-pair, and diagnosis together served to specify a single participant. The conditions of gaze and pitch variation were assessed by using within-subject permutations, while the effect of diagnosis was assessed by permuting data between participants matched for group and age-pair.

For each condition, we ran 1,000,000 permutations. Permutation evaluations were treated as independent samples from a distribution, and the beta function was used to assess the extent of the 95% confidence interval for the likelihood p of a permutated value for the rating exceeding the actual value. This upper limit on the confidence value is reported as p below.

For the analysis of eye-tracking data three regions of interest were defined: The eye region was defined by a rectangle (212 × 110 px) containing the eyes and a small area around the eyes, including the eyebrows. The head region was defined by a rectangle (280 × 414 px) fitting the virtual character's head and excluding the region of interest defined for the eye region. The object region was defined as a square (280 × 280 px) that included the object and a small area around the object to account for the slightly different objects' proportions while at the same time keeping this region of interest constant across trials. Further, for the analysis of eye-tracking data and the recognition task (i.e., the correctness of the responses as to whether an object had appeared in the main experiment or not), Bayesian models (package *brms*; Bürkner, 2017; Bürkner and Vuorre, 2019) were fitted to the data. If not otherwise stated, dichotomous factors were deviation-coded, and continuous factors were z-transformed. In each model, we included random intercepts and slopes for *subject* as well as random intercepts for *object*. Estimated parameters are reported in terms of posterior means and 95% credibility intervals. The *emmeans* package (Lenth et al., 2021) was used to extract contrast coefficients. To investigate the evidence for or against the investigated effects, we compared models by calculating Bayes factors applying the *bayesfactor_models* function from the *bayestestR* package (Makowski et al., 2019) which uses bridge sampling (Gronau et al., 2020). We report respective Bayes factors of model comparisons and follow the interpretation by Lee and Wagenmakers (2014). All models ran with four sampling chains of 12,000 iterations each including a warm-up period of 2,000 iterations.

For the analysis of the influence of *diagnosis* and *cluster* and their *interaction* on the duration of fixations within the three regions of interest, we included eye-tracking data starting at the onset of the gaze cue (= onset of the auditory stimulus). We modelled a proportional value for fixation duration, namely fixation duration directed towards the region of interest divided by the video duration starting at cue onset, separately for each region. Bayesian linear zero-inflated beta models [r package *brms*; Bürkner, 2017; Bürkner and Vuorre, 2019) were fitted to the data. Fixed effects were *diagnosis* and *cluster*. Weakly informative priors were used (intercept prior: normal distribution, $M = 0.5$, $SD = 0.5$; slope priors: normal distribution, $M = 0$, $SD = 0.5$; SD priors: normal distribution, $M = 0$, $SD = 0.5$; phi priors: normal distribution, $M = 0.5$, $SD = 0.5$; zi prior: $M = 0.2$, $SD = 0.5$).

The Bayesian logistic binomial regression model for object recognition in the recognition task was fitted exclusively to data pertaining to stimuli presented in one of the four conditions. Thus, false positive responses or true rejections following the presentation of distractors were not analyzed. We included fixed effects previously identified as important in the general population: the untransformed, proportional values for *participant's gaze duration* towards the object region during the rating task; the logarithmized values of *word frequency*; and the *number of trials that had passed since object presentation*. We compared this model with models that additionally included *ASD diagnosis,* the virtual character's *gaze duration* towards the object and *pitch height*. Weakly informative priors were used (intercept prior: normal distribution, $M = 0$, $SD = 0.5$; slope priors: normal distribution, $M = 0$, $SD = 0.5$; SD priors: normal distribution, $M = 0$, $SD = 0.5$; LKJ prior: 1). Results are reported on the log-odds scale.

3 Results

Scores indicating autistic traits, measured with the AQ, were significantly higher in participants with autism compared to the control group [Table 3, two-samples t-test, $t(46) = -20.18$, $p < 0.001$]. Scores indicating empathetic traits, measured with the EQ, were significantly lower in autistic participants compared to the control group [Table 3, Welch two-samples t-test, $t(37.07) = 14.42$, $p < 0.001$]. Scores indicating tendencies to systemize, as measured with the *SQ*, were significantly higher in autistic participants compared to the control group [Table 3, two-samples t-test, $t(46) = -5.64$, $p < 0.001$]. Mentalizing abilities, as indicated by the *Eyes-test* scores, were

10.3389/fcomm.2024.1483135

TABLE 3 Psychological screening scores.

	AQ	EQ	SQ	Eyes-test
ASD ($N = 24$)	$M = 42.1$ ($SD = 4.3$)	$M = 11.5$ ($SD = 6.0$)	$M = 45.0$ ($SD = 13.8$)	$M = 16.0$ ($SD = 4.6$)
Control ($N = 24$)	$M = 14.2$ ($SD = 5.3$)	$M = 46.3$ ($SD = 10.2$)	$M = 23.6$ ($SD = 12.4$)	$M = 19.0$ ($SD = 2.8$)

AQ, autism spectrum quotient; EQ, empathy quotient; SQ, systemizing quotient; Eyes-test, "Reading the mind in the Eyes" test.

significantly higher in the control group than in the autistic group [two-samples t-test, $t(46) = 2.76$, $p = 0.008$]. These results further support the clinical diagnosis.

3.1 Rating behavior

The condition characterized by short gaze duration and low pitch height yielded the lowest mean ratings in both the autistic group ($M = 1.90$, $SD = 0.53$) and the control group ($M = 2.14$, $SD = 0.37$). The condition with both long gaze and high pitch yielded the highest mean ratings in both groups (ASD: $M = 2.70$, $SD = 0.54$; control persons: $M = 2.65$, $SD = 0.48$). The conditions with either longer gaze duration (ASD: $M = 2.55$, $SD = 0.61$; control persons: $M = 2.45$, $SD = 0.46$) or increased pitch height (ASD: $M = 2.13$, $SD = 0.74$; control persons: $M = 2.45$, $SD = 0.58$) yielded mean ratings between the two afore-mentioned conditions. Mean ratings (see Figure 4) therefore replicate the general pattern reported for a sample from the general population in our previous web-based study (Zimmermann et al., 2020).

We assessed the significance of the differences in ratings as a function of condition and diagnosis by means of permutation tests. Long gaze significantly increased participants' ratings ($p < 0.001$). This held true regardless of the combination of diagnosis and pitch, i.e., both in the autistic and non-autistic group, ratings in conditions in which the virtual character looked towards the object for a long duration were higher than those for conditions in which the gaze was short, both for the high-pitch and low-pitch conditions. Pitch height had a slightly weaker impact on the ratings but again significantly increased participants' ratings ($p < 0.001$) for all combinations of diagnosis and gaze duration, i.e., both in the autistic and non-autistic group, ratings in conditions in which pitch was high, were higher than those for conditions in which pitch was low, both for the long-gaze and the short-gaze conditions. The only exception from this general pattern were participants diagnosed with ASD looking at long gaze: for this latter combination, the effect was also significant ($p = 0.001$), but potentially more likely to have occurred by chance.

Finally, we examined the impact of diagnosis on distinct combinations of pitch height and gaze duration. For short gaze, regardless of pitch height, ratings of autistic participants were significantly lower than those of the control group ($p < 0.001$). When gaze was long but pitch was low, ratings of autistic participants were significantly higher than those of the control group ($p = 0.004$). When both gaze was long and pitch was high, ratings of autistic and non-autistic participants did not differ significantly ($p = 0.130$). Rating differences in response to the two different gaze cue durations were thus greater in the autistic group than in the non-autistic group, indicating that different gaze durations of the virtual character towards the object had a greater effect in the autistic group.

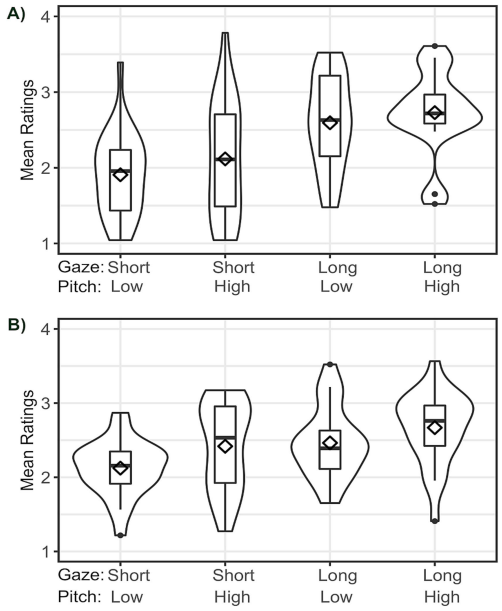

FIGURE 4
Distribution of participants' mean ratings of stimuli: **(A)** ASD group, **(B)** Control group. The range of the y-axis equals the total rating scale (1–4). Diamonds indicate means across subjects. Horizontal lines indicate medians.

These results reflect the visible differences by condition and diagnosis seen in Figure 4.

3.2 Individuality

Similar to the findings in our web-based study (Zimmermann et al., 2020), there was substantial inter-individual variability. Participants' ratings were predominantly influenced by either one or the other factor rather than by both factors in combination. Figure 5 shows each participant's individual cue use behavior regarding *gaze duration* and *pitch height,* indicated by the difference between their mean ratings for long vs. short gaze duration conditions and the difference between their mean ratings for high vs. low pitch height conditions. For each participant, we carried out two Wilcoxon rank sum tests—including the expectation that longer gaze and higher pitch would each increase ratings—on the ratings for the long- versus short-gaze and high- versus low-pitch conditions, respectively. The resulting

10.3389/fcomm.2024.1483135

p-values indicating significant differences at the 5% level were used as indicators that the individual made use of the respective cue. Participants were subsequently categorized as "Lookers" if ratings were significantly higher in the long-gaze conditions than in the short-gaze conditions, as "Listeners" if ratings were significantly higher in the high-pitch conditions than in the low-pitch conditions, as "Neithers" if ratings did not differ significantly for either factor, and as "Both" if ratings significantly differed for both factors. However, no participant was categorized as "Both" in this dataset, irrespective of diagnosis, mirroring results from our previous study (Zimmermann et al., 2020). The resulting three clusters are color-coded in Figure 5. Participants of both diagnostic groups can be found across all three clusters.

Interestingly, three participants that clustered as "Lookers" (two of these autistic) reported initially having used the virtual character's (tone of) voice for their ratings, but switching to concentrating on the character's gaze towards the objects, once they had detected this cue. In the "Listeners" cluster, only one participant reported also having used the character's gaze towards the object for their ratings.

Participants clustered as "Neithers" were not consistently influenced by either gaze duration or pitch height. However, when asked for their rating strategy in free-text form, some of the "Neithers" reported having taken into account the gaze behavior of the virtual character or the voice stimulus for their ratings, few specifically referred to the virtual character's gaze duration towards the objects or the tone of voice. However, none of the participants in the "Neithers" cluster reported exclusively having taken into account either intonation or gaze duration towards the object (or both). Instead, they attended to more than one source of information, amongst them the character's blinking behavior, the duration of the second idle gaze phase, gaze direction and loudness. One participant reported that the different durations of the character's gaze towards the object did not influence their rating behavior as it did not affect their perception as to how important the objects appeared to be for the character. Only one (autistic) participant in the "Neithers" cluster reported sometimes having guessed. Across both autistic and non-autistic participants, some reported having concentrated on the object itself (its animacy, entertaining quality, potential benefit or danger) or their personal perception of the object's importance as well as the object's presumed importance for the virtual character based on her age, gender and appearance.

3.3 Fixation durations

Overall, both the autism group and the control group spent more time looking at the eye region (ASD: $M = 3.88$ s, $SD = 1.61$; control group: $M = 4.10$ s, $SD = 1.26$) than at the object (ASD: $M = 1.01$ s, $SD = 0.74$; control group: $M = 1.04$ s, $SD = 0.74$) and head region (ASD: $M = 1.05$ s, $SD = 0.65$; control group: $M = 1.05$ s, $SD = 0.60$) (see Figure 6).

Across diagnostic groups, within the three clusters, rating behavior was reflected by fixation durations within the three regions: Compared to the other groups, the group of "Lookers" looked longer at the eye region ("Lookers": $M = 4.85$ s, $SD = 1.04$; "Listeners": $M = 3.01$ s, $SD = 1.36$; "Neithers": $M = 3.67$ s, $SD = 1.30$), but spent less time fixating the object region ("Lookers": $M = 0.55$ s, $SD = 0.24$; "Listeners": $M = 1.49$ s, $SD = 0.85$; "Neithers": $M = 1.30$ s, $SD = 0.69$). Fixation durations within the head region (not including the eye region) were similar between clusters ("Lookers": $M = 0.92$ s, $SD = 0.64$; "Listeners": $M = 1.09$ s, $SD = 0.61$; "Neithers": $M = 1.24$ s, $SD = 0.59$). Visual inspection suggested that within the "Listeners" cluster, the difference between mean fixation durations towards the eye region for participants with an ASD diagnosis and control participants tended to be greater than the respective difference within the other two

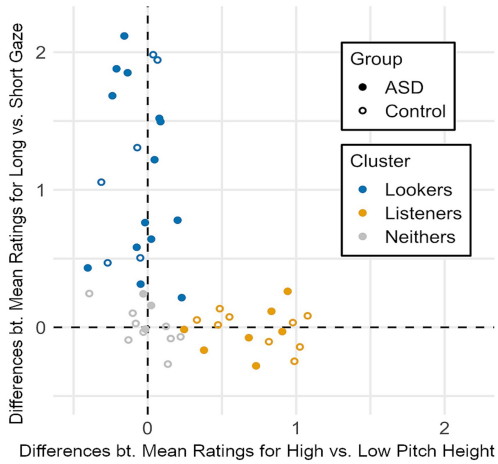

FIGURE 5
Individual differences between mean ratings for high vs. low pitch height conditions (*x*-axis) and individual differences between mean ratings for long vs. short gaze duration conditions (*y*-axis) combined to one coordinate for each participant. Cluster labelling is based on the significance of these differences at the 5% level. There are no participants with significant differences on both axes.

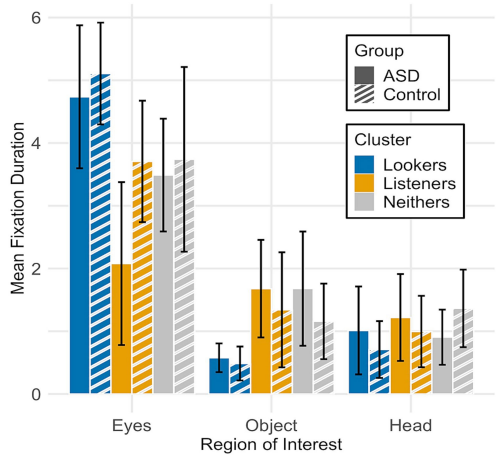

FIGURE 6
Mean fixation durations and standard deviations for the ASD and control group and the three clusters within the three regions of interest. The total possible per-trial fixation duration is 6.6 s.

clusters. This observation was, however, not supported by statistical analysis.

We analyzed the influence of *diagnosis* and *cluster* on fixation durations beginning at gaze cue onset (which coincides with the onset of the auditory stimulus), separately for each region of interest. In sum, *cluster* was identified as a statistical reliable influence on total fixation duration within the eye region and the object region, however, *diagnosis* was not: we found only anecdotal evidence for a *diagnosis* effect for the eye region ($b = -0.36$; 95% CI = [−0.77, 0.05], BF = 1.93), but anecdotal evidence against an effect in the object region ($b = 0.19$; 95% CI = [−0.10, 0.50], BF = 0.73) and in the head region ($b = 0.01$; 95% CI = [−0.37, 0.40], BF = 0.39). In support of the finding of cluster-dependent gaze patterns reported above, extreme evidence for an effect of *cluster* was found in the eye region (BF > 100) and object region (BF > 100), while anecdotal evidence against an effect was found in the head region (BF = 0.57). Specifically, and irrespective of diagnostic group, "Lookers" spent more time fixating the eye region than "Listeners" ($b = 0.95$; 95% CI = [0.50, 1.39], BF > 100) and tended to also spend more time fixating this region than "Neithers" ($b = 0.61$; 95% CI = [0.09, 1.14], BF = 4.90), while "Listeners" tended to spend less time fixating the eye region than "Neithers" ($b = -0.35$; 95% CI = [−0.82, 0.16], BF = 1.33). In comparison to the "Lookers," the "Listeners" ($b = 0.63$; 95% CI = [0.30, 0.96], BF > 100) and "Neithers" ($b = 0.67$; 95% CI = [0.28, 1.05], BF = 70.42) spent more time fixating the object region, while there was no difference between "Listeners" and "Neithers" ($b = -0.04$; 95% CI = [−0.40, 0.33], BF = 0.38). We found anecdotal evidence against an interaction effect of *diagnosis* and *cluster* in the eye region (BF = 0.70) and object region (BF = 0.40). Moderate evidence for an interaction effect was found in the head region (BF = 3.71). Further investigation of this effect revealed that it was mainly driven by tendencies within the control sample: participants in the "Neither" cluster tended to fixate the head region for a longer duration than both "Lookers" (BF = 4.97) and—to a lesser extent—"Listeners" (BF = 3.27), while no difference was found between the "Listeners" and "Lookers" (BF = 0.62). Within the autism sample, no statistically reliable differences between clusters were found for the head region (0.48 < BFs < 1.0).

3.4 Object recognition

Recognition rates for target words tended to be slightly lower in the autistic group compared to the control group, but were similar within groups for all four conditions (Table 4). Correct identification of distractor words was comparable between groups (ASD: $M = 93.0\%$, $SD = 5.4$; Controls: $M = 93.3\%$, $SD = 5.4$).

For target words, we found extreme evidence for an effect of *participants' fixation* duration towards the object on their memory performance, with longer fixation of an object increasing recognition ($b = 0.49$; 95% CI = [−0.03, 0.98], BF > 1,000). The *number of trials that had passed since object presentation* also had a statistically robust effect on memory performance: The fewer trials passed since object presentation, the greater the likelihood the respective word was recognized correctly in the recognition task ($b = -0.28$, 95% CI = [−0.37, −0.19]; BF > 1,000). Additionally, very strong evidence was found for an effect of *word frequency*: more frequent words tended to lead to better recognition ($b = 0.06$, 95% CI = [−0.09, 0.22]; BF > 100). Anecdotal evidence was found for including the factor *ASD diagnosis* (BF = 1.44). Including the factors *gaze duration* or *pitch height* did not improve model fit (gaze duration: BF = 0.07; pitch height: BF = 0.22).

3.5 Exploratory correlation analysis

Within the two diagnostic groups, we performed exploratory correlation analyses for differences between mean ratings (for high vs. low *pitch height* conditions and for long vs. short *gaze duration* conditions; see Figure 5) in combination with the *SPQ visual* and *auditory* scores. We found a statistically noteworthy relationship for the *SPQ* regarding *gaze duration*: In the control group, higher *SPQ visual* scores (indicating lower visual sensitivity) were significantly linked to taking gaze duration into account to lesser extent ($r_S = -0.444$, $p = 0.030$), which was not the case in the autistic group ($r_S = -0.183$, $p = 0.393$). No significant correlation between *SPQ visual* scores and differences between mean ratings for *pitch height* conditions was observed in the autistic and control group (ASD: $r_S = 0.012$, $p = 0.956$; Controls: $r_S = 0.290$, $p = 0.169$). No significant correlation was found between *SPQ auditory* scores and the differences between mean ratings for *gaze duration* conditions (ASD: $r_S = 0.165$, $p = 0.442$; Controls: $r_S = -0.092$, $p = 0.669$) and *pitch height* (ASD: $r_S = 0.047$, $p = 0.828$; Controls: $r_S = 0.150$, $p = 0.486$).

4 Discussion

4.1 Rating behavior

At the group-level, participants from both the autism group and the control group rated the importance of the object to the virtual character to be higher when any of the two deictic signals (gaze or pitch accent) suggested that the virtual character was more interested in the particular object (through longer gaze or higher pitch), confirming the results of our previous web-based study (Zimmermann et al., 2020).

Compared to the control group, autistic participants took gaze duration into account to a greater extent than pitch height: They

TABLE 4 Object recognition rates.

	Group	Low pitch	High pitch
Short gaze duration	ASD ($N = 24$)	$M = 54.2\%$ ($SD = 21.0$)	$M = 56.6\%$ ($SD = 24.5$)
	Control ($N = 24$)	$M = 67.6\%$ ($SD = 19.5$)	$M = 68.5\%$ ($SD = 21.8$)
Long gaze duration	ASD ($N = 24$)	$M = 60.0\%$ ($SD = 22.3$)	$M = 59.2\%$ ($SD = 20.3$)
	Control ($N = 24$)	$M = 66.0\%$ ($SD = 18.6$)	$M = 68.5\%$ ($SD = 20.0$)

judged the object's importance to the virtual character to be lower than the control group when it was gazed at for a short duration. They rated the importance higher than the control group when the object was gazed at for a long duration if presented with low pitch – and as high as the control group if it was presented with high pitch.

One explanation for the fact that participants with autism in our paradigm assigned more weight to the virtual character's gaze (as opposed to pitch height) might be an impaired interpretation of vocal pitch, both in speech (Grice et al., 2016, 2023; Schelinski and von Kriegstein, 2019) and non-speech (Schelinski et al., 2017). The study by Grice et al. (2023) suggests that the interpretation of prosody (amongst others intonation) is similar in autistic listeners and non-autistic listeners when it is used by the speaker to convey rule-based information such as syntactic structure. However, when it is used to convey less rule-based and more intuitive pragmatic aspects, such as the importance of a certain word, the interpretation of prosodic information seems to be more difficult for autistic listeners. An example for the latter is an investigation of intonation perception in autism (Grice et al., 2016): In this study, autistic listeners were less sensitive to intonation than the non-autistic group. Instead, they used other information about the words themselves, such as semantic information (human-non-human for instance), to judge whether a word presented in an auditorily presented sentence was new information or not. If participants in our paradigm found it difficult to interpret pitch height, this might be a reason for them to instead search for other information to solve the task.

Another reason for autistic participants to more strongly weigh the gaze cue rather than the pitch cue could be greater auditory capacity in comparison to control participants (Remington and Fairnie, 2017). In this study, autistic listeners were able to detect more auditory stimuli than the non-autistic group, regardless of whether they were distractors to the main task or not. Perceiving a wealth of auditory information might be beneficial in certain scenarios but could also be detrimental or exhausting in others. In our paradigm, the auditory information is arguably more complex than the visual information: The speech stimulus was a different one in each trial. Furthermore, since we used natural speech, the intonation pattern slightly varied for each item: Even if the accented syllable of each high-pitch stimulus is always 45 Hz higher relative to its low-pitch counterpart, low-pitch stimuli exhibit small fluctuations in their absolute Hz values. Additionally, other prosodic factors might influence prominence perception, such as the length of the utterance. The gaze cue, on the other hand, is comparably simple to perceive and categorize, as it was always set to either 0.6 or 1.8 s in a binary fashion. Therefore, a person processing the abundance of information presented with the auditory cue might find it easier to pay attention to the gaze cue instead, either because they do not detect the manipulated cue amongst the noise of other auditory information, or because this is more effortful than focusing on gaze duration.

The findings of the exploratory analyses showed that, in part, rating tendencies could plausibly be linked to sensory perception: within the control group, lower visual sensitivity was linked to less focus on gaze. This suggests that general visual sensitivity affects participants' ratings. One reason for a lack of this relationship in the autistic group could be that – instead of relying on their default perception – they attuned to the task's systematic structure more strongly than the control group did, which could also explain why they weighed the gaze cue more strongly than the pitch cue.

Other studies have reported that autistic participants had difficulties in solving mentalizing tasks that rely on nonverbal information (Baron-Cohen et al., 2001a; Baron-Cohen et al., 2001b; Ponnet et al., 2004; Dziobek et al., 2006; White et al., 2011). Those tasks involved more than two signals that varied in more than two steps, so that it was unclear which cue was informative. Additionally, the response required more complex mentalizing tasks than the current experiment (e.g., identifying different mental states from a selection of alternatives, or freely inferring mental states). In contrast, our task provides a much more structured setting, with only two cues varying by two different degrees. Moreover, the simple question to be answered is the same throughout. The most obvious strategy to solve the task is to identify (at least) one varying source of information and preferentially rely on that source.

4.2 Individuality

Gaze cues (Bayliss et al., 2007) and pitch height cues (Roy et al., 2017; Baumann and Winter, 2018) are not perceived and processed in the same way by every individual. Participants' ratings in our study tended to be influenced by either one or the other factor rather than by both factors in combination. Based on their rating behavior, participants clustered into three subgroups: (i) "Lookers," who based their ratings primarily on gaze duration, (ii) "Listeners," who based their ratings primarily on pitch height, and (iii) "Neithers," whose ratings were not predominantly influenced by either of these two cues. Participants of both diagnostic groups were found across all three clusters. The observation discussed above that autistic participants were more strongly influenced by the gaze cue was reflected in the distribution of clusters as well: autistic participants were identified as "Lookers" twice as often as they were identified as "Listeners." This pattern was not visible in control participants: six participants were categorized as "Lookers," whereas nine were categorized as "Listeners" in the control group. Based on previous findings of the high relevance of verbal at the expense of nonverbal information in autism (Kuzmanovic et al., 2011; Stewart et al., 2013) and of a reliance on invariant characteristics of words at the expense of intonation (Grice et al., 2016), we expected more autistic participants to cluster as "Neithers." However, this was not the case.

It is striking that none of the participants was considerably influenced by both gaze duration and pitch height together. Several studies that investigated the perception of pitch accents in combination with salient facial movement, head or hand gestures in the general population have shown that they can, in fact, lead to greater prominence perception compared to the presentation of only one modality (Krahmer et al., 2002; Swerts and Krahmer, 2008; Mixdorff et al., 2013; Prieto et al., 2015; Ambrazaitis et al., 2020). A possible explanation for the finding that, at the individual level, a combination of long gaze and high pitch in the current paradigm did not lead to higher ratings of object importance compared to when only one of the two cues was rendered prominent, might be that participants default to efficient cue use in this task. The instruction did not specify whether the virtual character would communicate via eye gaze, prosody or other behavior. Accordingly, participants had the freedom to use one, two, multiple or no cues at all. Increased multimodal cue use has been

reported in audiovisual studies in which auditory information is insufficient or difficult to understand (Munhall et al., 2004; Dohen and Lœvenbruck, 2009; Moubayed and Beskow, 2009; Macdonald and Tatler, 2013). For example, in a demanding, but highly structured task (Macdonald and Tatler, 2013), participants from the general population made use of the instructor's gaze behavior only, if the auditory information was not informative enough. Comparably, in the current paradigm, there was no need for participants to identify additional cues, as long as they found at least one cue that helped them solve the task. Identifying one cue and sticking to it may be the most efficient way to solve this task. Participants' feedback regarding their rating strategies lends anecdotal support for this idea: Four participants explicitly reported having focused on the virtual character's gaze towards the object as well as intonation. Three of these participants (two of them autistic) reported having used primarily the gaze cue for the remainder of the experiment, which exemplifies the efficiency of participants' cue use in this task.

The finding that participants in the "Neithers" cluster did not demonstrate a preference for either the gaze cue or the pitch cue does not necessarily imply that these did not affect their ratings at all, but that they weighed other cues more strongly. Feedback from these participants on their rating strategies suggests that some focused on the object's properties and the virtual character's characteristics when carrying out their ratings. Others did, in fact, attend to the character's gaze behavior and the voice stimuli, but considered aspects of gaze and voice other than the manipulated cues, such as gaze directions, blinking or voice loudness. Those that actually took into account the manipulated cues, additionally paid attention to other cues that were not manipulated, which may have attenuated potential effects of gaze duration or pitch height on their ratings.

4.3 Eye-tracking

Both diagnostic groups spent more time fixating the eye region than the object and head region. This finding is in line with previous eye-tracking studies: in the general population, a tendency to fixate the eye region for longer than either other parts of the face or objects in the environment has been reported across different tasks (Henderson et al., 2005; Freeth et al., 2010; Fedor et al., 2018). Similar fixation tendencies have been reported for individuals on the autism spectrum (Dalton et al., 2005; Hernandez et al., 2009; Freeth et al., 2010; Auyeung et al., 2015; Fedor et al., 2018).

We did not find reliable statistical evidence for differences between the autism and the control group regarding fixation durations for the eye region or the object region. A meta-analysis of 22 studies has reported shorter fixation durations for the eye region as opposed to objects in adult participants with autism in free viewing tasks (Setien-Ramos et al., 2022). Our paradigm was not suited to induce gaze aversion in autism as it required participants to search for potentially informative cues. Information variation was limited to the eyes, voice and object, and only the eye region showed visual change within a given trial (eye blinks, changing gaze direction). Thus, avoiding the character's gaze (and assuming the eye region is not processed via peripheral vision) would entail ignoring one of three relevant channels of information. Presenting only one rather static virtual character as well as a relatively long trial duration may further have shifted attention towards the eye region in our study.

Across both diagnostic groups, we were able to show that participants' rating behavior was in line with their gaze behavior: "Lookers" spent more time looking at the virtual character's eyes than "Listeners" and tended to also spend more time looking at the eyes than "Neithers." "Listeners" spent less time looking at the eyes than "Neithers." "Lookers" spent less time looking at the object than both other clusters, which did not differ in this regard. This finding corroborates the well-established notion that attention is closely linked to gaze direction (Buswell, 1935; Yarbus, 1967; DeAngelus and Pelz, 2009), leaving the "Lookers" no choice but to fixate the eye region and mostly ignore the object, while "Listeners" and "Neithers" were free to visually explore other areas as well.

Exclusively for the eye region, visual inspection—but not the statistical analysis—showed a small tendency for shorter fixation durations in "Listeners" with an ASD diagnosis compared to "Listeners" from the control group. It is possible that an underlying trend was not detected in the analysis. If present, it could suggest different strategies for solving the task: "Listeners" need to pay attention to the acoustic signal and do not depend on gathering information from the eyes. Especially for people with autism, who may experience mutual gaze as threatening or stressful, this could result in avoiding mutual gaze (Tottenham et al., 2014). In our study, we did not ask about uneasiness while fixating the eye region. Only one participant in the autism group reported exhaustion due to looking at the virtual character's face and the eye region in particular. A tendency within the autism group for the "Listeners" to look at the eye region for a shorter total duration could also indicate that persons with autism by default perceive the eyes as deictic cues but not as mutual gaze, which is a stronger social cue (Ristic et al., 2005; Caruana et al., 2018). Riby et al. (2013), who included children and adolescents with autism in their study, reported that the eye region was fixated for a shorter duration by their autistic group in comparison to a control group. In the autistic group, fixation duration towards the eye region, unsurprisingly, increased upon instruction to detect what the person in the photo was looking at.

In the control group, we found a tendency towards shorter fixations of the head region (not including the eyes) in the "Lookers" and "Listeners" compared to the "Neithers." The behavior in the rating task and our eye-tracking data support the idea that participants were actively monitoring their chosen input modality, searching for informativeness in these cues. Accordingly, we interpret the tendency of the "Neither" cluster as more strongly than the other clusters using the head region as a source of information. Three participants in the "Neither" cluster reported having taken into account virtual-character-related characteristics such as gender and age for their rating. Only one subject from the "Listeners" cluster reported potentially having been influenced in a similar fashion.

4.4 Object recognition

To detect possible memory traces of attention directed towards the objects, we included an object recognition task after completion of the rating task. Findings regarding word or object recognition in autistic adults without intelligence deficits have been mixed so far, with some studies reporting comparable performance in autism (Bowler et al., 2000; Boucher et al., 2005; Ring et al., 2015) and others showing worse recognition rates in autism (O'Hearn et al.,

2014). We found no reliable evidence for different recognition rates in participants with autism compared to the control group. Across groups, object recognition was better for objects that had previously been fixated by participants for a longer duration, which is in line with previous research on visual memory: the longer we look at an object, scene or face, the better we can later remember it (Melcher, 2001, 2006; Droll and Eckstein, 2009; Martini and Maljkovic, 2009). We also found a serial-position effect: participants could better recognize objects they had seen more recently, which is in line with previous research (Brady et al., 2008; Konkle et al., 2010). Importantly for our purposes, implicit memory in autism is considered comparable to that in the non-autistic population (Ring et al., 2015). Our results stand in contrast to other studies that reported an influence of gaze and pitch on object memory (Fraundorf et al., 2010, 2012; Dodd et al., 2012; Adil et al., 2018; Wahl et al., 2019; Ito et al., 2022). However, these studies are not directly comparable to our study because they manipulate neither gaze duration nor pitch excursion specifically.

In our study, low word frequency did not improve object memory. Instead, participants could better recognize objects described with more frequent words. Our paradigm is primarily designed to probe a very simple mentalization task requiring a judgment of how important an object appears to be for the virtual "person." Comparably, in our online study (Zimmermann et al., 2020), participants' ratings for the importance of the object for the virtual character increased with higher word frequency. We assume that word frequency in our stimulus set is confounded with other object properties (Zimmermann et al., 2020). The five most frequent words in our dataset were the German words for "car," "plane," "window," "sun" and "church." The five least frequent words were the German words for "spinning wheel," "doorknob," "spinning top," "chisel" and "roller skate." We assume that, among other factors, general object importance might have affected the ratings.

4.5 Limitations

To summarize the limitations of our paradigm discussed in previous work (Zimmermann et al., 2020), the most pertinent issue is the reductionistic design employed and its effect on the perception of the virtual character's mental state. This entails that the relevant experimental findings cannot be easily transferred to everyday social situations, which are much more complex. Additionally, the task instructions may have led participants to actively search for a cue and to then stop searching once a valid cue was found. In the following section, we will focus on issues specific to autism and the findings of the current study.

It could be argued that autistic participants may concentrate on gaze more than on intonation, because eye contact is a common target in early interventions in autism. However, most participants in our study were recruited in the outpatient clinic for autism in adulthood. This implies that they did not receive any autism-specific therapy before their diagnosis in adulthood. As we have not systematically asked every participant whether he or she received any specific training in nonverbal communication skills including mutual gaze, it cannot be ruled out that they did, but it is unlikely. Assessing a potential influence of such a training may nevertheless be informative in future studies.

It is possible that the external validity of our results is not only limited, but also differs between diagnostic groups. A study including children and adolescents has shown that the gaze behavior of participants with autism in reaction to a computer screen differed from gaze behavior in reaction to a live interaction, which was not the case for the control group (Grossman et al., 2019). No difference was, however, reported for gaze behavior in reaction to static images of virtual characters' faces as compared to photographs of real people in autistic adolescents and adults (Hernandez et al., 2009). In real-life scenarios, the problems persons with autism face when interpreting eye gaze do not only arise from difficulties with deciphering the "correct" social implications, but also from understanding when eye gaze may contain social implications in the first place, beyond, e.g., deictic information, which is itself problematic in autism (Pantelis and Kennedy, 2017; Griffin and Scherf, 2020). The latter was not part of this experiment, as the task (according to the interpretation of most participants) implicitly called for a social reading. Due to the limited stimulus variability, attending to the relevant cues was, moreover, easier than in real-life scenarios.

Future studies investigating the perception of gaze duration and intonation in a non-verbal mentalizing task in autism should aim to increase ecological validity by (1) using more natural social scenarios as stimulus material that does not only vary with respect to two isolated cues, and (2) by using different instructions or questions in each trial. We expect that this may reduce potential strategic cue searching strategies.

5 Conclusion

The current study aimed to investigate mentalizing behavior based on eye gaze and speech intonation in autism. Comparably to control participants, autistic persons used both gaze duration and intonation as cues for inferring the importance of an object for another (virtual) person. Compared to the control group, autistic participants were, however, influenced more strongly by gaze duration than by pitch height. Across both diagnostic groups, participants used either gaze or intonation as predominant cues, while some did not show this cue preference but might have used other cues predominantly to make their decision. Further investigations are required to accurately characterize differences in mentalizing abilities in autism in the nonverbal domain.

Data availability statement

The datasets presented in this article are not readily available because data handling is restricted to the collaborating institutes by our ethics proposal to secure sensitive data such as psycho(patho-) logical data. Requests to access the datasets should be directed to kai.vogeley@uk-koeln.de.

Ethics statement

The studies involving humans were approved by the Ethics Committee of the University Hospital Cologne. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

JZ: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Validation, Visualization, Writing – original draft, Writing – review & editing, Software, Supervision. TE: Data curation, Formal analysis, Writing – review & editing. FC: Methodology, Software, Writing – review & editing, Conceptualization. SW: Methodology, Writing – review & editing, Conceptualization. KV: Conceptualization, Funding acquisition, Methodology, Resources, Supervision, Writing – review & editing. MG: Conceptualization, Funding acquisition, Methodology, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. The study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) -Project-ID 281511265 -SFB 1252.

Acknowledgments

We would like to thank C. Bloch, N. Oliveira-Ferreira, P. Da Silva Vilaca, and C. Esser for their support during data-acquisition, C. Röhr for producing the audio stimuli, H. Hanekamp for extracting acoustic parameters from these stimuli and the staff of the Phonetics Department of the University of Cologne for rating the processed audio stimuli.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomm.2024.1483135/full#supplementary-material

SUPPLEMENTARY TABLE 1
Details on the pretest's methods and results.

SUPPLEMENTARY TABLE 2
Object names and translations.

SUPPLEMENTARY DATE SHEET 1
Auditory stimuli.

SUPPLEMENTARY DATE SHEET 2
Video stimuli.

References

Adil, S., Lacoste-Badie, S., and Droulers, O. (2018). Face presence and gaze direction in print advertisements: how they influence consumer responses—an eye-tracking study. *J. Advert. Res.* 58, 443–455. doi: 10.2501/JAR-2018-004

Ambrazaitis, G., Frid, J., and House, D. (2020). Word prominence ratings in Swedish television news readings – effects of pitch accents and head movements, *The respective conference was the 10th International Conference on Speech Prosody in* Tokyo, Japan. 314–318. doi: 10.21437/SpeechProsody.2020-64

American Psychiatric Association (2013). Diagnostic and statistical manual of mental disorders. DSM-5. Washington, DC: American Psychiatric Association.

Arnold, D., Wagner, P., and Baayen, H. (2013). "Using generalized additive models and random forests to model prosodic prominence in German" in INTERSPEECH 2013, Eds. F. Bimbot, C. Cerisara, C. Fougeron, G. Gravier, L. Lamel, F. Pellegrino, P. Perrier (Lyon, France: ISCA) 272–276.

Aster, M., Neubauer, A., and Horn, R. (2006). Hamburg-Wechsler-Intelligenz-Test für Erwachsene III. Bern, Switzerland: Huber.

Aurchitect Audio Software, LLC (2018). Myriad. Aurchitect Audio Software is now onwed by Zynaptiq, Hannover, Germany: Aurchitect Audio Software, LLC.

Auyeung, B., Lombardo, M. V., Heinrichs, M., Chakrabarti, B., Sule, A., Deakin, J. B., et al. (2015). Oxytocin increases eye contact during a real-time, naturalistic social interaction in males with and without autism. *Transl. Psychiatry* 5, e507. doi: 10.1038/tp.2014.146

Baron-Cohen, S., Campbell, R., Karmiloff-Smith, A., Grant, J., and Walker, J. (1995). Are children with autism blind to the mentalistic significance of the eyes? *Br. J. Dev. Psychol.* 13, 379–398. doi: 10.1111/j.2044-835X.1995.tb00687.x

Baron-Cohen, S., Leslie, A. M., and Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition* 21, 37–46. doi: 10.1016/0010-0277(85)90022-8

Baron-Cohen, S., Richler, J., Bisarya, D., Gurunathan, N., and Wheelwright, S. (2003). The systemizing quotient: an investigation of adults with Asperger syndrome or high-functioning autism, and normal sex differences. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 358, 361–374. doi: 10.1098/rstb.2002.1206

Baron-Cohen, S., and Wheelwright, S. (2004). The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *J. Autism Dev. Disord.* 34, 163–175. doi: 10.1023/B:JADD.0000022607.19833.00

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., and Plumb, I. (2001a). The "Reading the mind in the eyes" test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *J. Child Psychol. Psychiatry* 42, 241–251. doi: 10.1111/1469-7610.00715

Baron-Cohen, S., Wheelwright, S., and Jolliffe, T. (1997). Is there a "language of the eyes"? Evidence from Normal adults, and adults with autism or Asperger syndrome. *Vis. Cogn.* 4, 311–331. doi: 10.1080/713756761

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., and Clubley, E. (2001b). The autism-Spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *J. Autism Dev. Disord.* 31, 5–17. doi: 10.1023/A:1005653411471

Baumann, S., and Winter, B. (2018). What makes a word prominent? Predicting untrained German listeners' perceptual judgments. *J. Phon.* 70, 20–38. doi: 10.1016/j.wocn.2018.05.004

Bayliss, A. P., Frischen, A., Fenske, M. J., and Tipper, S. P. (2007). Affective evaluations of objects are influenced by observed gaze direction and emotional expression. *Cognition* 104, 644–653. doi: 10.1016/j.cognition.2006.07.012

Beck, A. T., Steer, R. A., and Brown, G. K. (1996). Manual for the Beck depression inventory-II. San Antonio, TX: Psychological Corporation.

Ben-David, B. M., Ben-Itzchak, E., Zukerman, G., Yahav, G., and Icht, M. (2020). The perception of emotions in spoken language in undergraduates with high functioning autism Spectrum disorder: a preserved social skill. *J. Autism Dev. Disord.* 50, 741–756. doi: 10.1007/s10803-019-04297-2

Berry, K. J., Johnston, J. E., and Mielke, P. W. Jr. (2011). Permutation methods. *WIREs Comput. Statistic.* 3, 527–542. doi: 10.1002/wics.177

Bierlich, A. M., Bloch, C., Spyra, T., Lanz, C., Falter-Wagner, C. M., and Vogeley, K. (2024). An evaluation of the German version of the sensory perception quotient from an expert by experience perspective. *Front. Psychol.* 15:1252277. doi: 10.3389/fpsyg.2024.1252277

Bishop, J. (2016). Focus projection and prenuclear accents: evidence from lexical processing. *Lang. Cognit. Neurosci.* 32, 236–253. doi: 10.1080/23273798.2016.1246745

Bishop, J., Kuo, G., and Kim, B. (2020). Phonology, phonetics, and signal-extrinsic factors in the perception of prosodic prominence: evidence from rapid prosody transcription. *J. Phon.* 82:100977. doi: 10.1016/j.wocn.2020.100977

Boersma, P., and Weenink, D. (2018). Praat: doing phonetics by computer [Computer program]. Available at: http://www.praat.org/ (Accessed April 3, 2020)

Boucher, J., Cowell, P., Howard, M., Broks, P., Farrant, A., Roberts, N., et al. (2005). A combined clinical, neuropsychological, and neuroanatomical study of adults with high functioning autism. *Cogn. Neuropsychiatry* 10, 165–213. doi: 10.1080/13546800444000038

Bowler, D. M. (1992). "Theory of mind" in Asperger's syndrome. *J. Child Psychol. Psychiatry* 33, 877–893. doi: 10.1111/j.1469-7610.1992.tb01962.x

Bowler, D. M., Gardiner, J. M., and Grice, S. J. (2000). Episodic memory and remembering in adults with Asperger syndrome. *J. Autism Dev. Disord.* 30, 295–304. doi: 10.1023/A:1005575216176

Brady, T. F., Konkle, T., Alvarez, G. A., and Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proc. Natl. Acad. Sci.* 105, 14325–14329. doi: 10.1073/pnas.0803390105

Brickenkamp, R. (2002). Test d2: Aufmerksamkeits-Belastungs-test. *9th, revised and newly standardized* Edn. Göttingen: Hogrefe.

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., and Böhl, A. (2011). The word frequency effect. *Exp. Psychol.* 58, 412–424. doi: 10.1027/1618-3169/a000123

Bürkner, P.-C. (2017). Brms: an R package for Bayesian multilevel models using Stan. *J. Stat. Softw.* 80, 1–28. doi: 10.18637/jss.v080.i01

Bürkner, P.-C., and Vuorre, M. (2019). Ordinal regression models in psychology: a tutorial. *Adv. Methods Pract. Psychol. Sci.* 2, 77–101. doi: 10.1177/2515245918823199

Buswell, G. T. (1935). How people look at pictures: A study of the psychology of perception in art. Chicago: University of Chicago Press.

Cangemi, F. (2015). Mausmooth [Praat script]. Available at: http://ifl.phil-fak.uni-koeln.de/sites/linguistik/Phonetik/mitarbeiterdateien/fcangemi/mausmooth.praat (Accessed April 13, 2019).

Caruana, N., Stieglitz Ham, H., Brock, J., Woolgar, A., Kloth, N., Palermo, R., et al. (2018). Joint attention difficulties in autistic adults: an interactive eye-tracking study. *Autism* 22, 502–512. doi: 10.1177/1362361316676204 (Accessed April 13, 2019).

Chita-Tegmark, M. (2016). Social attention in ASD: a review and meta-analysis of eye-tracking studies. *Res. Dev. Disabil.* 48, 79–93. doi: 10.1016/j.ridd.2015.10.011

Chuk, T., Chan, A. B., Shimojo, S., and Hsiao, J. H. (2016). Mind reading: discovering individual preferences from eye movements using switching hidden Markov models., in Proceedings of the 38th Annual Conference of the Cognitive Science Society 2016, 182–187.

Dahan, D., Tanenhaus, M. K., and Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *J. Mem. Lang.* 47, 292–314. doi: 10.1016/S0749-596X(02)00001-3

Dalton, K. M., Nacewicz, B. M., Johnstone, T., Schaefer, H. S., Gernsbacher, M. A., Goldsmith, H. H., et al. (2005). Gaze fixation and the neural circuitry of face processing in autism. *Nat. Neurosci.* 8, 519–526. doi: 10.1038/nn1421

DeAngelus, M. A., and Pelz, J. (2009). Top-down control of eye movements: Yarbus revisited.

Del Bianco, T., Mazzoni, N., Bentenuto, A., and Venuti, P. (2018). An investigation of attention to faces and eyes: looking time is task-dependent in autism Spectrum disorder. *Front. Psychol.* 9:2629. doi: 10.3389/fpsyg.2018.02629

Dodd, M. D., Weiss, N., McDonnell, G. P., Sarwal, A., and Kingstone, A. (2012). Gaze cues influence memory…but not for long. *Acta Psychol.* 141, 270–275. doi: 10.1016/j.actpsy.2012.06.003

Dohen, M., and Lœvenbruck, H. (2009). Interaction of audition and vision for the perception of prosodic contrastive focus. *Lang. Speech* 52, 177–206. doi: 10.1177/0023830909103166

Doughty, M. J. (2001). Consideration of three types of spontaneous eyeblink activity in normal humans: during reading and video display terminal use, in primary gaze, and while in conversation. *Optom. Vis. Sci.* 78, 712–725. doi: 10.1097/00006324-200110000-00011

Droll, J. A., and Eckstein, M. P. (2009). Gaze control and memory for objects while walking in a real world environment. *Vis. Cogn.* 17, 1159–1184. doi: 10.1080/13506280902797125

Dziobek, I., Fleck, S., Kalbe, E., Rogers, K., Hassenstab, J., Brand, M., et al. (2006). Introducing MASC: a movie for the assessment of social cognition. *J. Autism Dev. Disord.* 36, 623–636. doi: 10.1007/s10803-006-0107-0

Eckert, H. (2017). Erzeugung von Blickreizen virtueller Charaktere mit ambiger kommunikativer Absicht mittels systematischer Variierung zweier Faktoren einer Blickbewegung - Anfangsblick und Blickziel (Medical dissertation, University of Cologne). University of Cologne.

Einav, S., and Hood, B. M. (2006). Children's use of the temporal dimension of gaze for inferring preference. *Dev. Psychol.* 42, 142–152. doi: 10.1037/0012-1649.42.1.142

Fedor, J., Lynn, A., Foran, W., DiCicco-Bloom, J., Luna, B., and O'Hearn, K. (2018). Patterns of fixation during face recognition: differences in autism across age. *Autism* 22, 866–880. doi: 10.1177/1362361317714989

Ferreira, F., Apel, J., and Henderson, J. M. (2008). Taking a new look at looking at nothing. *Trends Cogn. Sci.* 12, 405–410. doi: 10.1016/j.tics.2008.07.007

Féry, C., and Kügler, F. (2008). Pitch accent scaling on given, new and focused constituents in German. *J. Phon.* 36, 680–703. doi: 10.1016/j.wocn.2008.05.001

FFmpeg Developers (2018). FFmpeg Tool [Software]. Available at: http://ffmpeg.org/

Fletcher-Watson, S., Leekam, S. R., Benson, V., Frank, M. C., and Findlay, J. M. (2009). Eye-movements reveal attention to social information in autism spectrum disorder. *Neuropsychologia* 47, 248–257. doi: 10.1016/j.neuropsychologia.2008.07.016 (Accessed June 15, 2019).

Fraundorf, S. H., Watson, D. G., and Benjamin, A. S. (2010). Recognition memory reveals just how CONTRASTIVE contrastive accenting really is. *J. Mem. Lang.* 63, 367–386. doi: 10.1016/j.jml.2010.06.004

Fraundorf, S. H., Watson, D. G., and Benjamin, A. S. (2012). The effects of age on the strategic use of pitch accents in memory for discourse: a processing-resource account. *Psychol. Aging* 27, 88–98. doi: 10.1037/a0024138

Freeth, M., Chapman, P., Ropar, D., and Mitchell, P. (2010). Do gaze cues in complex scenes capture and direct the attention of high functioning adolescents with ASD? Evidence from eye-tracking. *J. Autism Dev. Disord.* 40, 534–547. doi: 10.1007/s10803-009-0893-2

Freire, A., Eskritt, M., and Lee, K. (2004). Are eyes windows to a Deceiver's soul? Children's use of Another's eye gaze cues in a deceptive situation. *Dev. Psychol.* 40, 1093–1104. doi: 10.1037/0012-1649.40.6.1093

Frith, C. D., and Frith, U. (2006). The neural basis of mentalizing. *Neuron* 50, 531–534. doi: 10.1016/j.neuron.2006.05.001

Frith, U., Morton, J., and Leslie, A. M. (1991). The cognitive basis of a biological disorder: autism. *Trends Neurosci.* 14, 433–438. doi: 10.1016/0166-2236(91)90041-R

Genzel, S., Kerkhoff, G., and Scheffter, S. (1995). PC-gestützte Standardisierung des Bildmaterials von Snodgrass & Vanderwart (1980): I. Deutschsprachige Normierung. *Neurolinguistik* 9, 41–53.

Gernsbacher, M. A., and Yergeau, M. (2019). Empirical failures of the claim that autistic people lack a theory of mind. *Arch. Sci. Psychol.* 7, 102–118. doi: 10.1037/arc0000067

Globerson, E., Amir, N., Kishon-Rabin, L., and Golan, O. (2015). Prosody recognition in adults with high-functioning autism spectrum disorders: from psychoacoustics to cognition. *Autism Res.* 8, 153–163. doi: 10.1002/aur.1432

Golan, O., Baron-Cohen, S., Hill, J. J., and Rutherford, M. D. (2007). The "Reading the mind in the voice" test-revised: a study of complex emotion recognition in adults with and without autism spectrum conditions. *J. Autism Dev. Disord.* 37, 1096–1106. doi: 10.1007/s10803-006-0252-5

Good, P. (2013). Permutation tests: A practical guide to resampling methods for testing hypotheses (springer series in statistics) - Good, Phillip: 9783540940975 - ZVAB. Berlin and Heidelberg: Springer-Verlag.

Grice, M., and Baumann, S. (2007). "An introduction to intonation – functions and models" in Non-native prosody (Berlin: De Gruyter Mouton), 25–52.

Grice, M., Baumann, S., and Benzmüller, R. (2005). German intonation in autosegmental-metrical phonology - Oxford scholarship. *Prosodic Typol.* 1, 55–83. doi: 10.1093/acprof:oso/9780199249633.003.0003

Grice, M., Krüger, M., and Vogeley, K. (2016). Adults with Asperger syndrome are less sensitive to intonation than control persons when listening to speech. *Cult. Brain* 4, 38–50. doi: 10.1007/s40167-016-0035-6

Grice, M., Ritter, S., Niemann, H., and Roettger, T. B. (2017). Integrating the discreteness and continuity of intonational categories. *J. Phon.* 64, 90–107. doi: 10.1016/j.wocn.2017.03.003

Grice, M., Wehrle, S., Krüger, M., Spaniol, M., Cangemi, F., and Vogeley, K. (2023). Linguistic prosody in autism spectrum disorder—an overview. *Lang. Linguist. Compass* 17:e12498. doi: 10.1111/lnc3.12498

Griffin, J. W., and Scherf, K. S. (2020). Does decreased visual attention to faces underlie difficulties interpreting eye gaze cues in autism? *Mol. Autism.* 11:60. doi: 10.1186/s13229-020-00361-2

Gronau, Q. F., Singmann, H., and Wagenmakers, E. (2020). Bridgesampling: an R package for estimating normalizing constants. *J. Stat. Softw.* 92, 1–29. doi: 10.18637/jss.v092.i10

Grossman, R. B., Zane, E., Mertens, J., and Mitchell, T. (2019). Facetime vs. Screentime: gaze patterns to live and video social stimuli in adolescents with ASD. *Sci. Rep.* 9:12643. doi: 10.1038/s41598-019-49039-7

Guillon, Q., Hadjikhani, N., Baduel, S., and Rogé, B. (2014). Visual social attention in autism spectrum disorder: insights from eye tracking studies. *Neurosci. Biobehav. Rev.* 42, 279–297. doi: 10.1016/j.neubiorev.2014.03.013

Happé, F. G. E. (1994). An advanced test of theory of mind: understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *J. Autism Dev. Disord.* 24, 129–154. doi: 10.1007/BF02172093

Henderson, J. M., Williams, C. C., and Falk, R. J. (2005). Eye movements are functional during face learning. *Mem. Cogn.* 33, 98–106. doi: 10.3758/BF03195300

Hernandez, N., Metzger, A., Magné, R., Bonnet-Brilhault, F., Roux, S., Barthelemy, C., et al. (2009). Exploration of core features of a human face by healthy and autistic adults analyzed by visual scanning. *Neuropsychologia* 47, 1004–1012. doi: 10.1016/j.neuropsychologia.2008.10.023

Hesling, I., Dilharreguy, B., Peppé, S., Amirault, M., Bouvard, M., and Allard, M. (2010). The integration of prosodic speech in high functioning autism: a preliminary fMRI study. *PLoS One* 5:e11571. doi: 10.1371/journal.pone.0011571

Hobson, R. P., Ouston, J., and Lee, A. (1988). What's in a face? The case of autism. *Br. J. Psychol.* 79, 441–453. doi: 10.1111/j.2044-8295.1988.tb02745.x

Hurley, R., and Bishop, J. (2016). Prosodic and individual influences on the interpretation of only. *Speech Prosody*. 8, 193–197. doi: 10.21437/SpeechProsody.2016-40

Itier, R. J., Villate, C., and Ryan, J. D. (2007). Eyes always attract attention but gaze orienting is task-dependent: evidence from eye movement monitoring. *Neuropsychologia* 45, 1019–1028. doi: 10.1016/j.neuropsychologia.2006.09.004

Ito, K., Kryszak, E., and Ibanez, T. (2022). Effect of prosodic emphasis on the processing of joint-attention cues in children with ASD. Lisbon, Portugal: ISCA. 110–114.

Ito, K., and Speer, S. R. (2008). Anticipatory effects of intonation: eye movements during instructed visual search. *J. Mem. Lang.* 58, 541–573. doi: 10.1016/j.jml.2007.06.013

Jording, M., Engemann, D., Eckert, H., Bente, G., and Vogeley, K. (2019a). Distinguishing social from private intentions through the passive observation of gaze cues. *Front. Hum. Neurosci.* 13:442. doi: 10.3389/fnhum.2019.00442

Jording, M., Hartz, A., Bente, G., Schulte-Rüther, M., and Vogeley, K. (2019b). Inferring interactivity from gaze patterns during triadic person-object-agent interactions. *Front. Psychol.* 10:1913. doi: 10.3389/fpsyg.2019.01913

Klami, A. (2010). "Inferring task-relevant image regions from gaze data" in 2010 IEEE international workshop on machine learning for signal processing, Eds. S.l. Kaski, D. J. Miller, E. Oja, A. Honkela (IEEE). 101–106.

Klami, A., Saunders, C., de Campos, T. E., and Kaski, S. (2008). "Can relevance of images be inferred from eye movements?" in Proceedings of the 1st ACM international conference on Multimedia information retrieval (New York, NY, USA: Association for Computing Machinery), 134–140.

Kleinman, J., Marciano, P. L., and Ault, R. L. (2001). Advanced theory of mind in high-functioning adults with autism. *J. Autism Dev. Disord.* 31, 29–36. doi: 10.1023/a:1005657512379

Klin, A., Jones, W., Schultz, R., Volkmar, F., and Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Arch. Gen. Psychiatry* 59, 809–816. doi: 10.1001/archpsyc.59.9.809

Konkle, T., Brady, T. F., Alvarez, G. A., and Oliva, A. (2010). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *J. Exp. Psychol. Gen.* 139, 558–578. doi: 10.1037/a0019165

Krahmer, E., Ruttkay, Z., Swerts, M., and Wesselink, W. (2002). "Perceptual evaluation of audiovisual cues for prominence" in INTERSPEECH. Denver, Colorado, USA: ISCA.

Kuzmanovic, B., Schilbach, L., Lehnhardt, F.-G., Bente, G., and Vogeley, K. (2011). A matter of words: impact of verbal and nonverbal information on impression formation in high-functioning autism. *Res. Autism Spectr. Disord.* 5, 604–613. doi: 10.1016/j.rasd.2010.07.005

Leding, J. K. (2020). Animacy and threat in recognition memory. *Mem. Cogn.* 48, 788–799. doi: 10.3758/s13421-020-01017-5

Lee, K., Eskritt, M., Symons, L. A., and Muir, D. (1998). Children's use of triadic eye gaze information for "mind reading.". *Dev. Psychol.* 34, 525–539. doi: 10.1037//0012-1649.34.3.525

Lee, M. D., and Wagenmakers, E.-J. (2014). Bayesian cognitive modeling: a practical course. Cambridge: Cambridge University Press.

Lenth, R. V., Bürkner, P., Herve, M., Love, J., Riebl, H., and Singmann, H. (2021). Emmeans: estimated marginal means, aka least-squares means. Available at: https://cran.r-project.org/web/packages/emmeans/index.html (Accessed February 3, 2023).

Macdonald, R. G., and Tatler, B. W. (2013). Do as eye say: gaze cueing and language in a real-world social interaction. *J. Vis.* 13:6. doi: 10.1167/13.4.6

Makowski, D., Ben-Shachar, M. S., and Lüdecke, D. (2019). bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *JOSS*, 4, 1541. doi: 10.21105/joss.01541

Martini, P., and Maljkovic, V. (2009). Short-term memory for pictures seen once or twice. *Vis. Res.* 49, 1657–1667. doi: 10.1016/j.visres.2009.04.007

Melcher, D. (2001). Persistence of visual memory for scenes. *Nature* 412, 401. doi: 10.1038/35086646

Melcher, D. (2006). Accumulation and persistence of memory for natural scenes. *J. Vis.* 6, 2–17. doi: 10.1167/6.1.2

Mixdorff, H., Hönemann, A., and Fagel, S. (2013). Integration of acoustic and visual cues in prominence perception. Proceedings of AVSP 2013. Available at: https://pub.uni-bielefeld.de/record/2752439 (Accessed August 16, 2023).

Moubayed, S., and Beskow, J. (2009). Effects of visual prominence cues on speech intelligibility. *Proc Int Conf Auditory Visual Speech Process.* (Norwich, UK) 9:16.

Müller, N., Baumeister, S., Dziobek, I., Banaschewski, T., and Poustka, L. (2016). Validation of the movie for the assessment of social cognition in adolescents with ASD: fixation duration and pupil dilation as predictors of performance. *J. Autism Dev. Disord.* 46, 2831–2844. doi: 10.1007/s10803-016-2828-z

Munhall, K. G., Jones, J., Callan, D., Kuratate, T., and Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility head movement improves auditory speech perception. *Psychol. Sci.* 15, 133–137. doi: 10.1111/j.0963-7214.2004.01502010.x

Nakano, T., Tanaka, K., Endo, Y., Yamane, Y., Yamamoto, T., Nakano, Y., et al. (2010). Atypical gaze patterns in children and adults with autism spectrum disorders dissociated from developmental changes in gaze behaviour. *Proc. Biol. Sci.* 277, 2935–2943. doi: 10.1098/rspb.2010.0587

O'Connor, K. (2012). Auditory processing in autism spectrum disorder: a review. *Neurosci. Biobehav. Rev.* 36, 836–854. doi: 10.1016/j.neubiorev.2011.11.008

O'Hearn, K., Tanaka, J., Lynn, A., Fedor, J., Minshew, N., and Luna, B. (2014). Developmental plateau in visual object processing from adolescence to adulthood in autism. *Brain Cogn.* 90, 124–134. doi: 10.1016/j.bandc.2014.06.004

Odén, A., and Wedel, H. (1975). Arguments for Fisher's permutation test. *Ann. Stat.* 3, 518–520. doi: 10.1214/aos/1176343082

Pantelis, P. C., and Kennedy, D. P. (2017). Deconstructing atypical eye gaze perception in autism spectrum disorder. *Sci. Rep.* 7:14990. doi: 10.1038/s41598-017-14919-3

Pelphrey, K. A., Sasson, N. J., Reznick, J. S., Paul, G., Goldman, B. D., and Piven, J. (2002). Visual scanning of faces in autism. *J. Autism Dev. Disord.* 32, 249–261. doi: 10.1023/A:1016374617369

Pesarin, F., and Salmaso, L. (2010). The permutation testing approach: a review. *Statistica* 70, 481–509. doi: 10.6092/issn.1973-2201/3599

Pfeiffer-Lessmann, N., Pfeiffer, T., and Wachsmuth, I. (2012). An operational model of joint attention - timing of gaze patterns in interactions between humans and a virtual human., in Proceedings of the 34th annual conference of the Cognitive Science Society, 851–856.

Ponnet, K. S., Roeyers, H., Buysse, A., De Clercq, A., and Van der Heyden, E. (2004). Advanced mind-reading in adults with Asperger syndrome. *Autism* 8, 249–266. doi: 10.1177/1362361304045214

Prieto, P., Puglesi, C., Borràs-Comes, J., Arroyo, E., and Blat, J. (2015). Exploring the contribution of prosody and gesture to the perception of focus using an animated agent☆. *J. Phon.* 49, 41–54. doi: 10.1016/j.wocn.2014.10.005

R Core Team (2019). R: A language and environment for statistical computing. Available at: https://www.R-project.org (Accessed May 30, 2019).

R Core Team (2023). R: A Language and Environment for Statistical Computing. Available at: https://www.R-project.org/ (Accessed April 10, 2023).

Remington, A., and Fairnie, J. (2017). A sound advantage: increased auditory capacity in autism. *Cognition* 166, 459–465. doi: 10.1016/j.cognition.2017.04.002

Riby, D. M., Hancock, P. J., Jones, N., and Hanley, M. (2013). Spontaneous and cued gaze-following in autism and Williams syndrome. *J. Neurodevelop. Disord.* 5:13. doi: 10.1186/1866-1955-5-13

Riddiford, J. A., Enticott, P. G., Lavale, A., and Gurvich, C. (2022). Gaze and social functioning associations in autism spectrum disorder: a systematic review and meta-analysis. *Autism Res.* 15, 1380–1446. doi: 10.1002/aur.2729

Ring, M., Gaigg, S. B., and Bowler, D. M. (2015). Object-location memory in adults with autism spectrum disorder. *Autism Res.* 8, 609–619. doi: 10.1002/aur.1478

Ristic, J., Mottron, L., Friesen, C. K., Iarocci, G., Burack, J. A., and Kingstone, A. (2005). Eyes are special but not for everyone: the case of autism. *Cogn. Brain Res.* 24, 715–718. doi: 10.1016/j.cogbrainres.2005.02.007

Rosenblau, G., Kliemann, D., Dziobek, I., and Heekeren, H. R. (2017). Emotional prosody processing in autism spectrum disorder. *Soc. Cogn. Affect. Neurosci.* 2, 224–239. doi: 10.1093/scan/nsw118

Rossion, B., and Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: the role of surface detail in basic-level object recognition. *Perception* 33, 217–236. doi: 10.1068/p5117

Roy, J., Cole, J., and Mahrt, T. (2017). Individual differences and patterns of convergence in prosody perception. *Lab. Phonol.* 8:22. doi: 10.5334/labphon.108

RStudio Team (2016). RStudio: Integrated Development for R. Available at: http://www.rstudio.com (Accessed April 25, 2019).

Rutherford, M. D., Baron-Cohen, S., and Wheelwright, S. (2002). Reading the mind in the voice: a study with normal adults and adults with Asperger syndrome and high functioning autism. *J. Autism Dev. Disord.* 32, 189–194. doi: 10.1023/a:1015497629971

Scheeren, A. M., Rosnay, M.De, Koot, H. M., and Begeer, S. (2013). Rethinking theory of mind in high-functioning autism spectrum disorder. *J. Child Psychol. Psychiatry* 54, 628–635. doi: 10.1111/jcpp.12007

Schelinski, S., Roswandowitz, C., and von Kriegstein, K. (2017). Voice identity processing in autism spectrum disorder. *Autism Res.* 10, 155–168. doi: 10.1002/aur.1639

Schelinski, S., and von Kriegstein, K. (2019). The relation between vocal pitch and vocal emotion recognition abilities in people with autism Spectrum disorder and typical development. *J. Autism Dev. Disord.* 49, 68–82. doi: 10.1007/s10803-018-3681-z

Schneider, D., Slaughter, V. P., Bayliss, A. P., and Dux, P. E. (2013). A temporally sustained implicit theory of mind deficit in autism spectrum disorders. *Cognition* 129, 410–417. doi: 10.1016/j.cognition.2013.08.004

Schuwerk, T., Vuori, M., and Sodian, B. (2015). Implicit and explicit theory of mind reasoning in autism spectrum disorders: the impact of experience. *Autism* 19, 459–468. doi: 10.1177/1362361314526004

Senju, A., Southgate, V., White, S., and Frith, U. (2009). Mindblind eyes: an absence of spontaneous theory of mind in Asperger syndrome. *Science* 325, 883–885. doi: 10.1126/science.1176170

Setien-Ramos, I., Lugo-Marín, J., Gisbert-Gustemps, L., Díez-Villoria, E., Magán-Maganto, M., Canal-Bedia, R., et al. (2022). Eye-tracking studies in adults with autism Spectrum disorder: a systematic review and Meta-analysis. *J. Autism Dev. Disord.* 53, 2430–2443. doi: 10.1007/s10803-022-05524-z

Shimojo, S., Simion, C., Shimojo, E., and Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nat. Neurosci.* 6, 1317–1322. doi: 10.1038/nn1150

Stewart, M. E., McAdam, C., Ota, M., Peppé, S., and Cleland, J. (2013). Emotional recognition in autism spectrum conditions from voices and faces. *Autism* 17, 6–14. doi: 10.1177/1362361311424572

Swerts, M., and Krahmer, E. (2008). Facial expression and prosodic prominence: effects of modality and facial area. *J. Phon.* 36, 219–238. doi: 10.1016/j.wocn.2007.05.001

Theeuwes, J., Belopolsky, A., and Olivers, C. N. L. (2009). Interactions between working memory, attention and eye movements. *Acta Psychol.* 132, 106–114. doi: 10.1016/j.actpsy.2009.01.005

Tottenham, N., Hertzig, M. E., Gillespie-Lynch, K., Gilhooly, T., Millner, A. J., and Casey, B. J. (2014). Elevated amygdala response to faces and gaze aversion in autism spectrum disorder. *Soc. Cogn. Affect. Neurosci.* 9, 106–117. doi: 10.1093/scan/nst050

Wahl, S., Marinović, V., and Träuble, B. (2019). Gaze cues of isolated eyes facilitate the encoding and further processing of objects in 4-month-old infants. *Dev. Cogn. Neurosci.* 36:100621. doi: 10.1016/j.dcn.2019.100621

Wang, S., Jiang, M., Duchesne, X. M., Laugeson, E. A., Kennedy, D. P., Adolphs, R., et al. (2015). Atypical visual saliency in autism Spectrum disorder quantified through model-based eye tracking. *Neuron* 88, 604–616. doi: 10.1016/j.neuron.2015.09.042

Wang, L., Pfordresher, P. Q., Jiang, C., and Liu, F. (2021). Individuals with autism spectrum disorder are impaired in absolute but not relative pitch and duration matching in speech and song imitation. *Autism Res.* 14, 2355–2372. doi: 10.1002/aur.2569

Watson, D. G., Tanenhaus, M. K., and Gunlogson, C. A. (2008). Interpreting pitch accents in online comprehension: H* vs. L+H*. *Cogn. Sci.* 32, 1232–1244. doi: 10.1080/03640210802138755

Weber, A., Braun, B., and Crocker, M. W. (2006). Finding referents in time: eye-tracking evidence for the role of contrastive accents. *Lang. Speech* 49, 367–392. doi: 10.1177/00238309060490030301

White, S. J., Coniston, D., Rogers, R., and Frith, U. (2011). Developing the Frith-Happé animations: a quick and objective test of theory of mind for adults with autism. *Autism Res.* 4, 149–154. doi: 10.1002/aur.174

Winn, M. (2014). Fade in, Fade out [Praat script]. Available at: http://www.mattwinn.com/praat/RampOnsetAndOrOffset.txt (Accessed April 13, 2020).

World Medical Association (2013). World medical association declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* 310, 2191–2194. doi: 10.1001/jama.2013.281053

Yarbus, A. L. (1967). "Eye movements during perception of complex objects" in Eye movements and vision. ed. A. L. Yarbus (Boston, MA: Springer US), 171–211. doi: 10.1007/978-1-4899-5379-7_8

Zhang, M., Xu, S., Chen, Y., Lin, Y., Ding, H., and Zhang, Y. (2022). Recognition of affective prosody in autism spectrum conditions: a systematic review and meta-analysis. *Autism* 26, 798–813. doi: 10.1177/1362361321995725

Zimmermann, J. T., Wehrle, S., Cangemi, F., Grice, M., and Vogeley, K. (2020). Listeners and lookers: using pitch height and gaze duration for inferring mental states., in Proceedings of the 10th International Conference on Speech Prosody 2020, Tokyo, Japan.

5 Study 2

Zimmermann, J. T., Meuser, S., Hinterwimmer, S., & Vogeley, K. (2021). Preserved perspective taking in free indirect discourse in autism spectrum disorder. *Frontiers in Psychology*, *12*, Article 675633. https://doi.org/10.3389/fpsyg.2021.675633

Supplementary material for study 2 can be found in appendix 9.2.

frontiers in Psychology

ORIGINAL RESEARCH
published: 07 July 2021
doi: 10.3389/fpsyg.2021.675633

Preserved Perspective Taking in Free Indirect Discourse in Autism Spectrum Disorder

Juliane T. Zimmermann[1], Sara Meuser[2], Stefan Hinterwimmer[3] and Kai Vogeley[1,4]*

[1] Department of Psychiatry, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany, [2] Institute of Language and Literature I, University of Cologne, Cologne, Germany, [3] Institute of Language and Literature – Linguistics, University of Wuppertal, Wuppertal, Germany, [4] Institute of Neuroscience and Medicine, Cognitive Neuroscience (INM-3), Research Centre Juelich, Juelich, Germany

Perspective taking has been proposed to be impaired in persons with autism spectrum disorder (ASD), especially when implicit processing is required. In narrative texts, language perception and interpretation is fundamentally guided by taking the perspective of a narrator. We studied perspective taking in the linguistic domain of so-called Free Indirect Discourse (FID), during which certain text segments have to be interpreted as the thoughts or utterances of a protagonist without explicitly being marked as thought or speech representations of that protagonist (as in direct or indirect discourse). Crucially, the correct interpretation of text segments as *FID* depends on the ability to detect which of the protagonists "stands out" against the others and is therefore identifiable as implicit thinker or speaker. This so-called "prominence" status of a protagonist is based on linguistic properties (e.g., *grammatical function*, *referential expression*), in other words, the perspective is "hidden" and has to be inferred from the text material. In order to test whether this implicit perspective taking ability that is required for the interpretation of *FID* is preserved in persons with ASD, we presented short texts with three sentences to adults with and without ASD. In the last sentence, the perspective was switched either to the more or the less prominent of two protagonists. Participants were asked to rate the texts regarding their naturalness. Both diagnostic groups rated sentences with *FID* anchored to the less prominent protagonist as less natural than sentences with *FID* anchored to the more prominent protagonist. Our results that the high-level perspective taking ability in written language that is required for the interpretation of *FID* is well preserved in persons with ASD supports the conclusion that language skills are highly elaborated in ASD so that even the challenging attribution of utterances to protagonists is possible if they are only implicitly given. We discuss the implications in the context of claims of impaired perspective taking in ASD as well as with regard to the underlying processing of *FID*.

Keywords: autism spectrum disorder (ASD), perspective taking, free indirect discourse (FID), perspectival centers, mentalizing, theory of mind (ToM)

OPEN ACCESS

Edited by:
Antonio Benítez-Burraco, University of Seville, Spain

Reviewed by:
Mikhail Kissine, Université libre de Bruxelles, Belgium
Aparna Nadig, McGill University, Canada

****Correspondence:***
Juliane T. Zimmermann
juliane.zimmermann@uk-koeln.de

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *04 March 2021*
Accepted: *11 June 2021*
Published: *07 July 2021*

Citation:
Zimmermann JT, Meuser S, Hinterwimmer S and Vogeley K (2021) Preserved Perspective Taking in Free Indirect Discourse in Autism Spectrum Disorder. Front. Psychol. 12:675633. doi: 10.3389/fpsyg.2021.675633

INTRODUCTION

One of two key symptoms of autism spectrum disorder (ASD) refers to social communication and interaction disturbances (American Psychiatric Association, 2013). One explanation for these phenomena is an impaired ability to take the perspective of others (Baron-Cohen et al., 1985; Frith et al., 1991), also referred to as *Theory of Mind* (*ToM*; Premack and Woodruff, 1978) or *mentalizing* (Fonagy et al., 2004; Frith and Frith, 2006). This impairment has often been demonstrated in language-based tasks with children with ASD[1] (Baron-Cohen et al., 1985; Baron-Cohen, 1989; Leslie and Thaiss, 1992; Swettenham, 1996; Hutchins et al., 2012; Begeer et al., 2014). Adolescents or adults with ASD and normal intelligence usually pass comparable false-belief tasks designed to probe second-order *ToM* tasks as successfully as control participants (Bowler, 1992; Happé, 1994; Ponnet et al., 2004; Scheeren et al., 2013). In these tasks, participants are prompted with explicit questions regarding the mental state of a protagonist. These tasks probe an explicit and, hence, better accessible type of perspective taking. On the other hand, tasks that require a more implicit type of perspective taking appear to be problematic for adolescents or adults with ASD, even under conditions of normal intelligence. This is especially revealed when participants are asked to not only infer a protagonist's mental state, but also to provide reasons for their attributions (Ozonoff et al., 1991; Bowler, 1992; Happé, 1994; Dziobek et al., 2006; Callenmark et al., 2014), similarly, when eye movement is measured to assess overt attention in false-belief tasks (Senju et al., 2009; Schneider et al., 2013; Schuwerk et al., 2015). Impairments are also visible when inferring a protagonist's mental state based on photo or video material (Baron-Cohen et al., 2001; Ponnet et al., 2004; Dziobek et al., 2006), which might explain why participants with ASD rely in their impressions formation of others significantly more on verbal than on non-verbal information (Kuzmanovic et al., 2011).

The interpretation of an utterance does not only depend on the linguistic content and its context, but also and essentially on the person of the speaker. An utterance of a sentence containing a so-called predicate of personal taste (e.g., "licorice is tasty"; Lasersohn (2005)) might be true for one, but not for another person. Furthermore, utterances including deictic expressions referring to persons ("I", "you"), places ("here," "there") and/or time ("now", "then") can only be successfully interpreted in their context (i.e., speaker, reader/listener, location, time). In contrast to spoken language, written text does not always allow for an unambiguous identification of the speaker or perspectival center. It has been proposed (Zeman, 2017) that processing of so-called *Free Indirect Discourse* (*FID*; Banfield (1982)) shares an important aspect with perspective taking involved in *ToM* as operationalized in many false belief tasks, namely the ability to identify and differentiate between separate viewpoints at the same time. Importantly, we believe that *FID* processing differs from false-belief tasks insofar as perspective taking in FID is implicit. While in false-belief tasks commonly mastered by adults with ASD and normal intelligence the instruction to take a perspective is explicit, in *FID* it is implicit as readers are not instructed to take the perspective of a certain protagonist, but rather switch perspectives automatically in order to reach a sensible interpretation. Harris and Potts (2009) showed that certain context-sensitive markers have the potential to alter text interpretation so that perspective is shifted away from the first-person narrator to a competing protagonist. Kaiser (2015) demonstrated that *FID* cues increase perspectival-center-oriented text interpretation. However, these studies do not consider contexts in which multiple protagonists can serve as potential anchors for the utterance in *FID* mode.

In *FID*, utterances or thoughts are to be ascribed to a protagonist without explicitly mentioning her/him as the source of the utterance or thought. In the following example: "When Thomas entered the pub a guy in a black coat punched him right in the face with his bare hand. Ouch, how that hurt!" the reader will most likely understand that the last sentence expresses the experience of Thomas, whereas it is much less likely that the punching guy complains about his hand hurting. Without any explicit linguistic markers (e.g., quotation marks), *FID* is commonly indicated by the use of more subtle signals (Banfield, 1982; Steube, 1985), such as an exclamative ("Ouch!") or a judgmental statement ("that hurt"). Often, *FID* can only be interpreted correctly when certain parts of the sentence such as deictic adverbials of space and time or expressions such as "Ouch" are anchored to the protagonist's perspective (e.g., it is Thomas who feels pain, not the narrator) while others such as pronouns and tenses are anchored to the narrator's perspective (e.g., for Thomas, being punched does not hurt in the past, but in the present) (Schlenker, 2004; Eckardt, 2014). In other words, the interpretation of *FID* requires the identification of the implicit anchor for a specific thought or utterance and, hence, taking the perspective of one protagonist as opposed to another (Example *1*).

(A) *On Monday morning Jaqueline was running to the classroom in a hurry. In the hallway she bumped into her classmate*$_{m}$. *Now she would have to go to the nurse with that clumsy oaf.*
(B) *On Monday morning Arne was running to the classroom in a hurry. In the hallway he bumped into his classmate*$_{f}$. *Now she would have to go to the nurse with that clumsy oaf.*
(C) *On Monday morning Arne was running to the classroom in a hurry. In the hallway he bumped into his classmate*$_{f}$. *She went to the nurse with him.*
(D) *On Monday morning Arne was running to the classroom in a hurry. On the hallway he bumped into his classmate*$_{f}$. *He went to the nurse with her.*

Example 1: *One variation of a scenario as it appeared in our study in the four different conditions A, B, C, and D. The last sentence of item A and B is an instance of FID that needs to be anchored to one of the two protagonists of the preceding sentences to be interpreted sensibly. Items C and D do not*

[1]The use of "person-first" terminology in the context of ASD is controversial (Kenny et al., 2016; Vivanti, 2020). We apply a clinical perspective that focuses on common symptoms (or the absence thereof), which has been argued to be adequate depending on the context (Tepest, 2021).

contain FID. All texts were presented in German, followed the same structure and were similar in style. German words may denote a specific gender (e.g., classmate, German: "Klassenkameradin," or "Klassenkamerad"), indicated with "f" (female), and "m" (male).

In our study we follow a so-called prominence-based account for *FID* anchoring (Hinterwimmer, 2019), according to which the prominence status is the key for perspective ascription. Prominence refers to the property of a linguistic element (e.g., a syllable, a word, a sentence) as "standing out" in contrast to a group of similar elements (Streefkerk, 2002; Himmelmann and Primus, 2015). The protagonist who is more prominent in terms of *grammatical function* and *type of referential expression* (i.e., the expression we use to refer to an object or a person, e.g., "Thomas", "he") is more plausible as the anchor for *FID* than a competing protagonist (Hinterwimmer and Meuser, 2019). Based on the assumption that *FID* anchoring requires implicit perspective taking, these findings indicate that *FID* anchored to the more prominent protagonist is perceived as more natural and therefore receives higher ratings on a scale indicating acceptability by test persons, because it is easier or more common to take the prominent protagonist's perspective. For the purpose of our study we systematically varied *grammatical function* and *type of referential expression* as influential factors for a protagonist's prominence status. In the hierarchy of *grammatical functions,* a subject is more prominent than an indirect object, which is in turn more prominent than a direct object and so forth (Himmelmann and Primus, 2015). With respect to *referential expression* a protagonist that is familiar to the reader is more prominent than a protagonist that is unfamiliar (Jasinskaja et al., 2015). We make use of these prominence-lending features by claiming that a protagonist who is introduced with her/his first name and picked up by a pronoun in subject position is easier identified as the perspectival center for *FID* ascription than a competing protagonist who is introduced with an indefinite noun phrase in object position, which was already shown to be the case in an acceptability rating study by Hinterwimmer and Meuser (2019).

So far, it has not been clarified which particular linguistic types of perspective taking are consistently affected in exactly what way in ASD during speech and language production and perception, especially with regard to the shifting of perspectival centers. While *FID* perception has not been investigated in ASD so far, the production and perception of *referential expressions* has been studied already. While people with ASD and normal intelligence perform well in verbal perspective taking tasks, subtle differences indicate problems with respect to *ToM* in language production. The general population tends to adjust their choice of *referential expressions* to the listener or reader (i.e., depending on the context, we choose to substitute names with pronouns; Achim et al. (2017)). Adults with ASD use more full noun phrases during narratives when they could use pronouns instead, while, on the other hand, they use more pronouns when full noun phrases would be less ambiguous and hence would make it easier to understand the narration (Colle et al., 2008). This finding could indicate a reduced *ToM* in ASD with regard to the listener (Colle et al., 2008). This behavior has, however, not consistently been reported (Arnold et al., 2009). In a perception study investigating spatial perspective taking, participants with ASD showed unimpaired performance and neural activation comparable to a control group during the perception of written text referring to two people by their first names in third person, namely the participant and another person. On the other hand, when the task required perspective shifts induced by references to the participant as "you", performance decreased and neural patterns differed compared to the control group (Mizuno et al., 2011).

In our web-based study, we investigate for the first time the perception of shifting perspectival centers by means of *FID* in written language in adults with ASD. This implicit form of perspective taking might not be as easily accomplished by adults with ASD as by adults without ASD. Therefore, we expected to identify difficulties in *FID* processing in persons with ASD. In our study, participants judged the naturalness of sentences including *FID* anchored to protagonists of different prominence status. Based on the idea that texts in which the required perspective taking is easier to accomplish are linked to higher naturalness ratings, and considering the reported perspective taking difficulties in people with ASD in implicit *ToM* tasks, we anticipated lower naturalness ratings in people with ASD for texts associated with implicit perspective taking, especially if the required perspective shift is an unusual one. More specifically, we pursued the following hypotheses:

H1: The difference between naturalness ratings for texts including *FID* (here: condition A) and ratings for texts not including *FID* (here: condition D) will be greater in the ASD group in comparison to the control group.

H2: The difference between naturalness ratings for texts including *FID* anchored to the less prominent protagonist (here: condition B) and ratings for texts including *FID* anchored to the more prominent protagonist (here: condition A) will be greater in the ASD group in comparison to the control group. If H1 is supported, differences between ratings for condition A and B might play a minor role.

MATERIALS AND METHODS

Participants

Only participants who were monolingual native speakers of German were included in the study. For the ASD group, we recruited 45 adults with ASD via a mailing list of the Outpatient Clinic for Autism in adulthood at the University Hospital of Cologne. Of these, 41 participants had a diagnosis of Asperger syndrome (F.84.5 according to ICD-10), four participants indicated a diagnosis of high-functioning autism, one of these a diagnosis of childhood autism (F.84.0). For the control group, we recruited 45 adults without a diagnosis of ASD via the intranet of the University Hospital Cologne, publicly

TABLE 1 | Sample characteristics.

	Gender	Age	WST	BDI-II	AQ	EQ
ASD (*N* = 45)	25 men 20 women	20 - 82 years men: *M* = 48.2 (*SD* = 13.9) women: *M* = 42.6 (*SD* = 10.9)	*M* = 112.3 (*SD* = 10.00)	M = 13.8 (*SD* = 9.30)	*M* = 42.5 (*SD* = 4.25)	*M* = 13.8 (*SD* = 5.95)
Control (*N* = 45)	25 men 20 women	20 - 80 years men: *M* = 47.7 (*SD* = 14.7) women: *M* = 41.0 (*SD* = 12.3)	*M* = 111.0 (*SD* = 9.35)	M = 8.2 (*SD* = 6.27)	*M* = 15.5 (*SD* = 6.60)	*M* = 47.0 (*SD* = 12.5)

accessible notice boards and personal contacts (**Table 1** for sample characteristics).

In the group of participants with ASD, 25 of 45 participants with ASD reported that they had experienced depressive episodes. Participants with ASD indicated the following medication for the treatment of psychological, psychiatric and neurological conditions: antidepressants (15 participants), mood stabilizer (1), neuroleptic medication (1). Control participants indicated no history of neurological or psychiatric disorders. No psychotropic medication was reported by any participant in the control group. Scores for verbal intelligence as measured with the *Wortschatztest* (*WST*, Schmidt and Metzler (1992)) indicated average or above-average verbal intelligence in all participants (**Table 1**) and did not differ between groups (two-samples *t*-test, $t(88) = -0.63$, $p = 0.530$). Depressive symptoms measured with the *Beck depression inventory II* (*BDI-II*, Beck et al. (1996)) were significantly higher in participants with ASD than in control participants (**Table 1**, Welch two-samples *t*-test, $t(77.1) = -3.34$, $p = 0.001$), with symptoms ranging from none to clinically relevant symptoms in both groups. Scores indicating autistic traits measured with the *autism quotient* (*AQ*, Baron-Cohen et al. (2001)) were significantly higher in participants with ASD compared to the control group (**Table 1**, Welch two-samples *t*-test, $t(75.2) = -23.08$, $p < 0.001$). Scores indicating empathetic traits measured with the *empathy quotient* (*EQ*, Baron-Cohen and Wheelwright (2004)) were significantly lower in participants with ASD compared to the control group (**Table 1**, Welch two-samples *t*-test, $t(63.1) = 16.15$, $p < 0.001$).

Text Material

We presented short German narrative texts with three sentences each. We developed 24 different scenarios with a common theme. Each scenario was varied systematically in four different conditions, resulting in a total of 96 different texts. The conditions varied with respect to utterances with *FID* (conditions A and B; see example *1*) or without *FID* in neutral story continuation (conditions C and D; see example *1*. See **Table 2** for an overview of experimental conditions and the **Supplementary Material** for the complete list of texts). The content of the utterance with *FID* was thematically ambiguous with respect to two protagonists that were both potential candidates for the perspectival center, i.e., the thought presented as *FID* in the last sentence of the text could plausibly be linked to either one of the two protagonists, if the pronoun in the third sentence did not allow for unambiguous resolution. The utterance with *FID* thus varied with respect to the pronoun that indicated which one of the two protagonists was the anchor of the thought.

TABLE 2 | Overview of experimental conditions; "P" stands for "protagonist".

Condition	Subject in S1	Subject in S2	Subject/ Perspective in S3
A: *FID*, prominent	P1	P1	P1
B: *FID*, non-prominent	P2	P2	P1
C: Control, subject change	P2	P2	P1
D: Control, no subject change	P2	P2	P2

In the first sentence (S1) of each text, one of two protagonists was introduced by a proper name in subject position, and an explicit reference to the past (e.g., "Monday morning") was included. In the second sentence (S2) the protagonist introduced in S1 was picked up with a personal pronoun in subject position interacting with a second protagonist who was referred to with a full noun phrase and who was anchored to the first protagonist with a possessive pronoun (e.g., "her/his classmate"). Contrary to the English equivalent, the German noun phrases used in our stimuli were each linked to a specific gender (female/male). Therefore, both protagonists (P1 and P2) differed with regard to gender so that the *FID* in S3 could only reasonably be anchored to either P1 or P2.

The target sentence (S3) in condition A and condition B was an utterance in *FID* mode. It featured three indicators of *FID*: (i) a temporal adverbial referring to the present (e.g., "now," "today") or an immediate or close future (e.g., "soon," "tomorrow") contrasting with the temporal adverbials in S1, (ii) a verb in subjunctive II mode (e.g., "would"), and (iii) a colloquial term or qualitative noun (e.g. "clumsy oaf"). Conditions C and D served as control conditions. Unlike the target sentence S3 in *FID* conditions, S3 in control conditions did not feature any markers of *FID*. The target sentence continued the story in neutral narrative story mode. Control condition D continued with P1 in subject position while in condition C, P2 was the subject. Thus, the two neutral conditions resembled the test conditions regarding content and syntactic structure.

In order to investigate the anchoring of *FID* we manipulated our texts with regard to the two protagonists in three different ways, with respect to (i) the grammatical function of the first expression referring to them (subject or object),

(ii) the number of references (two or three), and iii) the type of referring expression (first name and pronouns or noun phrase and colloquial term). Based on previous findings (Hinterwimmer and Meuser, 2019) we predicted that in control participants *FID* anchored to the more prominent protagonist, i.e., the one in subject position, referred to with their first name and picked up by an adequate personal pronoun (condition A), would more likely be accepted as the perspectival center of a sentence in *FID* than the competing protagonist who was introduced with a noun phrase in object position in the second sentence (condition B). Texts in condition A should thus be rated more natural than texts in condition B.

As our manipulation of the utterance in *FID* mode involved a change or continuation of the subject with respect to one of the two protagonists, we included two control conditions C and D to account for the effect of subject change based on differences in referential chains: In condition C, the pronoun in subject position of the final sentence picked up the object of the preceding sentence, while in condition D, it picked up the subject. If texts of condition C would be rated comparable to texts of condition D, we might conclude that differences between the two *FID* conditions cannot be explained by (dis)continuity of referential chains alone. As both story continuations were equally coherent in terms of content, both control conditions C and D should be equally acceptable.

We included 40 filler texts similar to the 96 target texts in length and complexity (see **Supplementary Material**). In order to mask our manipulation, some filler texts were deliberately designed to yield low acceptability by an odd choice of pronouns, i.e., in the last sentence, a personal pronoun was used which referred back to an inanimate entity that occurred in object position in the previous sentence in which a personal pronoun was used to refer to the protagonist ("[. . .] He ate the cake_{m}. He was made of marzipan."). All four conditions were equally distributed across four lists so that every participant was presented with only one condition (A, B, C, or D) of each of the 24 scenarios and the total set of 40 filler texts, resulting in 64 texts in total that were presented to each participant in random order.

Procedure

The experiment was programmed and presented on Ibex farm, a platform for online experiments (Drummond, 2020). It was conducted in accordance with the Declaration of Helsinki (World Medical Association, 2013) and approved by the Ethics Committee of the Medical Faculty of the University of Cologne. Informed consent was obtained from all participants prior to participation. Demographic data and information on clinical diagnoses and medication was collected. In the following rating, participants were instructed to judge the naturalness of the third sentence in the context of the first two sentences of each presented text on a scale from 1 (labeled "very unnatural") to 7 (labeled "very natural"). For each text, presentation duration including response time was limited to 25 seconds. After completing the naturalness ratings, participants were given the opportunity to report what they noticed about the task in an open format. Psychological questionnaires were obtained afterward: *WST, AQ, EQ, BDI-II*. The *BDI-II* was included due to the high incidence of depressive symptoms in persons with ASD (Ghaziuddin et al., 2002). Finally, participants had the opportunity to make assumptions with regard to the aims of the study. The whole procedure engaged participants for approximately one hour. They were debriefed and compensated for their participation with a gift voucher of ten Euro.

Analysis

Data was analyzed using R (R Core Team, 2019) in RStudio (RStudio Team, 2016). We fitted Bayesian ordinal models using the *brms* package (Bayesian Regression Models using Stan, v2.10.0; Bürkner (2017); Bürkner and Vuorre (2019)). Factors were sum-coded. Weakly informative priors were used for group-level effects as well as for random intercepts (normal distribution; mean = 0; standard deviation = 2) and fixed intercepts (normal distribution; mean = 4, i.e., the center of the rating scale; standard deviation = 2). Estimated parameters are reported in terms of posterior means and 95% credibility intervals. To investigate the evidence for or against the investigated effects, we compared models by calculating Bayes factors applying the *bayesfactor_models* function from the *bayestestR* package (Makowski et al., 2019) which uses bridge sampling (Gronau et al., 2020). All models ran with four sampling chains of 12,000 iterations each including a warm-up period of 2,000 iterations.

Models

To test hypothesis 1 and thus the influence of *FID* and diagnosis, i.e., to identify differences between the groups regarding naturalness ratings for texts with *FID* and ratings for comparable texts without *FID*, a Bayesian ordinal mixed model was fitted to the ratings from conditions A and D. Fixed effects used in the model were *FID*, *group* and their interaction. Additionally, we included random intercepts and slopes for the factor *subject* as well as random intercepts for *text*. To test hypothesis 2 and thus the influence of protagonist prominence and diagnosis, i.e., to identify differences between the groups regarding naturalness ratings for texts with *FID* anchored to the more prominent protagonist and ratings for texts with *FID* anchored to the less prominent protagonist, a Bayesian ordinal mixed model was fitted to the data of the acceptability ratings for conditions A and B. Fixed effects used in the model were *prominence*, *group* and their interaction. Additionally, we included random intercepts and slopes for the factor *subject* as well as random intercepts for *text*. To demonstrate that a subject shift toward the less prominent protagonist does not in general lead to lower ratings, but only in *FID* conditions, we ran a Bayesian ordinal mixed model for naturalness ratings of our control conditions that did not include *FID*, i.e., neutral condition C including a subject shift toward the less prominent protagonist and neutral condition D not including a subject shift. Fixed effects used in the model were *subject shift*, *group* and their interaction. Additionally, we included random intercepts and slopes for the factor *subject* as well as random intercepts for *text*. Because

texts in conditions C and D are minimally different, which is not the case for texts in conditions A and B, differences between C and D are not fully equivalent to differences in A and B. Thus, the resulting conditions do not allow to test our hypotheses in a single model. Therefore, we addressed our hypotheses in separate models. To investigate evidence for or against the presence of effects, we additionally ran the following models for comparison with each of these models: the respective null model not including the group level factors; the model including only one of either factor; and the model including the linear combination of both factors. We report respective Bayes factors of model comparisons and follow the interpretation by Jeffreys (1939).

Explorative Analyses

We carried out correlational explorative analyses to identify possible relationships between the naturalness ratings and parameters we collected in addition to the ratings, i.e., psycho(patho)logical measures and age. To account for individual rating behavior, we standardized the ratings for each participant applying a rank-based non-linear transformation to the ratings of all four conditions, which for each participant results in normally distributed rating values centered around zero. Influences due to individually different scale use are therefore minimized. We investigated correlations across and within the two groups for the difference between ratings for condition A and D with our parameters (i.e., *AQ, EQ, BDI-II, WST,* AQ-scores for the subscales *attention switching*, *communication* and *imagination, age*). Differences between conditions were calculated by subtracting the standardized ratings for condition D from the standardized ratings for condition A. Likewise, correlations were investigated between our parameters and the difference between the standardized ratings for conditions A and B. We report Pearson correlation coefficients or Spearman's rank correlation coefficients reaching significance at the 5% confidence-level.

RESULTS

In general, texts in conditions A and B including *FID* were rated less natural (condition A: M = 4.34, SD = 2.23; condition B: M = 2.85, SD = 1.91) than conditions C and D not including *FID* (condition C: M = 4.96, SD = 2.00; condition D: M = 5.17, SD = 2.04). Across the whole sample naturalness ratings for texts in condition A were higher than for texts in condition B. Ratings did not show any statistically meaningful difference between both diagnostic groups. See **Figure 1** for an overview of mean ratings per condition for both diagnostic groups.

Comparison of FID Condition A and Neutral Condition D

FID affected the ratings by lowering the units on the latent rating scale by 0.48 (95% CI = [−0.64, −0.32]). An ASD diagnosis showed a general tendency to lower the ratings (b = −0.18, 95% CI = [−0.43, 0.07]), however, the influence of a diagnosis on the ratings was smaller than that of *FID*. The interaction of *FID* and *group* hardly affected the ratings (b = −0.03, 95% CI = [−0.36, 0.29]). Model comparisons indicated extreme evidence only for an influence of *FID*. They further revealed moderate evidence for the absence of a *group* effect. Strong evidence was found against an interaction effect. Bayes factors for the models in comparison to the null model: full model: BF > 1000; model with linear combination: BF > 1000; model including only the factor *group*: BF = 0.16; model including only the factor *FID*: BF > 1000. We further investigated rating patterns in the two groups using the *marginal_effects* function from the *brms* package. The model results and the rating behavior within the two groups did not support the assumption of *FID* affecting rating behavior of the two groups differently. Thus, hypothesis 1 is not supported by our data.

Correlation analyses (see **Table 3**) showed that *age* was positively correlated with the difference between standardized

TABLE 3 | Correlations between psycho(pathological) measures and rating differences for compared conditions.

		AQ	EQ	BDI-II	WST	AQ attention switching	AQ communication	AQ imagination	age
A minus D (Difference bt. rank-based-standardized ratings)	Both groups	r_S = −0.07 (*p = 0.531*)	r_S = 0.02 (*p* = 0.836)	r_S = −0.08 (*p* = 0.480)	r_P = −0.13 (*p* = 0.217)	r_S = −0.08 (*p* = 0.431)	r_S = −0.04 (*p* = 0.710)	r_S = *−0.04* (*p* = 0.695)	r_S = 0.12 (*p* = 0.260)
	ASC	r_S = −0.19 (*p = 0.218*)	r_S = −0.11 (*p = 0.489*)	r_P = 0.02 (*p* = 0.889)	r_P = −0.24 (*p = 0.111*)	r_S = −0.06 (*p* = 0.708)	r_S = −0.15 (*p = 0.314*)	r_S = −0.16 (*p* = 0.295)	r_P = −0.17 (*p* = 0.277)
	Control	r_P = 0.18 (*p* = 0.224)	r_P = −0.04 (*p = 0.778*)	r_S = −0.03 (*p = 0.856*)	r_P = 0.00 (*p = 0.980*)	r_S = 0.03 (*p = 0.848*)	r_S = 0.28 (*p = 0.065*)	r_S = 0.23 (*p = 0.135*)	r_P = 0.39 (*p* = 0.008)*
A minus B (Difference bt. rank-based-standardized ratings)	Both groups	r_S = −0.10 (*p* = 0.327)	r_S = 0.05 (*p = 0.666*)	r_S = −0.01 (*p = 0.943*)	r_P = −0.09 (*p = 0.400*)	r_S = −0.23 (*p = 0.031*)*	r_S = −0.11 (*p* = 0.297)	$r_{S=}$ *−0.06* (*p* = 0.593)	r_S = 0.11 (*p = 0.302*)
	ASC	r_S = −0.06 (*p* = 0.686)	r_S = −0.15 (*p* = 0.327)	r_P = −0.09 (*p = 0.577*)	r_P = −0.16 (*p = 0.308*)	r_S = −0.35 (*p* = 0.018)*	r_S = −0.18 (*p* = 0.235)	r_S = 0.16 (*p = 0.298*)	r_P = −0.06 (*p = 0.709*)
	Control	r_P = *0.06* (*p* = 0.695)	r_P = −0.15 (*p = 0.320*)	r_S = 0.26 (*p* = 0.080)	r_S = 0.01 (*p* = 0.958)	r_S = 0.01 (*p* = 0.929)	r_S = 0.29 (*p* = 0.050)	r_S = 0.09 (*p* = 0.554)	r_P = 0.30 (*p* = 0.045)*

P-values <0.05 are marked with an asterisk.

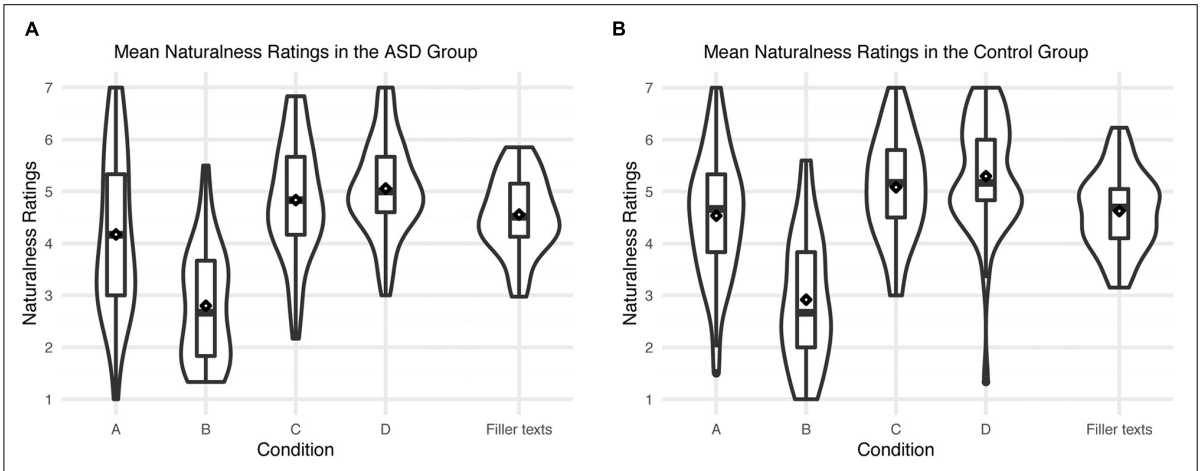

FIGURE 1 | Mean naturalness ratings for the four conditions and filler texts in the ASD group **(A)** and the control group **(B)**. Diamonds indicate means.

ratings for conditions A and D in the control group (r_P = 0.39, p = 0.008; ASD group: r_P = −0.17, p = 0.277). This indicates that the difference between ratings for sentences including *FID* and ratings for sentences without *FID* decreases with age in the control group.

Comparison of FID Conditions A and B

Reducing protagonist prominence generally affected the ratings by lowering the units on the latent rating scale by 0.85 (95% CI = [−1.06, −0.65]). ASD diagnosis lowered the ratings. However, this tendency was smaller than the effect of protagonist prominence (b = −0.12, 95% CI = [−0.37, 0.14]). The interaction showed that reduced *prominence* tended to result in higher ratings in the ASD group in comparison to the control group (b = 0.17, 95% CI = [−0.23, 0.57]). Model comparisons indicated extreme evidence for an influence of reduced protagonist prominence. They further revealed moderate evidence for an absence of a *group* effect as well as for an absence of an interaction effect (Bayes factors for the models in comparison to the null model: full model: BF > 1000; model with linear combination: BF > 1000; model including only the factor *group*: BF = 0.1; model including only the factor *prominence*: BF > 1000). The model results and the rating behavior within the two groups did not support the assumption of prominence affecting rating behavior of the two groups differently. Thus, hypothesis 2 is not supported by our data.

Comparable to the correlation analysis for the comparison of conditions A and D, correlation analyses showed that *age* was positively correlated with the difference between standardized ratings for conditions A and B in the control group only (r_P = 0.30, p = 0.045; ASD group: r_P = −0.06, p = 0.709). This indicates that the difference between ratings for *FID* anchored to the less prominent protagonist and ratings for *FID* anchored to the more prominent protagonist decreases with age in the control group. Moreover, correlations of our psycho(patho)logical measures with the difference between standardized ratings for conditions A and B showed a statistically significant correlation across the sample (r_S = −0.23, p = 0.031), which appears to mainly be driven by the sample with ASD: In this group, the scores of the *AQ* subscale *attention switching* were moderately negatively correlated with the difference between standardized ratings for conditions A and B (r_S = −0.35, p = 0.018). This indicates that participants with ASD reporting more problems regarding attention switching tend to give less divergent ratings for conditions A and B.

Comparison of Neutral Conditions C and D

The analysis suggests that a subject shift alongside the respective *referential expression* lowered the ratings (b = −0.16, 95% CI = [−0.30, −0.02]). The factor *group* showed a tendency to also lower the ratings (b = −0.15, 95% CI = [−0.39, 0.09]). The interaction hardly influenced the ratings (b = 0.03, 95% CI = [−0.25, 0.32]). Model comparisons, however, showed no reliable evidence for the presence of any of these effects and tendencies in our data, as indicated by Bayes factors favoring the null model over all other models while at the same time lacking robustness (Bayes factors for the models in comparison to the null model: full model: BF < 0.001; model with linear combination: BF = 0.01; model including only the factor *group*: BF = 0.13; model including only the factor *subject shift*: BF = 0.08).

Further Explorative Analyses

Visual inspection of naturalness ratings distributions suggested bimodality. To test if bimodality was present in our data, we tested for each ratings distribution in each condition in each group the deviance from unimodality. We used the R package *diptest* (v0.75-7; Mächler, 2015) which applies Hartigan's dip test (Hartigan and Hartigan, 1985). The results indicated that unimodality was not given at a 95%-confidence level in condition

C in the ASD group as well as in condition B and C in the control group. In the remaining conditions, unimodality was not given at a 90%-confidence level. Therefore, the visual impression was corroborated by the test. We performed a median split of the data to identify if there was a difference between people that tend to give higher ratings and people that tend to give lower ratings. To this end, we split the groups into two subgroups (high-rating subjects and low-rating subjects) based on their ratings in condition D, which we set as the reference condition for this analysis, because it does not contain *FID* nor a subject shift. We then ran the models already introduced above again with the additional factor *subgroup* (high-rating vs low-rating) along its interaction terms with the other factors.

The results of this analysis of subgroups showed that the negative effect of *FID* on the ratings in condition A as opposed to condition D seemed to be mediated mostly by participants who rated high in condition D. Most importantly, this pattern did not differ statistically in the two subgroups of both the ASD and the control group. Further, the results indicated that *FID* anchoring to the less prominent protagonist lead to lower ratings in all subgroups. Most importantly, this pattern did not differ statistically for the ASD subgroups and the control subgroups, indicating that the tendency for high or low ratings is more fundamental than the differential response behavior due to diagnostic groups.

Participants' Feedback

Most participants found the texts – at least to some degree – confusing, stylistically clumsy, illogical, and/or grammatically wrong. Several participants perceived a lack of coherence due to sudden subject or perspective shifts (supposedly in the case of neutral and *FID* texts) or due to the third sentence containing ambiguous reference (supposedly in the case of filler texts). Six participants (five with ASD) noticed and/or found the shifts of perspective in the third sentence confusing (supposedly with regard to *FID* texts), referring to this factor as "perspective shift", "shifting perspective" to the protagonist that the story was not about, "shift of (emotional) narrative perspective", "brutal shift of the narrative perspective," "illogical perspective," and "ambiguous perspective." The markers we used to indicate *FID* were partly perceived as unnatural, both by people with ASD and control participants. Not only markers of *FID* were mentioned in the feedback, but also our markers of prominence. Three participants with ASD reported problems with the interpretation of task instructions for the judgment of naturalness or a difficulty to integrate naturalness regarding the narrative style and naturalness regarding the content into a comprehensive rating of naturalness. Some participants felt torn between what to base their rating on, e.g., whether they should base their rating on what would be considered natural with regard to the behavior of the protagonists and the content of the story, or rather on whether this was a narrative form that could naturally be encountered. Three participants in the ASD group reported that they found it hard to make a decision within the time limit.

DISCUSSION

The aim of this study was to investigate the perception of *FID* and prominence in the context of *FID* in participants with ASD. In contrast to our hypotheses, we did not observe any difference in the performance between persons with ASD in comparison to unaffected control persons. The first focus related to hypothesis H1 was the acceptability of *FID* in the ASD group compared to the control group based on naturalness ratings of sentences including *FID* (condition A) as opposed to neutral sentences not containing *FID* (condition D). The second focus related to hypothesis H2 was the study of the difference between naturalness of *FID* anchored to a more prominent protagonist as opposed to a less prominent one (conditions A and B). Contrary to both hypotheses, the ratings were comparable and did not differ between the diagnostic groups, neither with respect to the presence of *FID* (conditions A vs. D, hypothesis H1) nor with respect to anchoring to more or less prominent protagonists (conditions A vs. B, hypothesis H2). Technically speaking, the factor *group* did not improve the adequacy of the statistical model. Taken together, both hypotheses had to be rejected.

FID Processing

Across the whole sample, naturalness ratings were lower for sentences with *FID* (conditions A and B) compared to sentences without *FID* (conditions C and D). This result is in accordance with findings of a previous study in the general population in which test items with *FID* received lower ratings in general. Additionally, in that study test items with *FID* anchored to the perspective of a more prominent protagonist yielded higher acceptability ratings than test items with *FID* anchored to the perspective of a less prominent protagonist (Hinterwimmer and Meuser, 2019). We could replicate this effect in our study, further supporting the notion of prominence as a relevant factor for anchoring *FID*. In our control analysis in neutral conditions, i.e., non-*FID* sentences, we showed that a subject shift as manipulated via *grammatical function* and *referential expression* shows a tendency, but not a reliable decrease, to lower acceptability ratings when pronouns need to be resolved. This indicates that the effect of *prominence* reported above cannot be explained by subject shift alone.

Interestingly, we found that the control sample as well as the ASD sample could both be divided into two subgroups with different rating tendencies. Persons who generated high ratings in condition D were more strongly affected by *FID,* whereas the effect of prominence for *FID* anchoring was comparable across subgroups. The explanation for these subgroups' behavior might be trivial: High-raters might tend to rate the acceptability of texts with *FID* worse compared to low-raters, because they have more rating variance available to indicate their perception. However, individual factors might also play a role such as different perspective taking abilities (Kaiser and Cohen, 2012) or language dexterity. Interestingly, this pattern was visible across the control and the ASD sample, which further underlines that rating patterns for *FID* in general and prominence-dependent *FID* anchoring in particular are not affected in ASD.

With respect to the processes involved, we propose that the anchoring of *FID* depends both on perspective taking as well as on linguistic markers, more specifically, on perspective taking and the ascription of the perspectival center of a text which in turn depends on the linguistic notion of prominence (Hinterwimmer, 2019). That leaves two strategies to anchor an utterance in *FID* mode which may be both involved: (i) the reader may ascribe an utterance in *FID* mode to the perspectival center of the text and/or (ii) they may ascribe an utterance in *FID* mode to a protagonist based on linguistic markers i.e., prominence-lending cues.

Influence of Age

Another interesting observation was the correlation of the ratings with age. In the control group, we report a relationship of *age* with the naturalness-ratings for sentences with *FID* as opposed to sentences without *FID*, in other words, both types of sentences are rated more similar with increasing *age*. The same relationship was found for *age* and the naturalness-ratings for *FID* anchored to the less prominent protagonist as opposed to *FID* anchored to the more prominent protagonist. This might be related to a cognitive decline that also involves language comprehension (Burke and Shafto, 2007) as well as referential processing such as in anaphor resolution based on problems recalling contextual information (Light and Capps, 1986). *FID* processing might be affected in older participants in a similar fashion, since it requires anchoring to a protagonist previously introduced in the context. Furthermore, tracking of protagonist prominence relations has been suggested to be affected in older adulthood (Hendriks et al., 2014). More generally, studies in older participants show that *ToM* abilities decrease with *age* across different experimental tasks (Henry et al., 2013).

In contrast to these aforementioned aspects that putatively explain the reduced *FID* sensitivity in older participants, greater linguistic experience could on the other hand allow for easier processing (Crocker and Keller, 2006) which could in turn lead to an increased acceptance of sentences in *FID* mode in older people, but also to easier processing of *FID* anchoring to less prominent protagonists as opposed to more prominent ones. Additionally, psycho-affective changes associated with higher *age* might play a role, such as a more positive mindset in general (Carstensen et al., 2010). Finally, *age*-associated cognitive decline affecting text processing may be compensated for by other abilities that improve with *age* such as crystallized abilities like vocabulary, or change with *age* such as allocation of attention during reading (Stine-Morrow et al., 2008).

Notably, we did not observe any such relationship with age in persons with ASD. Research on aging in people with ASD is sparse in general and often inconsistent (Happé and Charlton, 2012; Howlin and Magiati, 2017). While some cognitive abilities seem to decline in ASD similarly to the general population (Howlin and Magiati, 2017), others are less affected than in the general population, such as working memory (Lever et al., 2015) or align with control participants with age resulting in comparable abilities in both groups, such as *ToM* abilities (Lever and Geurts, 2016). Thus, different lifetime trajectories of cognitive abilities responsible for *FID* processing might explain the different rating behavior in ASD with increasing age.

Conceptual Issues

Theory of Mind (ToM)

One key capacity associated with perspective taking is *ToM*, the ability to ascribe mental states to oneself and others, also closely related to language (Gernsbacher and Yergeau, 2019). In adolescents or adults with ASD, language abilities can partly explain performance in *ToM* tasks (Peterson and Miller, 2012; Lombardo et al., 2015) and strange stories tasks (Abell and Hare, 2005). Based on clinical diagnoses and *WST* performances, we can make sure that participants with ASD did not display any substantial language problems.

Our findings are in concordance with research showing that text-based second-order *ToM* abilities in high-functioning adults and adolescents with ASD are largely unimpaired (Bowler, 1992; Happé, 1994; Ponnet et al., 2004; Scheeren et al., 2013; Schuwerk et al., 2015; Murray et al., 2017). However, in contrast to our data, second-order implicit *ToM* abilities have indeed been reported to be affected in ASD in some studies (Ponnet et al., 2004; Dziobek et al., 2006; Murray et al., 2017). Our data show that persons with ASD are not compromised in this specific *FID* task. Language-related *ToM* impairments have been argued to be subtle (Colle et al., 2008). The most obvious interpretation seems to be that adult persons with ASD with good verbal intelligence are obviously able to learn the complex processes of perspective taking that can be expressed via written language, even if implicit perspective taking is required.

However, it is also possible that our purely behavioral measures in this web-based study were not sensitive enough to identify group differences. Previous studies have shown difficulties associated with second-order *ToM* tasks despite correct task responses, e.g., regarding the causal reasoning about others' mental states (Ozonoff et al., 1991; Bowler, 1992; Happé, 1994; Dziobek et al., 2006), eye movements (Senju et al., 2009; Scheeren et al., 2013; Schneider et al., 2013; Schuwerk et al., 2015; Murray et al., 2017) as well as regarding the attribution of belief which has been shown to not happen automatically (Senju et al., 2009) and to be more difficult for adults with ASD than for control participants (Bradford et al., 2018). Future studies on *FID* in ASD should therefore also include either a non-text-based *ToM* task to assess if persons with ASD show second-order *ToM* impairments in other domains or a *FID* component that requires faster responses, possibly as a task in an ongoing interaction with another person.

Embodiment

There is strong evidence that readers tend to create complex mental models of the presented situation including the protagonists' experiences (e.g., Zwaan and Radvansky (1998)) for which also spatial grounding is a necessary prerequisite (Beveridge and Pickering, 2013). Listeners or readers might even embody the protagonists to re-experience their actions (Kiefer and Pulvermüller, 2012) which possibly facilitates empathizing with them (van Berkum, 2019). Furthermore, participants also adopt a story's timeline as they need more

time to remember events if more time has passed in the story's timeline (Zwaan, 1996; Carreiras et al., 1997). If taking the perspective of a protagonist is accompanied by *embodiment*, *FID* anchoring could possibly be embodied, too. Spatial perspective taking related to embodiment seems to play a role in FID interpretation as indicated by its correlation with FID sensitivity (Kaiser and Cohen, 2012). Furthermore, embodiment has been shown to be relevant for *referential expressions*: In written texts, processing of singular second person pronouns (Brunyé et al., 2011; Gianelli et al., 2011) as well as third person pronouns, but the latter only with spatial anchoring (Gianelli et al., 2011), are usually accompanied by *embodiment* in control participants. This effect seems to happen automatically (Ditman et al., 2010).

If embodiment is indeed involved in *FID* anchoring, the use of third-person pronouns such as in our texts might pose an obstacle for identifying its influence on *FID* anchoring, because embodiment seems to be limited in this case (Gianelli et al., 2011). In a study investigating different text styles on spatial grounding, Salem et al. (2017) found that *FID* alongside spatial anchors presented within the text did not increase self-reported identification with the protagonist nor did it affect spatial perspective taking of participants.

A disturbance of embodiment was proposed to offer an explanation for problems adults with ASD have with certain mentalizing tasks especially in the spatial domain (Pearson et al., 2013). But embodiment does not appear to be necessary, depending on the task, mental rotation processes could be employed (Pearson et al., 2013; Conson et al., 2015). In such a spatial task, participants with ASD showed mostly unimpaired performance when written texts referred to the participant or the other person with first names (Mizuno et al., 2011). In our study, we assume that participants did not make use of any such strategies related to visual perspective taking, as we have not systematically varied spatial information in our texts.

Executive Control

Basic abilities required for perspective taking are inhibitory control (Brown-Schmidt, 2009; Wardlow, 2013) and working memory capacity (Lin et al., 2010; Wardlow, 2013). Both of these executive abilities have been reported to be impaired in persons with ASD (Demetriou et al., 2018; Habib et al., 2019).

The ability to shift between or integrate different perspectives requires the balanced inhibition of one or more of potentially competing perspectives (MacWhinney, 2000; Frith and de Vignemont, 2005). Competing tasks demanding *executive control* hinder the correct selection of perspectives (Qureshi et al., 2010). Schwarzkopf et al. (2014) hypothesized for the visual domain that persons with ASD do in fact implicitly take the perspective of others. However, to decode behaviorally relevant interpretations of the perspective of another person, an attentional shift away from their own perspective toward another person's perspective is necessary, which might be less easily accomplished in ASD (Schwarzkopf et al., 2014). Our explorative correlation analysis suggests that people with ASD reporting more problems with attention switching tend to give less divergent ratings for conditions A and B. One related explanation could be that impaired attention switching might lead to less perspective taking and to reduced sensitivity for cognitively effortful *FID* anchoring as opposed to effortless *FID* anchoring. However, because we did not investigate executive functions, these claims are speculative. Potentially, implicit methods could in principle reveal processing differences in ASD while behavior is otherwise unimpaired (e.g., Bradford et al. (2018)).

Executive control is not only relevant for the shifting of perspective, but also for keeping track of a story or a conversation, and thus for establishing and maintaining prominence relations, accordingly, working memory abilities have been shown to have a positive effect on the cognitive maintenance of shared conversational information or "common ground" in ASD (Schuh et al., 2016). Other abilities impaired in ASD such as planning and fluency (Demetriou et al., 2018) might play a role in predicting, updating and maintaining common ground, and thus the tracking of prominence relations. Our results suggest largely preserved abilities regarding inhibiting less prominent anchors for the interpretation of *FID*, of storing information in working memory to predict upcoming information and of shifting attention toward the different perspectival centers to interpret *FID*. Thus, in our task, participants with ASD appear to track prominence relations comparable to control participants.

LIMITATIONS

One limitation of the study was that we did not test any of the capacities discussed under the umbrella terms of *ToM, embodiment* or *executive control*. Our results therefore offer a first insight into how *FID* is processed at the behavioral level, but cannot yet inform us about potential differences regarding their underlying cognitive processes.

Our web-based study did not allow us to measure reaction times. Considering the issue of response time, further studies investigating persons with ASD should potentially allow for longer time frames for the participants' response or use different methods like self-paced reading to accommodate different needs regarding the duration of stimulus presentation. To stimulate embodied text processing and thus increase perspective taking, longer and more vivid texts might be helpful (MacWhinney, 2000).

CONCLUSION

In this paper, we have shown that implicit perspective taking based on verbal abilities in the context of *FID* is fully preserved in ASD. We replicated the results of previous studies in healthy control persons (Hinterwimmer and Meuser, 2019) that the prominence status of protagonists in written short stories affects acceptability judgments of *FID* anchored to these protagonists. Our results suggest intact processing of *FID* in adults with ASD. We speculate that a possible impairment with respect to second-order *ToM* in ASD can possibly be compensated or can be successfully dealt with in the verbal domain when conventionalized linguistic operations are applied.

Further investigations of *FID* interpretation in ASD will benefit from additional measures beyond naturalness ratings, such as implicit measures like reaction time, eye movement, neurophysiological measures or neuroimaging that might shed light on specific processes involved in perspective taking such as *ToM*, *embodiment* or *executive control*, possibly with a focus on discerning attention switching abilities and conventionalized linguistic operations. With respect to treatment, this result implies that interventions can potentially make use of these language-based resources when focusing on impairments, such as inferring mental states from photos or video animations (Baron-Cohen et al., 2001; Ponnet et al., 2004; Dziobek et al., 2006) or beliefs (Bradford et al., 2018) and intentions (Ozonoff et al., 1991; Bowler, 1992; Happé, 1994; Dziobek et al., 2006).

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because data handling is restricted to the collaborating institutes by our ethics proposal to secure sensitive data such as psycho(patho-)logical data. Requests to access the datasets should be directed to JZ, juliane.zimmermann@uk-koeln.de.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Commission of Cologne University's Faculty of Medicine. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SH and KV contritbuted to theoretical discussions. JZ and SM designed and conducted the study. JZ analyzed the data and wrote the first manuscript version with contributions from SM. All authors read and modified the manuscript several times. All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

The study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 281511265 – SFB 1252.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.675633/full#supplementary-material

REFERENCES

Abell, F., and Hare, D. J. (2005). An experimental investigation of the phenomenology of delusional beliefs in people with Asperger syndrome. *Autism* 9, 515–531. doi: 10.1177/1362361305057857

Achim, A. M., Achim, A., and Fossard, M. (2017). Knowledge likely held by others affects speakers' choices of referential expressions at different stages of discourse. *Lang. Cogn. Neurosci.* 32, 21–36. doi: 10.1080/23273798.2016.1234059

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*. Available Online at: http://dsm.psychiatryonline.org/book.aspx?bookid=556 (accessed February 21, 2018)

Arnold, J. E., Bennetto, L., and Diehl, J. J. (2009). Reference production in young speakers with and without autism: effects of discourse status and processing constraints. *Cognition* 110, 131–146. doi: 10.1016/j.cognition.2008.10.016

Banfield, A. (1982). *Unspeakable Sentences: Narration and Representation in the Language of Fiction*. Boston: Routledge.

Baron-Cohen, S. (1989). The autistic child's theory of mind: a case of specific developmental delay. *J. Child Psychol. Psychiatry* 30, 285–297. doi: 10.1111/j.1469-7610.1989.tb00241.x

Baron-Cohen, S., and Wheelwright, S. (2004). The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *J. Autism Dev. Disord.* 34, 163–175. doi: 10.1023/b:jadd.0000022607.19833.00

Baron-Cohen, S., Leslie, A. M., and Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition* 21, 37–46. doi: 10.1016/0010-0277(85)90022-8

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., and Clubley, E. (2001). The autism-spectrum quotient (AQ): evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *J. Autism Dev. Disord.* 31, 5–17.

Beck, A. T., Steer, R. A., and Brown, G. K. (1996). *Manual for the Beck Depression Inventory-II*. San Antonio: Psychological Corporation.

Begeer, S., De Rosnay, M., Lunenburg, P., Stegge, H., and Terwogt, M. M. (2014). Understanding of emotions based on counterfactual reasoning in children with autism spectrum disorders. *Autism* 18, 301–310. doi: 10.1177/1362361312468798

Beveridge, M. E. L., and Pickering, M. J. (2013). Perspective taking in language: integrating the spatial and action domains. *Front. Hum. Neurosci.* 7:577.

Bowler, D. M. (1992). "Theory of mind" in Asperger's syndrome. *J. Child Psychol. Psychiatry* 33, 877–893.

Bradford, E. E. F., Hukker, V., Smith, L., and Ferguson, H. J. (2018). Belief-attribution in adults with and without autistic spectrum disorders. *Autism Res.* 11, 1542–1553. doi: 10.1002/aur.2032

Brown-Schmidt, S. (2009). The role of executive function in perspective taking during online language comprehension. *Psychon. Bull. Rev.* 16, 893–900. doi: 10.3758/pbr.16.5.893

Brunyé, T. T., Ditman, T., Mahoney, C. R., and Taylor, H. A. (2011). Better you than I: perspectives and emotion simulation during narrative comprehension. *J. Cogn. Psychol.* 23, 659–666. doi: 10.1080/20445911.2011.559160

Burke, D. M., and Shafto, M. A. (2007). "Language and aging," in *The Handbook of Aging and Cognition*, eds F. I. M. Craik and T. A. Salthouse (East Sussex: Psychology Press).

Bürkner, P.-C. (2017). brms: an r package for bayesian multilevel models using stan. *J. Stat. Softw.* 80, 1–28.

Bürkner, P.-C., and Vuorre, M. (2019). Ordinal regression models in psychology: a tutorial: advances in methods and practices in psychological science. *Adv. Methods Pract. Psychol. Sci.* 2:251524591882319.

Callenmark, B., Kjellin, L., Rönnqvist, L., and Bölte, S. (2014). Explicit versus implicit social cognition testing in autism spectrum disorder. *Autism* 18, 684–693. doi: 10.1177/1362361313492393

Carreiras, M., Carriedo, N., Alonso, M. A., and Fernández, A. (1997). The role of verb tense and verb aspect in the foregrounding of information during reading. *Mem. Cogn.* 25, 438–446. doi: 10.3758/bf03201120

Carstensen, L., Turan, B., Scheibe, S., Ram, N., Hershfield, H., Samanez-Larkin, G., et al. (2010). Emotional experience improves with age: evidence based on over 10 years of experience sampling. *Psychol. Aging* 26, 21–33. doi: 10.1037/a0021285

Colle, L., Baron-Cohen, S., Wheelwright, S., and van der Lely, H. K. J. (2008). Narrative discourse in adults with high-functioning autism or Asperger syndrome. *J. Autism Dev. Disord.* 38, 28–40.

Conson, M., Mazzarella, E., Esposito, D., Grossi, D., Marino, N., Massagli, A., et al. (2015). "Put myself into your place": embodied simulation and perspective taking in autism spectrum disorders. *Autism Res.* 8, 454–466. doi: 10.1002/aur.1460

Crocker, M. W., and Keller, F. (2006). "Probabilistic grammars as models of gradience in language processing," in *Gradience in Grammar: Generative Perspectives*, eds R. Vogel and M. Schlesewsky (Oxford: Oxford University Press).

Demetriou, E. A., Lampit, A., Quintana, D. S., Naismith, S. L., Song, Y. J. C., Pye, J. E., et al. (2018). Autism spectrum disorders: a meta-analysis of executive function. *Mol. Psychiatry* 23, 1198–1204.

Ditman, T., Brunyé, T. T., Mahoney, C. R., and Taylor, H. A. (2010). Simulating an enactment effect: pronouns guide action simulation during narrative comprehension. *Cognition* 115, 172–178. doi: 10.1016/j.cognition.2009.10.014

Drummond, A. (2020). *Ibex Farm. Ibex Farm*. Available Online at: https://spellout.net/ibexfarm (accessed May 15, 2020).

Dziobek, I., Fleck, S., Kalbe, E., Rogers, K., Hassenstab, J., Brand, M., et al. (2006). Introducing MASC: a movie for the assessment of social cognition. *J. Autism Dev. Disord.* 36, 623–636. doi: 10.1007/s10803-006-0107-0

Eckardt, R. (2014). *The Semantics of Free Indirect Discourse: How Texts Allow Us to Mind-read and Eavesdrop*. Leiden: Brill.

Fonagy, P., Gergely, G., Jurist, E., and Target, M. (2004). *Affektregulierung, Mentalisierung und die Entwicklung des Selbst*. Milton Park: Routledge.

Frith, C. D., and Frith, U. (2006). The neural basis of mentalizing. *Neuron* 50, 531–534.

Frith, U., and de Vignemont, F. (2005). Egocentrism, allocentrism, and Asperger syndrome. *Conscious Cogn.* 14, 719–738. doi: 10.1016/j.concog.2005.04.006

Frith, U., Morton, J., and Leslie, A. M. (1991). The cognitive basis of a biological disorder: autism. *Trends Neurosci.* 14, 433–438. doi: 10.1016/0166-2236(91)90041-r

Gernsbacher, M. A., and Yergeau, M. (2019). Empirical failures of the claim that autistic people lack a theory of mind. *Arch. Sci. Psychol.* 7, 102–118. doi: 10.1037/arc0000067

Ghaziuddin, M., Ghaziuddin, N., and Greden, J. (2002). Depression in persons with autism: implications for research and clinical care. *J. Autism Dev. Disord.* 32, 299–306.

Gianelli, C., Farnè, A., Salemme, R., Jeannerod, M., and Roy, A. C. (2011). The agent is right: when motor embodied cognition is space-dependent. *PLoS One* 6:e25036. doi: 10.1371/journal.pone.0025036

Gronau, Q. F., Singmann, H., and Wagenmakers, E. (2020). bridgesampling: an r package for estimating normalizing constants. *J. Stat. Softw.* 10, 1–29.

Habib, A., Harris, L., Pollick, F., and Melville, C. (2019). A meta-analysis of working memory in individuals with autism spectrum disorders. *PLoS One* 14:e0216198. doi: 10.1371/journal.pone.0216198

Happé, F. G. E. (1994). An advanced test of theory of mind: understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *J. Autism Dev. Disord.* 24, 129–154. doi: 10.1007/bf02172093

Happé, F., and Charlton, R. A. (2012). Aging in autism spectrum disorders: a mini-review. *Gerontology* 58, 70–78.

Harris, J., and Potts, C. (2009). "Predicting perspectival orientation for appositives," in *Proceedings from the Annual Meeting of the Chicago Linguistic Society*, Vol. 2009, (Chicago: Chicago Linguistic Society).

Hartigan, J. A., and Hartigan, P. M. (1985). The dip test of unimodality. *Ann. Stat.* 13, 70–84.

Hendriks, P., Koster, C., and Hoeks, J. C. (2014). Referential choice across the lifespan: why children and elderly adults produce ambiguous pronouns. *Lang. Cogn. Neurosci.* 29, 391–407. doi: 10.1080/01690965.2013.766356

Henry, J. D., Phillips, L. H., Ruffman, T., and Bailey, P. E. (2013). A meta-analytic review of age differences in theory of mind. *Psychol. Aging* 28, 826–839. doi: 10.1037/a0030677

Himmelmann, N. P., and Primus, B. (2015). "Prominence beyond prosody - a first approximation," in *Proceedings of the International Conference*, ed. A. De Dominicis (Viterbo: DISUCOM Press), 38–58.

Hinterwimmer, S. (2019). Prominent protagonists. *J. Pragmatics* 154, 79–91. doi: 10.1016/j.pragma.2017.12.003

Hinterwimmer, S., and Meuser, S. (2019). "Erlebte rede und protagonistenprominenz," in *Rede- und Gedankenwiedergabe in narrativen Strukturen – Ambiguitäten und Varianz Linguistische Berichte Sonderheft 27*, eds S. Engelberg, C. Fortmann, and I. Rapp (Tübingen: Helmut Buske Verlag), 177–200.

Howlin, P., and Magiati, I. (2017). Autism spectrum disorder: outcomes in adulthood. *Curr. Opin. Psychiatry* 30, 69–76. doi: 10.1097/yco.0000000000000308

Hutchins, T. L., Prelock, P. A., and Bonazinga, L. (2012). Psychometric evaluation of the theory of mind inventory (ToMI): a study of typically developing children and children with autism spectrum disorder. *J. Autism Dev. Disord.* 42, 327–341. doi: 10.1007/s10803-011-1244-7

Jasinskaja, K., Chiriacescu, S. I., Donazzan, M., von Heusinger, K., and Hinterwimmer, S. (2015). "Prominence in discourse," in *Prominences in Linguistics. Proceedings of the pS-prominenceS International Conference*, ed. A. De Dominicis (Viterbo: DISUCOM Press), 134–153.

Jeffreys, H. (1939). *Theory of Probability*. Oxford: The Clarendon Press.

Kaiser, E. (2015). Perspective-shifting and free indirect discourse: experimental investigations. *Semant. Linguist. Theory* 25, 346–372. doi: 10.3765/salt.v25i0.3436

Kaiser, E., and Cohen, A. (2012). In someone else's shoes: a psycholinguistic investigation of FID and perspective taking. in *Proceedings of the Talk Presented at the Conference "Quotation: Perspectives from Philosophy and Linguistics*. Bochum: Ruhr-University-Bochum.

Kenny, L., Hattersley, C., Molins, B., Buckley, C., Povey, C., and Pellicano, E. (2016). Which terms should be used to describe autism? perspectives from the UK autism community. *Autism* 20, 442–462. doi: 10.1177/1362361315588200

Kiefer, M., and Pulvermüller, F. (2012). Conceptual representations in mind and brain: theoretical developments, current evidence and future directions. *Cortex* 48, 805–825. doi: 10.1016/j.cortex.2011.04.006

Kuzmanovic, B., Schilbach, L., Lehnhardt, F. G., Bente, G., and Vogeley, K. (2011). A matter of words: impact of verbal and nonverbal information on impression formation in high-functioning autism. *Res. Autism Spectr. Disord.* 5, 604–613. doi: 10.1016/j.rasd.2010.07.005

Lasersohn, P. (2005). Context dependence, disagreement, and predicates of personal taste. *Linguist. Philos.* 28, 643–686. doi: 10.1007/s10988-005-0596-x

Leslie, A. M., and Thaiss, L. (1992). Domain specificity in conceptual development: neuropsychological evidence from autism. *Cognition* 43, 225–251. doi: 10.1016/0010-0277(92)90013-8

Lever, A. G., and Geurts, H. M. (2016). Age-related differences in cognition across the adult lifespan in autism spectrum disorder. *Autism Res.* 9, 666–676. doi: 10.1002/aur.1545

Lever, A. G., Werkle-Bergner, M., Brandmaier, A. M., Ridderinkhof, K. R., and Geurts, H. M. (2015). Atypical working memory decline across the adult lifespan in autism spectrum disorder? *J. Abnorm. Psychol.* 124, 1014–1026. doi: 10.1037/abn0000108

Light, L. L., and Capps, J. L. (1986). Comprehension of pronouns in young and older adults. *Dev. Psychol.* 22, 580–585. doi: 10.1037/0012-1649.22.4.580

Lin, S., Keysar, B., and Epley, N. (2010). Reflexively mindblind: using theory of mind to interpret behavior requires effortful attention. *J. Exp. Soc. Psychol.* 46, 551–556. doi: 10.1016/j.jesp.2009.12.019

Lombardo, M. V., Lai, M. C., Auyeung, B., Holt, R. J., Allison, C., Smith, P., et al. (2015). Enhancing the precision of our understanding about mentalizing in adults with autism. *BioRxiv [Preprint]* doi: 10.1101/034454

Mächler, M. (2015). *diptest: Hartigan's Dip Test Statistic for Unimodality - Corrected*. Available Online at: https://github.com/mmaechler/diptest (accessed March 15, 2021).

MacWhinney, B. (2000). Perspective taking and grammar. *Jpn. Soc. Lang. Sci.* 1, 1–25.

Makowski, D., Ben-Shachar, M. S., Chen, S. H. A., and Lüdecke, D. (2019). Indices of effect existence and significance in the bayesian framework. *Front. Psychol.* 10:2767.

Mizuno, A., Liu, Y., Williams, D. L., Keller, T. A., Minshew, N. J., and Just, M. A. (2011). The neural basis of deictic shifting in linguistic perspective-taking in high-functioning autism. *Brain* 134, 2422–2435. doi: 10.1093/brain/awr151

Murray, K., Johnston, K., Cunnane, H., Kerr, C., Spain, D., Gillan, N., et al. (2017). A new test of advanced theory of mind: the "strange stories film task" captures social processing differences in adults with autism spectrum disorders. *Autism Res.* 10, 1120–1132. doi: 10.1002/aur.1744

Ozonoff, S., Pennington, B. F., and Rogers, S. J. (1991). Executive function deficits in high-functioning autistic individuals: relationship to theory of mind. *J. Child Psychol. Psychiatry* 32, 1081–1105. doi: 10.1111/j.1469-7610.1991.tb00351.x

Pearson, A., Ropar, D., and Hamilton, A. F. C. (2013). A review of visual perspective taking in autism spectrum disorder. *Front. Hum. Neurosci.* 7:652.

Peterson, E., and Miller, S. (2012). The eyes test as a measure of individual differences: how much of the variance reflects verbal IQ? *Front. Psychol.* 3:220.

Ponnet, K. S., Roeyers, H., Buysse, A., De Clercq, A., and Van der Heyden, E. (2004). Advanced mind-reading in adults with Asperger syndrome. *Autism* 8, 249–266. doi: 10.1177/1362361304045214

Premack, D., and Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behav. Brain Sci.* 1, 515–526. doi: 10.1017/s0140525x00076512

Qureshi, A. W., Apperly, I. A., and Samson, D. (2010). Executive function is necessary for perspective selection, not Level-1 visual perspective calculation: evidence from a dual-task study of adults. *Cognition* 117, 230–236. doi: 10.1016/j.cognition.2010.08.003

R Core Team (2019). *R: a Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

RStudio Team (2016). *RStudio: Integrated Development for R*. Boston, MA: RStudio, Inc.

Salem, S., Weskott, T., and Holler, A. (2017). Does narrative perspective influence readers' perspective-taking? an empirical study on free indirect discourse, psycho-narration and first-person narration. *Glossa J. Gen. Linguist.* 2:61. doi: 10.5334/gjgl.225

Scheeren, A. M., de Rosnay, M., Koot, H. M., and Begeer, S. (2013). Rethinking theory of mind in high-functioning autism spectrum disorder. *J. Child Psychol. Psychiatry* 54, 628–635. doi: 10.1111/jcpp.12007

Schlenker, P. (2004). Context of thought and context of utterance: a note on free indirect discourse and the historical present. *Mind Lang.* 19, 279–304. doi: 10.1111/j.1468-0017.2004.00259.x

Schmidt, K.-H., and Metzler, P. (1992). *WST-Wortschatztest*. Weinheim: Beltz Test GmbH.

Schneider, D., Slaughter, V. P., Bayliss, A. P., and Dux, P. E. (2013). A temporally sustained implicit theory of mind deficit in autism spectrum disorders. *Cognition* 129, 410–417. doi: 10.1016/j.cognition.2013.08.004

Schuh, J., Eigsti, I.-M., and Mirman, D. (2016). Discourse comprehension in autism spectrum disorder: effects of working memory load and common ground. *Autism Res.* 9, 1340–1352. doi: 10.1002/aur.1632

Schuwerk, T., Vuori, M., and Sodian, B. (2015). Implicit and explicit theory of mind reasoning in autism spectrum disorders: the impact of experience. *Autism* 19, 459–468. doi: 10.1177/1362361314526004

Schwarzkopf, S., Schilbach, L., Vogeley, K., and Timmermans, B. (2014). "Making it explicit" makes a difference: evidence for a dissociation of spontaneous and intentional level 1 perspective taking in high-functioning autism. *Cognition* 131, 345–35403. doi: 10.1016/j.cognition.2014.02.003

Senju, A., Southgate, V., White, S., and Frith, U. (2009). Mindblind eyes: an absence of spontaneous theory of mind in Asperger syndrome. *Science* 325, 883–885. doi: 10.1126/science.1176170

Steube, A. (1985). Erlebte rede aus linguistischer sicht. *Zeitschrift für Germanistik* 6, 389–406.

Stine-Morrow, E. A. L., Soederberg Miller, L. M., Gagne, D. D., and Hertzog, C. (2008). Self-regulated reading in adulthood. *Psychol. Aging* 23, 131–153. doi: 10.1037/0882-7974.23.1.131

Streefkerk, B. M. (2002). *Prominence. Acoustic and Lexical/Syntactic Correlates*. Available Online at: https://www.bibliotheek.nl/catalogus/titel.240864522.html/prominence--acoustic-and-lexical-syntactic-correlates/ (accessed January 15, 2019)

Swettenham, J. G. (1996). What's inside someone's head? conceiving of the mind as a camera helps children with autism acquire an alternative to a theory of mind. *Cogn. Neuropsychiatry* 1, 73–88. doi: 10.1080/135468096396712

Tepest, R. (2021). The meaning of diagnosis for different designations in talking about Autism. *J. Autism. Dev. Disord.* 51, 760–761. doi: 10.1007/s10803-020-04584-3

van Berkum, J. J. A. (2019). "Language comprehension and emotion," in *The Oxford Handbook of Neurolinguistics*, eds de Zubicaray and N. O. Schiller (Oxford: Oxford University Press).

Vivanti, G. (2020). Ask the editor: what is the most appropriate way to talk about individuals with a diagnosis of autism? *J. Autism Dev. Disord.* 50, 691–693. doi: 10.1007/s10803-019-04280-x

Wardlow, L. (2013). Individual differences in speaker's perspective taking: the roles of executive control and working memory. *Psychon. Bull. Rev.* 20, 766–772. doi: 10.3758/s13423-013-0396-1

World Medical, and Association. (2013). World medical association declaration of helsinki: ethical principles for medical research involving human subjects. *JAMA* 310, 2191–2194. doi: 10.1001/jama.2013.281053

Zeman, S. (2017). Confronting perspectives: modeling perspectival complexity in language and cognition. *Glossa A J. General Linguist.* 2:6.

Zwaan, R. A. (1996). Processing narrative time shifts. *J. Exp. Psychol. Learn. Memory Cogn.* 22, 1196–1207. doi: 10.1037/0278-7393.22.5.1196

Zwaan, R. A., and Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychol. Bull.* 123, 162–185. doi: 10.1037/0033-2909.123.2.162

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

6 General Discussion

This thesis aimed at investigating mentalizing and perspective taking in the perception of language in autism – in particular regarding the perception of prominence – by (i) establishing a paradigm to investigate mentalizing drawing on the perception of nonverbal information, namely gaze and intonation as used for highlighting objects (study 1a and 1b), (ii) investigating mentalizing abilities in autistic adults based on nonverbal information using the established paradigm (study 1c) and (iii) investigating perspective taking in autistic adults reading short stories including protagonists that stand out to different degree (study 2).

Albeit both investigating perspective taking in the perception of language, these two lines of studies focus on two different types of study material and tasks. Correspondingly, the perspective taking abilities required to perform the tasks differ.

6.1 Inferring Mental States from Nonverbal Information

6.1.1 Summary

A paradigm to investigate mentalizing drawing on the perception of nonverbal information, namely gaze and intonation, was established in online study 1a (Zimmermann et al., 2020). The main findings were subsequently replicated in a controlled laboratory setting (studies 1b and 1c). In studies 1a and 1b, non-autistic participants were asked to infer how important a presented object was to a virtual character. In study 1c (Zimmermann et al., 2024), autistic adults participated in this paradigm along with non-autistic adults.

At the group level, in both autistic and non-autistic participants, results of studies 1a–c showed an influence of the manipulated cues – gaze duration and pitch height – on mentalizing, more specifically on the ratings of object importance to the virtual character: The longer the virtual character looked at an object, and the higher the pitch excursion on the accented syllable of the corresponding utterance denoting the respective object, the higher participants rated its importance for the virtual character. The influence of gaze duration was greater in the autistic group compared to the non-autistic group suggesting that the autistic group paid more attention to the gaze cue as opposed to the pitch cue.

At the individual level, both autistic and non-autistic participants differed regarding which cue they based their ratings on. Three subgroups could be identified: (i) “Lookers” based their ratings on gaze duration, (ii) “Listeners” based their ratings on pitch height, (iii) “Neithers” did not use either of the two cues. Moreover, participants’ rating behavior was linked to their gaze

behavior: “Lookers” spent more time looking at the virtual character’s eye region than “Listeners”, whereas “Listeners” spent more time fixating the object region.

6.1.2 A Mentalizing Task amongst Others

In line with previous studies on the general population, results from studies 1a and 1b of the current thesis demonstrate that non-autistic participants can infer mental states of others based on gaze duration (Einav & Hood, 2006; Freire et al., 2004; Klami, 2010; Klami et al., 2008) and intonation (Dahan et al., 2002; Ito & Speer, 2008; Kaland et al., 2014; Watson et al., 2008; Weber et al., 2006). Contrary to accounts of impaired mentalizing abilities in autistic adults based on nonverbal information (Baron-Cohen et al., 2001; David et al., 2010; Dziobek et al., 2006; Ponnet et al., 2004; White et al., 2011), study 1c of this thesis demonstrates similar mentalizing abilities in autistic and non-autistic people. However, study 1c differs from these studies not only regarding the stimuli investigated: The previous studies did not exclusively manipulate gaze duration and pitch height. Study 1c also differs from these studies with respect to stimulus and task complexity: The studies referred to above include moving triangles, naturalistic photos of the eye region showing various expressions, naturalistic video material of two or more people interacting with each other, virtual characters expressing preference by means of facial expression and body language. For example, a study on inferring a virtual character’s preference for one object over a second object, showed that autistic participants were slower and made more errors during this task than non-autistic participants (David et al., 2010). The stimulus material in this study was arguably the most structured one of the studies mentioned above, manipulating object preference by varying three factors – facial expression, whole body rotation and hand gestures – on three levels each. In contrast, in the paradigm used in study 1c, which was also highly structured, complexity was even lower with only two cues varying on two levels. Additionally, the question participants were required to answer was the same in each trial. This likely led to a low level of difficulty and a good chance to identify informative cues by strategically using the two cues available.

This may be an important differentiation as the task can potentially be solved by means of intelligence and attention instead of mentalizing faculties that are likely needed to solve less explicit ToM tasks based on nonverbal information. A meta-analysis including autistic participants from childhood to adulthood showed that within the comparison group, performance on the “Reading the mind in the Eyes” test – referred to as the Eyes-test in study 1c – improved with age and intelligence, while this was not the case in the autistic group (Peñuelas-Calvo et al., 2019). This finding suggests that age and intelligence cannot

compensate for the difficulties autistic people face when tackling the Eyes-test. In contrast, the paradigm in study 1c can possibly be solved by investing attention and intelligence resources, i.e. without mentalizing abilities required to solve the Eyes-test. The assessment of mentalizing abilities which go beyond strategic cue use may be limited in study 1c by the relatively high structuredness of the task. Therefore, enhancing the paradigm to a level of less structured, more life-like variability may change the study's outcome and may lead to better insight into which mentalizing tasks may be particularly challenging for autistic people.

6.1.3 Relative Impact of Gaze Duration and Intonation

The impact of gaze duration and intonation on the ratings was not the same in the autistic and the non-autistic group but also differed between the web-based study and the laboratory study that included only non-autistic participants.

6.1.3.1 Relative Impact of Gaze Duration and Intonation in Autism

Compared to the non-autistic group, the autistic group took gaze duration into account to a greater extent than pitch height for their ratings. They rated the importance higher than the non-autistic group when the object was gazed at for a long duration if presented with low pitch – and as high as the non-autistic group if it was presented with high pitch. When gazed at for a short duration, they judged the object's importance to the virtual character to be lower than the non-autistic group.

One possible explanation for why autistic participants in study 1c assigned more weight to the virtual character's gaze, rather than pitch height, could be a difficulty in perceiving vocal pitch. This challenge has been observed in both speech (Grice et al., 2016, 2023; Schelinski & von Kriegstein, 2019) and non-speech contexts (Schelinski et al., 2017). Another factor could be an enhanced auditory processing capacity in autistic people compared to non-autistic people (Remington & Fairnie, 2017): In the paradigm used in studies 1a–c, the speech stimulus varied between trials due to the use of natural speech material, which resulted in slight differences in intonation across items. While the accented syllable in high-pitch stimuli was consistently 45 Hz higher than in low-pitch stimuli, there were minor fluctuations in the absolute Hz values of low-pitch items. Furthermore, other prosodic aspects, like the length of the utterance, could have influenced how prominence was perceived. In contrast, the gaze cue was relatively straightforward to process, as it had a fixed duration of either 0.6 or 1.8 seconds, making it easier to categorize. Consequently, individuals managing a large amount of auditory information may find it easier to focus on the gaze cue.

Based on the premise that the task induces top-down attention allocation, the results may tentatively be interpreted as a focus on eye gaze as opposed to intonation in autistic people with regard to themselves or to autism in general. Research on gaze behavior in autism by far exceeds research on intonation in autism. Gaze behavior may overall be perceived as a more prominent characteristic feature in autism than intonation by both autistic and non-autistic people. Autistic participants may therefore have focused on gaze more strongly than on intonation in study 1c. Additionally, participants that took a systematic approach to complete the task may rightfully have inferred from the eye-tracking device that the experimenter was interested in their visual attention, and thus in their interpretation of – amongst others – the visual stimuli. Starting the task with a focus on visual information, as opposed to auditory information, is therefore a logical and strategic approach to the task.

6.1.3.2 Relative Impact of Gaze Duration and Intonation in Non-autistic People

Whereas in study 1a, gaze duration and pitch height had a similar effect on non-autistic participants' responses, the results of study 1b suggest that the impact of pitch height on non-autistic participants' ratings of object importance was greater than the impact of the virtual character's gaze duration. In study 1c, in which the non-autistic comparison group comprised participants from study 1b, a greater impact of pitch height as opposed to gaze duration was less pronounced than in study 1b in the non-autistic comparison group as a coincidental result of sample size reduction. In the autistic group, a greater impact of pitch height as opposed to gaze duration was not present.

The reason for different weights of gaze duration and pitch height in the studies including non-autistic participants may be the different conditions under which participants worked on the paradigm: Study 1a was a web-based study, whereas study 1b was a laboratory study with controlled visual and acoustic conditions. However, as results of previous studies (Krahmer et al., 2002b; Mixdorff et al., 2013; Srinivasan & Massaro, 2003; Swerts & Krahmer, 2008) suggest, prosody – and pitch accents in particular – may have a perceptual advantage compared to visually perceived facial movements when it comes to prominence perception, which here refers to prosodically highlighted elements in speech. Like studies 1a and 1b, these studies did not systematically aim at balancing prominence perception following pitch accents and visual facial movements. As for the current paradigm, this could be achieved by incorporating different levels of prominence, i.e. different maximum pitch excursions on the accented syllable as well as different gaze durations of the virtual character. Additionally, in the studies referred to, participants were asked to straightforwardly infer focus, prominence, or discriminate

questions from statements without the need to additionally infer mental states. This can – under normal listening conditions – arguably be done based on prosodic information alone. It is relatively easier than additionally inferring mental states. This renders the visual cues a mere add-on in these tasks. However, in natural settings, the importance of an object to a person may well be accompanied by signs of affection or emotion, not just displays of attentional focus. Considering the important role of the visual modality when inferring emotions from prosody and facial information (Collignon et al., 2008; de Gelder & Vroomen, 2000; Vroomen et al., 2001), pitch height is unlikely to generally outweigh visual facial information when it comes to judging the importance of an object to a real-life person in a life-like scenario. Accordingly, while reports of greater impact of pitch height as opposed to eye gaze can explain the differences between findings of studies 1a and 1b, these differences may not expand to real-life scenarios and may be a mere result of the experimental setting.

Another reason for the different influence of the gaze and the pitch cue in study 1a and 1b could be the different samples. Because the web-based study did not include an extensive set of tests and questionnaires, which would have allowed for screening the non-autistic participants for signs of autism not only in the laboratory study but also in the web-based study, it cannot be ruled out that autistic people or people with high autistic traits entered the web-based study. In study 1b, non-autistic participants tended to use pitch height more the greater their empathic skills and the lower their autistic traits, which aligns with a greater impact of gaze duration on ratings of autistic participants in study 1c. The study information of the web-based study excluded neurological or psychiatric disorders in general, but not autism specifically. Even if no person with an autism diagnosis entered the web-based study, it may still be possible that participants with very high autistic traits or very low empathic skills entered the study undetected (which was not the case in the laboratory study). These participants may have attenuated a possible tendency in the non-autistic sample to take into account the pitch cue to a greater extent. Accordingly, further investigation is required to better understand the impact of pitch height and gaze durations on prominence perception in this paradigm both in autistic and non-autistic participants.

6.1.4 Audiovisual Interaction

The findings further provide insights about mentalizing in settings in which both cues are presented in combination. In the paradigm used in these studies, both autistic and non-autistic participants made use of either the visual information, i.e. gaze duration, or the acoustic information, i.e. pitch height, or none, but not both. Studies investigating the perception of pitch

accents in the general population have shown that a simultaneous presentation of facial movement, head or hand gestures can lead to greater prominence perception compared to the presentation of only one modality (Ambrazaitis et al., 2020; Krahmer et al., 2002a; Mixdorff et al., 2013; Prieto et al., 2015; Swerts & Krahmer, 2008), while an interaction effect is not to be expected (Prieto et al., 2015; Swerts & Krahmer, 2008).

The reason that, at the individual level, a combination of the auditory and the visual cue did not lead to higher ratings of object importance in the paradigm used in studies 1a–c may be the reductionistic study material and the task instruction. The latter introduced considerable cognitive load which can be detrimental when keeping track of simultaneously presented audio-visual information (Fougnie et al., 2018) as well as to audio-visual integration (Ren et al., 2023). Most importantly, however, the instruction did not specify whether the virtual character would communicate via eye gaze, prosody or other behavior.

Participants likely searched for a valid cue and stuck to it once they felt it was informative. Rarely, participants reported to have identified both cues, in which case this nevertheless did not result in them making use of both cues throughout the experiment, which exemplifies participants' efficiency regarding their cue use in this task. Because participants had the freedom to use one, two, multiple or even no cues at all, there was no need for them to further identify other cues that may have added to their perception if they only found one that was sufficient to solve the task. Participants would arguably feel the need to search other sources of information in this paradigm only if their selected cue was not informative enough. Equivalently, in another demanding, but highly structured task (Macdonald & Tatler, 2013), in which auditory instructions were either ambiguous or unambiguous, participants made use of the instructor's gaze behavior only, if the auditory information was not informative enough. Similarly, multimodal cue use has been reported in audiovisual studies in which for example auditory information is insufficient or difficult to understand: When trying to comprehend whispered speech, facial cues and head movements can help detecting focus (Dohen & Lœvenbruck, 2009). Likewise, intelligibility has been shown to be increased for speech in noise by head movements (Munhall et al., 2004) and for vocoded (i.e. synthetically distorted) speech by head nods and eye brow raises (Al Moubayed & Beskow, 2009).

The paradigm as applied does not support integration of prosodic and gaze cues at a detectable level. It is possible, however, that integration occurred which would be detectable with other measures. Comparable to findings of longer reaction times for incongruent vocal and facial information in emotion recognition (Föcker et al., 2011) and prominence perception (Swerts

& Krahmer, 2008), incongruent audio-visual information may also lead to longer response times in the current paradigm, i.e. when the audio-visual channels suggest different levels of prominence, participants may react slower than when both intonation and gaze duration suggest that an object is important to the virtual character, or when both signal that it is not. Detecting such a difference would likely require the paradigm to be changed so that response time variation is reduced and the informative value of response times is optimized, e.g. by reducing the forced-choice selection to two options ("important" and "not important") and by more strongly restricting response time windows.

6.1.5 Individual Behavior

Participants were clustered into three behaviorally different subgroups: "Lookers", "Listeners" and "Neithers". Across both the autistic and the non-autistic group, participants were found across all three clusters. In the light of the previously reported importance of verbal information at the expense of nonverbal information in autistic people (Kuzmanovic et al., 2011; Stewart et al., 2013) along with a focus on word characteristics at the expense of intonation (Grice et al., 2016), it would not have been surprising, if more autistic individuals would have been identified as "Neithers", which was however not the case in this paradigm. The stronger influence of the gaze cue on the ratings in the autistic group was mirrored in autistic participants being identified as "Lookers" twice as often (14 of 24 participants, i.e. 58 % of the autistic group) as they were identified as "Listeners" (7 participants, i.e. 29 %, of the autistic group). In contrast, six participants (25 % of the non-autistic group) in the non-autistic group were categorized as "Lookers", nine (38 % of the non-autistic group) as "Listeners".

The three behaviorally different subgroups feed into the literature on interindividual variability in the general population when it comes to perception of gaze in gaze cueing paradigms (Bayliss et al., 2005) as well as to the perception of pitch in prosodic prominence perception (Baumann & Winter, 2018; Roy et al., 2017; Wagner et al., 2019). Unlike these studies, studies 1a–c combined these factors in one paradigm and showed interindividual variability not within one domain, but across domains. Especially the factors contributing to the perception of prosodic prominence have been examined in detail (Baumann & Winter, 2018; Roy et al., 2017; Wagner et al., 2019). Findings indicate that pitch height is one of several factors contributing to the perception of prominence. Another acoustic cue to prominence is for example syllable duration (Baumann & Winter, 2018; Wagner et al., 2019). Studies 1a–c did not vary other acoustic factors besides pitch height. This clearly played to the strength of participants who were sensitive to pitch accents in prominence perception and likely made prominence detection in

the auditory cue more difficult for participants with certain perceptive prosodic profiles: For instance, participants who are generally less sensitive to pitch accents but more sensitive to other acoustic factors, such as syllable duration, likely found it harder to identify the auditorily presented prominence cue. Upcoming research could include different acoustic cues besides pitch height and take individual prosodic perceptive profiles into account to better understand and further dissect behavior in reaction especially to the acoustic stimuli. Likewise, including individual factors contributing to gaze perception, such as the perception of mutual and directional gaze in female and male characters, could be equally informative.

Previous findings suggest that participants paying less attention to pitch accents in the acoustic stimuli may instead pay more attention to lexical, semantic-syntactic aspects (Baumann & Winter, 2018). While the stimuli in studies 1a–c are not embedded in an informative syntactic context, their lexical or semantic meaning could be controlled or manipulated in the paradigm to a greater extent in the future to better characterize the "Neithers" subgroup. For example, contributing factors such as the importance of the object to the participant, general object importance as well as features of the virtual character could be assessed or accounted for.

It is unclear, to what degree the attentional preference of "Lookers", "Listeners" and "Neithers" was formed in the rating paradigm and to what degree they entered the studies with a corresponding default attentional preference with respective auditory and visual sensitivity or different preferences to fixate a person's eye region. Within the non-autistic group in studies 1b and 1c, lower visual sensitivity – assessed via questionnaires before the experiment – was associated with less focus on the gaze cue as judged by rating behavior, which offers some support for the idea of prior attentional defaults. In the autistic group, no such relationship was found, which may be indicative of autistic participants attuning to the task's systematic structure more strongly than the non-autistic group.

Additionally, participants differed with regard to their perception of the virtual character as being the source of the utterance as well as with regard to their perception of the virtual character herself. As mentalizing is influenced by these perceptions, a more life-like scenario including a real person may increase participants' mentalizing tendencies and thus lead to different results. Whereas comparable gaze behavior in autistic and non-autistic adolescents and adults has been reported in reaction to static images of virtual characters' faces as compared to photographs of real people (Hernandez et al., 2009), different settings (computer screen as opposed to a real-live interaction with an experimenter) changed the gaze behavior of autistic children and adolescents, but not that of the non-autistic comparison group (Grossman et al.,

2019). Equivalently, increasing external validity of the paradigm of studies 1a–c may paint a clearer picture of autistic (and non-autistic) behavior in more life-like settings.

6.1.6 Gaze Fixation Durations

With regard to gaze behavior, the results from study 1c in this thesis do not support the idea of generally attenuated attention towards the eye region in autistic adults, or towards gaze cues in particular (Itier et al., 2007). Both the autistic and the non-autistic group spent more time fixating the eye region than the object or the head region (excluding the eyes). Additionally, the time spent fixating these regions was similar in both diagnostic groups. These findings align with reports of longer fixations of the eye region as opposed to other parts of the face or to objects in the environment in autism (Auyeung et al., 2015; Dalton et al., 2005; Fedor et al., 2018; Freeth et al., 2010; Hernandez et al., 2009). This behavior has also been shown in the general population (Fedor et al., 2018; Freeth et al., 2010; Henderson et al., 2005).

Gaze behavior in autism is dependent on the stimulus (Chita-Tegmark, 2016; Guillon et al., 2014) and the task (Setien-Ramos et al., 2022). For example, it has been suggested that the amount of autistic participants' gaze towards people in the stimulus material is more likely to differ from a non-autistic group if the stimulus material includes more than one person (Guillon et al., 2014). This was not the case in study 1c and may explain why no statistical reliable differences were observed. Moreover, in tasks that require participants to gain information from other persons' eye gaze, autistic and non-autistic adults fixate the eye region, as opposed to objects, for a similar amount of time (Setien-Ramos et al., 2022). On the other hand, in free viewing tasks, i.e. in tasks that allow for visual exploration at will, autistic adults have, compared to non-autistic adults, been shown to spend less time fixating the eye region (Setien-Ramos et al., 2022). The paradigm used in study 1c was not a free viewing task as it required participants to search for potentially informative cues. Only if participants concentrated exclusively on the auditory information, they were able to visually explore freely in this paradigm. However, as the virtual character's face and the object were the only visual stimuli in this paradigm, the expected areas of interest were nevertheless limited. Additionally, focusing exclusively on the auditory information in this paradigm equals discarding half of the available information which presumably is an unlikely behavior, at least at the beginning of the rating task.

Across both the autistic and the non-autistic group, rating behavior was mirrored by participants' gaze behavior: "Lookers" spent more time looking at the virtual character's eyes than "Listeners" and tended to also spend more time looking at the eyes than "Neithers".

"Neithers" in turn spent more time looking at the eyes than "Listeners". "Neithers" and "Listeners", who spent a similar amount of time looking at the object, both spent more time looking at the object than "Lookers". As attention is closely linked to gaze direction (Buswell, 1935; DeAngelus & Pelz, 2009; Yarbus, 1967), these findings are plausibly aligned with the idea that "Lookers" seek information from the eyes, while "Listeners" and "Neithers" seek information elsewhere. Along with qualitative feedback participants made after the experiment this supports the idea that participants were actively monitoring their preferred input modality. In the non-autistic group, "Neithers" tended to spend more time looking at the head region than "Lookers" and "Listeners". This tendency is likely indicative of the "Neithers" searching for information not accessible in the eyes or the object region. For example, some participants from this subgroup reported to have taken into account the virtual character's age and sex for their rating.

In the eye region, visual inspection of the graphs from the fixation duration data – but not the statistical analysis – showed a small tendency for autistic "Listeners" to look at the virtual character's eyes for a shorter period of time than non-autistic "Listeners". It is possible that such a trend was not detected in the analysis due to e.g. great variation in fixation durations. In future investigations, it may be interesting to further explore this tendency as it could reveal different gaze behavior in the autistic and the non-autistic group. As has been demonstrated in different tasks, autistic participants flexibly adjust their gaze behavior to task-demands (Caruana et al., 2018; Riby et al., 2013; Setien-Ramos et al., 2022). As "Listeners" do not gather information from the eye region and are able to visually explore freely, this setting resembles a free-viewing task in which autistic participants are more likely to avoid looking at the eye region (Setien-Ramos et al., 2022). It seems plausible that autistic "Listeners" may default to avoiding the eye region in a situation in which a task does not require otherwise. One explanation for avoiding eye gaze in this situation could be an attenuated attention to social stimuli in general in autism (Chita-Tegmark, 2016). Autistic people may also avoid mutual gaze because they experience it as threatening or stressful (Tottenham et al., 2014). Moreover, it has been suggested that autistic participants perceive the eyes primarily as deictic cues instead of social cues (Caruana et al., 2018; Ristic et al., 2005), which may also explain this tendency. Systematically assessing uneasiness, stress or exhaustion in reaction to the virtual character's gaze in the paradigm used in studies 1a–c, as well as collecting respective qualitative feedback may lead to better understanding of gaze perception, especially in the autistic group.

6.1.7 Attention (Re-)Orienting

In the paradigm used in studies 1a–c, the duration of the virtual character's gaze towards the object and the pitch height of the utterances referring to the object affected participants' ratings of object importance for the virtual character. But it did not affect participants' fixation durations towards the virtual character, the character's eyes or the object. Moreover, it did not have an effect on later object recognition. Despite the paradigm not being aimed at producing these effects, participants' gaze behavior and object memory was investigated to identify potential traces of attention (re-)orienting by means of the virtual character's gaze duration towards the object and the intonation of the respective utterance.

6.1.7.1 Participants' Gaze Behavior

No gaze cueing or gaze following effects were detected in the paradigm used in studies 1a–c. Comparable to other studies that demonstrated the impact of task instructions on fixation durations of participants from the general population (Buswell, 1935; Klami, 2010; Klami et al., 2008), findings of studies 1a–c suggest a strong influence of the instruction to infer the virtual character's mental state on participants' gaze behavior – possibly at the expense of potential gaze cueing or gaze following effects.

In contrast, when tasks do not require participants to mentalize, a gaze following effect can be observed in the general population: If a person's gaze is directed towards an object, it increases an observer's gaze duration towards that object (Adil et al., 2018; Castelhano et al., 2007; Hutton & Nolte, 2011; Theuring et al., 2007). Most interestingly in the context of studies 1a–c: The longer a person is observed looking at an object, the longer the observer tends to look towards the respective object (Freeth et al., 2010). In the study by Freeth and colleagues (2010), autistic participants spent less time fixating the gaze-indicated object than non-autistic participants. Similarly, other studies show diminished gaze following or gaze cueing effects in autism: Autistic adults, for instance, look less at objects another person is looking at (Wang et al., 2015) and spend less time looking at such gaze-indicated objects (Fletcher-Watson et al., 2009; Freeth et al., 2010).

Similar to gaze cues, prosody can orient a listener's attention in the general population (Dahan et al., 2002; Ito & Speer, 2008; Watson et al., 2008; Weber et al., 2006). Unfortunately, respective studies in autistic adults have not been carried out yet. However, a study including autistic children suggests that pitch accents guide overt attention in autism similarly as in the general population: The study demonstrates that an utterance denoting an object, if highlighted by a pitch accent, increases the time spent looking at the indicated object both in autistic and non-autistic individuals (Ito et al., 2022).

6.1.7.2 Effect of Memory on Object Recognition

After participants rated the importance of objects to a virtual character, they completed a memory task, more precisely an object recognition task. This task was implemented to detect possible memory traces of attention directed towards the object during the rating period. The object recognition task successfully measured object memory as demonstrated by effects generally found in recognition tasks such as an effect of participants' fixation duration of a stimulus on its recognition (Droll & Eckstein, 2009; Martini & Maljkovic, 2009; Melcher, 2001, 2006) as well as a recency effect (Brady et al., 2008; Konkle et al., 2010).

The results of study 1c demonstrate similar performance regarding object recognition in autistic and non-autistic participants, which was reported by previous studies (Boucher et al., 2005; Bowler et al., 2000; Ring et al., 2015) – albeit not consistently (O'Hearn et al., 2014).

An effect of the virtual character's gaze duration or intonation on object recognition was, however, not found in studies 1b and 1c, neither in the autistic nor in the non-autistic group. In contrast, studies in autistic people (Ito et al., 2022) and in the general population (Adil et al., 2018; Dodd et al., 2012; Fraundorf et al., 2010, 2012; Gregory & Jackson, 2017; Gregory & Kessler, 2022; Kember et al., 2021; Kushch et al., 2018; Morett & Fraundorf, 2019; Sajjacholapunt & Ball, 2014) reported effects of prosodic prominence or gaze during stimulus encoding on object memory. These studies are, however, not directly comparable to studies 1b and 1c due to different methodologies. Apart from the studies' manipulation of prosodic prominence and gaze not being operationalized specifically and exclusively via pitch excursion or gaze duration, task instructions in the paradigm used in studies 1b and 1c arguably included higher task demands as they required participants to allocate attention to mental state attributions of the virtual character. Additionally, the later memory test, which was not announced before the object presentation phase, was presented more than ten minutes after object presentation had ended. If pitch height and gaze duration exert any cueing or gaze following effects in this paradigm, they may be short-lived and not be detected by merely collecting recognition responses, but may instead be better identified by means of implicit measures such as response times or pupil dilation. Further investigations of possible subtle effects of intonation and gaze on memory in this paradigm are necessary to understand under which conditions these factors may affect object memory in a task otherwise designed for probing mentalizing.

6.2 Perspective Taking in Written Text

6.2.1 Summary

In stories including more than one character, free indirect discourse (FID) requires the reader to identify the correct perspective for successful resolution. In study 2 (Zimmermann et al., 2021), naturalness ratings for written sentences including FID anchored to one of two protagonists of different prominence status (i.e. of protagonists that stood out in the story to different degree) as well as naturalness ratings for written sentences not including FID were assessed in autistic adults and a comparison group. Both groups rated sentences including FID as less natural than sentences not including FID. Moreover, both groups rated sentences including FID anchored to the more prominent character as more natural than sentences including FID anchored to the less prominent character. Importantly, no group differences due to an autism diagnosis were observed. These results suggest that the identification of perspectival centers for prominence-dependent FID anchoring in short stories is comparable in autistic adults with normal intelligence and non-autistic adults.

6.2.2 Free Indirect Discourse Perception

The overall pattern of naturalness ratings was in accordance with previous findings (Hinterwimmer & Meuser, 2019): Naturalness ratings were lower for sentences including FID compared to sentences not including FID. Additionally, naturalness ratings were lower for sentences including FID anchored to the less prominent protagonist compared to sentences including FID anchored to the more prominent protagonist.

In concordance with previous studies on explicit text-based second-order ToM abilities in autistic adults and adolescents (Bowler, 1992; Happé, 1994; Murray et al., 2017; Ponnet et al., 2004; Scheeren et al., 2013; Schuwerk et al., 2015), study 2 shows similar performance for the autistic and the non-autistic group. Arguably, the task in study 2 requires participants to implicitly take the perspective of a protagonist to successfully anchor FID to the most likely candidate. The similar rating behavior may suggest similar implicit perspective taking abilities of both groups. Study 2 did not include any further assessment of implicit or explicit ToM abilities. This limits the conclusions that can be drawn with regard to participants' perspective taking abilities. As it has been shown that perspective taking following certain FID markers is correlated with spatial perspective taking in the general population (Kaiser & Cohen, 2012), including measures of perspective taking (other than FID perception) in future studies may be informative. In other studies, implicit perspective taking has been shown to be attenuated in autistic adults compared

to non-autistic adults, even if explicit task performance is comparable. For example, when autistic participants are asked to explain their responses within the context of a strange stories task, they refer less to mental states and provide less reasons that would suggest perspective taking (Callenmark et al., 2014; Happé, 1994). When working on false-belief tasks, anticipatory gaze of autistic observers suggests a lacking or decreased implicit perspective taking compared to non-autistic observers (Schneider et al., 2013; Schuwerk et al., 2015; Senju et al., 2009).

6.2.3 Open Research Questions

The naturalness ratings employed in study 2 offer a first glance at FID perception in autism which had not been investigated before. While comparable rating patterns in the autistic and non-autistic group suggest similar FID processing, it must be noted that (i) FID anchoring does not necessarily require (implicit) perspective taking, which is particularly relevant within the context of the current thesis, and (ii) naturalness ratings are merely a superficial measure of FID anchoring processes that may still differ between autistic and non-autistic participants at other behavioral or non-behavioral levels.

Given the uncertainty about whether – and to what extent – perspective-taking is required for FID anchoring, the following discussion will explore both possibilities: that perspective-taking plays a role in FID processing (Section 6.2.3.1) and that it does not (Section 6.2.3.2), in relation to open research questions. A deeper understanding of the factors contributing to FID anchoring may help future research on FID anchoring in autism decide whether to focus primarily on verbal abilities or perspective taking abilities. In study 2, all participants had normal verbal intelligence. However, future studies may benefit from including a wider range of participants with varying verbal abilities to better explore the relationship between verbal skills and FID processing, both in the general population and in autism specifically.

Section 6.2.3.3 then examines how experimental methods beyond naturalness ratings can offer deeper insights into FID processing in both autistic and non-autistic participants.

6.2.3.1 What if Free Indirect Discourse Involves Perspective Taking?

It has been argued (Zeman, 2017) that FID processing and perspective taking required to pass false belief tasks, commonly used to assess ToM abilities, are similar in the sense that they draw on the same two fundamental components: (i) the ability to maintain different viewpoints at the same time, and (ii) the integration of these viewpoints from an external third viewpoint.

Shifting and integrating different perspectives requires the inhibition of one or more potentially competing perspectives (Frith & de Vignemont, 2005; MacWhinney, 2000). Equivalently, the

prerequisite for FID processing and ToM formulated above (Zeman, 2017), i.e. the ability to simultaneously maintain different viewpoints and integrating these from an external viewpoint requires sufficient basic executive function in terms of working memory capacity (Lin et al., 2010; Wardlow, 2013) and inhibitory control (Brown-Schmidt, 2009; Wardlow, 2013). Both working memory and inhibitory control have been shown to be challenging for autistic people (Demetriou et al., 2018; Habib et al., 2019). It has also been suggested to be particularly difficult for autistic participants to shift attention from one perspective to another (Schwarzkopf et al., 2014). The exploratory analysis of data obtained in study 2 suggests that autistic participants, that reported more difficulties with attention switching in the AQ questionnaire, showed smaller differences in their naturalness ratings between sentences with FID anchored to the more prominent protagonist and sentences with FID anchored to the less prominent protagonist. If perspective taking is required for FID processing and attention switching is linked to perspectival shifts, why do difficulties in attentional switches lead to this decrease in ratings in the autistic group? May difficulties switching attention be associated with decreased sensitivity to different types of FID anchoring? May difficulties switching attention foster a more linguistic approach, which may be less or differently affected by violations of FID anchoring expectations? These questions cannot be answered with the results of study 2. Executive functions were not assessed in this study beyond the self-reported questionnaire items of the AQ. Assessing executive functions may, however, be an interesting idea for future studies. Increasing task demands with regard to executive functions may also be informative. For example, it may be worthwhile investigating if autistic readers may be affected differently by a task with higher task demands, as competing tasks demanding executive control hinder the correct selection of perspectives (Qureshi et al., 2010).

It can be argued that perspective taking in FID processing not only shares similarities with ToM but also with embodiment and perceptual perspective taking: There is reason to believe that perspective taking during reading encompasses embodiment at least to a certain degree. Readers from the general population not only acquire a memory and a mental representation of the text itself, they also create mental models of the scenario described by the text (Zwaan & Radvansky, 1998). This can entail that readers mentally represent spatial locations of, for instance, things and people within a story. Such a spatial grounding has been argued to form the basis for embodied action simulation (Beveridge & Pickering, 2013), i.e. for embodied mental representations of a protagonist's actions. A finding from neuroimaging demonstrates that imagining a contextual scenario while reading about motor actions performed by a protagonist referred to by third-person pronouns can indeed enhance embodiment, as opposed

to conditions preventing imagining a contextual scenario or conditions not including motor action (Tomasino et al., 2007). Spatial grounding seems to be particularly important for embodiment when reading stories including third-person protagonists (as opposed to when the reader is included in the story by referring to them as "you") (Gianelli et al., 2011). Instead of outlining spatial relations in each story to induce embodiment it is also possible to include action descriptions that already contain spatial directionality (Taylor et al., 2008; Zwaan & Taylor, 2006). It may be interesting to investigate FID anchoring in stories that are more likely to enhance embodiment. Embodiment difficulties were, moreover, proposed to explain challenges autistic adults faced in some perspective taking tasks, particularly with regard to spatial perspective taking (Pearson et al., 2013). As discussed above, at least coarse-grained spatial relations may be necessary for embodiment during reading, and embodiment may be enhanced for motor action descriptions. Therefore, testing new items comparable to those of study 2 in autism alongside measures of embodiment such as participants' body movements, brain activity or response times may be informative with respect to perspective taking processes during FID anchoring in autistic and non-autistic readers. Additionally, longer and more vivid stories may increase embodiment during reading (MacWhinney, 2000). It might help to understand whether or to what degree embodiment is involved in FID anchoring.

It has been shown that – irrespective of the impact of FID – readers from the general population differ with regard to their tendency to take the perspective of a protagonist. Three subgroups have been identified that differ with regard to their perspective taking tendencies (Bimpikou, 2023; Hartung et al., 2017; Vogels et al., 2021): (i) participants that tend to take the perspective of a protagonist, (ii) participants that do not tend take a protagonist's perspective, i.e. that, depending on the study, maintain their own point of view, the point of view of a bystander or of the narrator, and (iii) participants that do not show a consistent tendency. The study by Bimpikou (2023) showed that participants with a tendency to take the protagonist's perspective were influenced by FID markers. In this study, faster response times and mouse trajectories suggested that the presence of FID cues further increased the tendency of these participants (however not of participants without a consistent perspective taking tendency) to take the perspective of the protagonist. The studies mentioned above demonstrate a high interindividuality not only in perspective taking tendencies (Bimpikou, 2023; Hartung et al., 2017; Vogels et al., 2021) but also with regard to the susceptibility to influences of FID (Bimpikou, 2023). Taking a step back and investigating individual perspective taking tendencies in autistic readers – irrespective of FID involvement – may be a fruitful base for further assessing FID perception in behaviorally different subgroups.

Markers of FID do likely not all induce the same type or degree of perspective taking. In a study with participants from the general population, Kaiser and Cohen (Kaiser & Cohen, 2012) found that evaluative adjectives (e.g. "*Poor* girl.") in FID contexts led to stronger indications of perspective taking than adverbials of possibility or doubt (e.g. "He'd *probably* put toothpaste in the shampoo bottle again"). The authors suggest that the stronger influence of evaluative adjectives may be explained by the emotional component of this FID marker and a correspondingly more "empathic perspective taking", whereas adverbials of possibility or doubt may rely on reasoning about other's knowledge. They further showed a link of the former with spatial perspective taking abilities, i.e. participants with better spatial perspective taking performance showed higher sensitivity to evaluative adjectives. This link was not observed for adverbials of possibility or doubt. What does this entail for FID research in autism? Investigating FID processing as a general construct may be too coarse-grained to paint a comprehensive picture. Since different FID markers may be linked to different types of perspective taking, autistic readers might be influenced by them differently than non-autistic readers.

In a similar fashion, it is possible that autistic readers perceive different types of prominence differently. Prominence-lending cues increase the likelihood of text interpretation from the perspective of a protagonist that is the perspectival center, i.e. the most likely anchor for any upcoming FID. When processing text that includes FID, readers arguably implicitly take the perspective of the character that is the perspectival center and anchor FID to it. Prominence can, however, arise from different qualities that may differ with regard to how they are perceived by autistic readers. Additionally, they may differ with regard to their potential interaction with perspective taking. For example, both a character's agentivity as well as their newness are associated with comparably high levels of prominence (Himmelmann & Primus, 2015). However, while some evidence suggests that readers or listeners (Buccino et al., 2005) tend to embody a protagonist's agentivity (Tomasino et al., 2007), newness may be less likely to elicit embodiment or other forms of perspective taking. FID in the text material used in study 2 was rated more natural if it was anchored to the most prominent protagonist, which here was always achieved by the protagonist being the subject of the sentence and being referred to by their name and pronouns. It may be interesting to systematically vary different prominence-lending cues – for instance, agentivity – to investigate potential effects on embodiment or other types of perspective taking.

Study 2 was the first investigation of FID perception in autism. To pave the way for systematic future investigations, qualitative interviews may help learn more about autistic participants'

experience while processing FID, e.g. similarly to an investigation done on perspective taking in autism when processing story excerpts (Chapple et al., 2023). It would also be helpful to include measures and questionnaires that allow for the quantification of the degree of participants' identification and relation with the protagonist as well as an associated spatial perspective taking, e.g. similarly to the measures used in a study by Salem and colleagues (2017) investigating FID perception in the general population. For example, with regard to study 2, item presentation may be followed by an identification scale, and items similar to those used in study 2 may be created which include specific spatial relations between the protagonists, enabling participants to spatially situate themselves.

In conclusion, the processing of FID appears to involve a complex interplay of cognitive and perceptual factors, including executive functions and different types of perspective taking abilities such as mentalizing and embodiment. Future research may benefit from qualitative and quantitative investigations into individual differences in perspective taking tendencies, the influence of various FID markers, and the impact of different prominence-lending cues, in both autistic and non-autistic readers.

6.2.3.2 What if Free Indirect Discourse does not Involve Perspective Taking?

Do readers actually adopt a protagonist's perspective when processing FID? A study examining text material including FID and spatial anchors (Salem et al., 2017) found that FID did not significantly enhance spatial perspective taking, the level of identification with the protagonist, or participants' feelings of how strongly they felt themselves to be entering into a relation with the protagonist. However, these findings are limited, as participants in the FID condition also exhibited lower thematic interest in the text. Despite this, the question remains: how is FID processed if perspective taking is not involved and to what extent do autistic and non-autistic readers anchor FID to the perspectival center identified through linguistic cues? If participants use a cognitively conscious approach based on linguistic markers for FID anchoring, they need to identify indicators of FID as well as the prominent protagonist. Both require keeping track of common ground, i.e. the knowledge shared between two interlocutors – or between the narrator and the reader. Unsurprisingly, working memory has been shown to have a positive influence on common ground maintenance in autistic children and adolescents (Schuh et al., 2016). Not only in this study, but also in other studies (Habib et al., 2019), lower working memory capacity has been observed in the autistic group. May common ground maintenance be associated with greater difficulty in autism in the case of FID processing? The results from study 2 do not suggest this idea, however, studies employing more extensive and detailed

measures might. In future studies, it may therefore be helpful to assess working memory capacity to investigate a possible influence of working memory on FID processing.

Some evidence suggests that autistic and non-autistic participants are both proficient in tracking prominence relations: In autistic adolescents and adults with normal intelligence, shifting attention towards prominent entities seems to be intact (however, sometimes slightly delayed) both when anticipating information following spoken and written language (Brock et al., 2008; Tresh, 2016; Black et al., 2019). The study by Brock and colleagues (2008) investigated the perception of spoken sentences in autistic adolescents and a non-autistic group. No difference was found between the groups regarding anticipatory eye movements towards images of the denoted objects which were presented alongside distractor images. This finding suggests that the context is taken into account for the prediction of the following speech content as well as for an attentional shift towards more prominent, i.e. accessible, verbal content. This notion is corroborated by a study by Tresh (2016) including autistic adults, in which moreover no influence of mentalizing ability on the interpretation of future text content was shown. The identification of FID indicators may, as it follows specific rules, be arguably mastered with comparable ease by autistic and non-autistic participants with normal intelligence and sufficient language abilities. Taken together, it may be assumed that FID anchoring – both in autistic and non-autistic readers – may, in principle, be possible without perspective taking abilities.

In the context of other ToM tasks (Livingston & Happé, 2017) and visual perspective taking (Conson et al., 2015), autistic participants have been suggested to opt for more conscious strategies more often than non-autistic participants. It has further been proposed that autistic people may in general have a more analytic style of decision making (Martino et al., 2008) and use more deliberate, less intuitive reasoning than non-autistic people (Brosnan et al., 2016, 2017; Brosnan & Ashwin, 2023). Correspondingly, they may solve perspective taking tasks by applying learned (social) rules of cause and effect (van Tiel et al., 2021). Similarly, it is possible that anchoring FID may be more cognitively conscious in autistic readers compared to non-autistic readers.

6.2.3.3 Beyond Naturalness Ratings: Measures and Material

Irrespective of whether FID anchoring requires perspective taking or not, it may be informative to include experimental methods other than naturalness ratings in future studies. Some studies investigating autistic and non-autistic behavior in adults in second-order ToM tasks show comparable performance with regard to participants' explicit responses, but the diagnostic groups differ with regard to accuracy, response times (Bradford et al., 2018) or eye gaze

behavior (Schneider et al., 2013; Schuwerk et al., 2015; Senju et al., 2009). Both perspective taking and a linguistic approach to FID anchoring may be accompanied by processes not detectable by assessing naturalness ratings, as FID effects can be subtle and may not necessarily show in primary measures (cf. Bimpikou, 2023). Implicit measures such as response times, mouse trajectories, eye-tracking, neuroimaging or neurophysiological measures such as EEG (electroencephalography) may be employed with the idea of further understanding cognitive processing and identifying attentional differences in reaction to FID following different perspectival centers. This may allow for detecting possible subtle differences between individuals or between groups of autistic and non-autistic readers.

In study 2, self-paced reading would not only accommodate individual reading times but may also reveal differences with regard to response times between groups following FID indicators or following the pronouns used for FID anchoring. Three of the 24 participants in the autistic group reported problems responding within the given time limit. It may be interesting to investigate whether their reading process took longer overall, or whether they spent particularly much time on specific parts of the text. Likewise, eye-tracking may be used to assess fixation durations for text parts. Longer reading times or longer fixation durations may be a sign of conscious or more effortful FID processing or anchoring. Further, eye-tracking or EEG may reveal attentional processes, for example with those associated with surprise or violation of expectations which may be assessed via pupil dilation or electrophysiological measures in response to FID in general and to FID anchored to the less prominent or more prominent protagonist. The greater rating differences for FID vs. non-FID conditions as well as for high vs. low prominence conditions found in younger participants in study 2, as opposed to older participants, may be linked to lifetime text exposure and be associated with less surprisal in higher age.

Using more ecologically valid stimulus material, may be interesting as well. Creating text items including FID indicators can lead to stylistically clumsy and unnatural results. This could in part be avoided by using different types of FID indicators as well as potentially reducing FID indicators. In study 2, the three FID indicators where the same in all FID items: a temporal adverbial contrasting with the temporal adverbials in the first sentence (e.g. “now”), a verb in subjunctive II mode (e.g. “would”), and a colloquial term or qualitative noun (e.g. “clumsy oaf”). For instance, avoiding subjunctive II mode may yield less stylistically clumsy text material.

6.2.4 Ageing in Non-Autistic Readers: Decline of Rating Differences

It may be informative to further investigate the effect of age found in study 2: The rating differences for the sentences including FID and those not including FID, as well as the rating differences for the sentences including FID anchored to the more prominent protagonist and those including FID anchored to the less prominent protagonist decreased with age in the non-autistic group. This age effect was not observed in the autistic group. Further understanding this effect in non-autistic readers may not only shed light on FID processing in the general population but also in autism. The age effect may be mediated by other variables that could influence FID processing: In the general population, age is associated with a decline of cognitive abilities including language comprehension (Burke & Shafto, 2008), referential processing such as in anaphora resolution (Light & Capps, 1986), and tracking prominence relations (Hendriks et al., 2014). Moreover, performance in different ToM tasks decreases with age (Henry et al., 2013). On the other hand, a generally more positive mindset (Carstensen et al., 2010) in older participants, as well as changed attention allocation during reading and greater linguistic experience (Crocker & Keller, 2005), may lead to easier processing of FID and FID anchored to a less prominent protagonist in particular. Establishing whether these abilities and age-associated changes are directly associated with FID processing could putatively explain, why FID perception in study 2 was not affected by age in the autistic group, as opposed to the non-autistic group.

Age-related cognitive decline in autism and the non-autistic population follows a similar pattern (Howlin & Magiati, 2017). Some abilities are, however, affected differently by age. For instance, visual memory is less affected by higher age in autistic compared to non-autistic people (Lever et al., 2015). One explanation for age affecting ratings in study 2 in the non-autistic group only, could be a different lifetime trajectory of cognitive abilities responsible for FID processing in autistic and non-autistic readers. Another explanation could be a different approach for FID processing in both diagnostic groups. Further, both explanations may be at play at the same time. For example, ToM abilities decline with age in the general population to a level found in both young and old autistic people (Lever & Geurts, 2016; Zıvralı Yarar et al., 2021). A potential association of ToM abilities with the age-related effects found in study 2 may explain why the ratings of autistic participants were not affected by age. Additionally, as it has been shown that autistic children show difficulties in classic ToM tasks whereas adolescents and adults show similar performance to non-autistic participants, it may be interesting to investigate if this may also apply to FID processing, i.e. to investigate if autistic

children and adolescents perceive FID anchoring similar to non-autistic children and adolescents.

6.3 Clinical Implications

Study 1 demonstrates that, in both the autistic and non-autistic group, there is a high individuality among participants with regard to their use of the virtual character's eye gaze behavior and their use of the intonation of the utterance presented alongside. Especially with regard to pragmatic intonation perception in autism and individual differences, research has yet to paint a clearer picture. When compared to research on gaze and gaze perception in autism, research on the production and perception of intonation in autism is sparse. A better understanding and quantification of individual use of intonation in autism – both for production and perception – may, however, be informative for diagnostics and potential speech training. The ICD-11 (World Health Organization, 2019) considers pragmatic language deficits one of the core features of autism (as opposed to other language deficits that may or may not accompany autism), and with respect to prosody states that speech in autism may lack "normal prosody and emotional tone and therefore [may] appear […] monotonous". However, findings from studies investigating prosody production in autistic children and adults from different language contexts (Chan & To, 2016; Diehl et al., 2009; Sharda et al., 2010; Wehrle et al., 2022) suggest this may not generally be a suitable – or specific enough – description of speech in autism, which in these studies has been described as being melodic rather than monotonous, albeit not consistently for all autistic participants. Neither the DSM-5 (American Psychiatric Association, 2013) nor the ICD-11 (World Health Organization, 2019) characterize intonation perception in autism. It may, however, be an informative autistic feature. While emotional prosody perception is an often-documented area of difficulty in autism, the extent of potential challenges in pragmatic prosody perception is not as clear-cut and subject to individual differences. Individuality may therefore be key for assessing impairments or unique features in the autistic individual. One application may be an individual chart of prosody perception fields of strengths and weaknesses which may not only enhance systematic assessment but may also help informing speech training.

6.4 Conclusion and Outlook

In this thesis, the perception of prominence-related perspective taking processes was investigated in autistic adults in two different experimental approaches: (i) audiovisual perception of intonation and gaze duration associated with attentional focus and respective

mentalizing, and (ii) perception of FID in written short stories in which protagonists vary with respect to their quality of standing out.

Both study 1 and study 2 of this thesis investigated perspective taking which is a broad concept and subsumes different types of perspective taking. Study 1 was designed to assess conceptual perspective taking abilities required to infer other people's mental states based on their gaze behavior and speech intonation. Study 2 investigated a type of perspective taking that does not require inferring others' mental states but requires the interpretation of text from the perspective of one of the story's protagonists. The former type of perspective taking may be considered an ability classically assessed within the scope of ToM – or mentalizing –, while the latter shares similarities with mentalizing, embodiment and perceptual perspective taking.

Despite similar behavior in autistic adults compared to non-autistic adults in the studies constituting this thesis, the question remains whether they applied the same strategy to solve these tasks. As argued above, the tasks may allow for different strategies such as a more intuitive approach (which arguably would assess ToM and implicit perspective taking abilities) and a more strategic and rule-based approach. Potential perspective taking difficulties in the autistic group may be compensated by these latter overt and conscious strategies. The results of study 1 and study 2 of the current thesis do, however, not allow for the systematic investigation of the strategies participants applied. Establishing measures to achieve this may help mapping autistic behavior in these tasks.

These studies offer a valuable starting point for future investigations that should optimally apply more life-like scenarios since external validity is limited for both investigated domains. Additionally, including measures allowing for improved assessment of processes underlying participants' behavior such as neurophysiological measures could elucidate whether different strategies were applied by the different diagnostic groups to solve the tasks.

7 Summary

Compared to non-autistic participants, autistic adults with normal intelligence show mentalizing and perspective taking difficulties mostly in tasks in which mental states are to be inferred from nonverbal information or in implicit tasks. i.e. in tasks in which perspective taking is not explicitly required. Research has yet failed to clearly demarcate the extent of these difficulties. The studies included in this thesis add to this body of literature by identifying important communicative competences both in the nonverbal and verbal domain that appear to be preserved in autistic adults. They show that mentalizing and perspective taking abilities in autistic adults are similar to non-autistic participants in two different domains of perception of language: (i) nonverbal language, i.e. body language (eye gaze) and intonation (pitch height) (studies 1a–c), as well as (ii) written short stories including free indirect discourse (FID) (study 2).

The paradigm developed for studies 1a–c revealed robust and replicable results: The main findings obtained in an online study (study 1a) were successfully replicated in a controlled laboratory setting (studies 1b and 1c), both in an autistic and a non-autistic sample. Taken together, studies 1a–c demonstrate proficient use of prominence-lending cues (i.e. cues that help highlighting an element such as a syllable, an object or a protagonist) when perceiving intonation and gaze duration for mentalizing in autistic and non-autistic participants. In this paradigm, participants' mentalizing ability was assessed by their capacity to infer the importance of an object to a virtual character. Both diagnostic groups rated an object as more important to a virtual character if the character's gaze towards the object was longer or if the pitch accent of the utterance denoting the object was higher. Across both diagnostic groups, three behaviorally different subgroups could be identified: "Lookers" that primarily took into account the gaze cue for their ratings, "Listeners" that predominantly concentrated on the intonation cue, and "Neithers" whose ratings were not significantly influenced by either of these two cues. Notably, compared to the non-autistic group, the autistic group took gaze duration into account to a greater extent than intonation.

Using nonverbal cues for mentalizing has often been shown to be difficult for autistic participants. The design of the present study did not reveal such difficulties in the autistic group. However, it must be noted that the external validity is limited. Further, the reductionistic and highly structured task design in principle allows for different task solving strategies (or a mixture thereof): Either, participants rely on intuitive impressions, or they identify patterns and use these to solve the task. The tendency of autistic participants to attend to the gaze cue rather

than the pitch cue may be a byproduct of such a more systematic than intuitive task solving strategy. In upcoming studies, higher external validity could be achieved by including more naturalistic and less structured scenarios.

Study 2 investigated the perception of FID in short written stories, which is suggested to be accompanied by implicit perspective taking. More precisely, participants rated how natural they found the last sentence of a short story. Autistic and non-autistic readers perceived the target sentences similarly: (i) Sentences including FID were perceived as less natural compared to sentences not including FID, and (ii) sentences including FID that needed to be anchored to the more prominent of two protagonists were perceived as more natural than sentences including FID that needed to the anchored to the less prominent protagonist.

These results suggest similar FID processing in autistic adults compared to non-autistic adults. However, although the involvement of perspective taking in FID has been a subject of discussion, it has not yet been fully understood, nor has it been investigated extensively. Moreover, the conclusions drawn from this experiment are focused on naturalness ratings, which capture only one aspect of FID processing. Nevertheless, the findings provide an important foundation for exploring potential differences between autistic and non-autistic readers of FID at other (behavioral or physiological) levels.

The studies in this thesis add to the body of literature on mentalizing and perspective taking in adults with autism as well as to research on prominence in the perception of language and general communication. The main findings of both investigated domains – nonverbal language (i.e. gaze and intonation) and FID perception – do not support the idea that autistic participants faced problems with perspective taking in these studies. As both tasks allowed for different task-solving strategies, further investigations of the behavior (in autism) in these tasks is necessary to understand (i) which role mentalizing and perspective taking play in these tasks, and (ii) to what extent autistic and non-autistic participants behave similarly in these tasks.

8 Zusammenfassung

Autistische Erwachsene (ohne Intelligenzeinschränkungen) haben im Vergleich zu nicht-autistischen Personen häufig Schwierigkeiten, mentale Zustände anderer Personen zu erfassen und deren Perspektive einzunehmen. Dies gilt insbesondere für Aufgaben, bei denen mentale Zustände aus nonverbalen Informationen ausgelesen werden sollen oder für implizite Aufgaben, d.h. Aufgaben, die eine Perspektivübernahme nicht explizit erfordern. Das Ausmaß dieser Schwierigkeiten ist in der bisherigen Forschung noch nicht klar herausgearbeitet worden. Die in dieser Dissertation vorgestellten Studien erweitern die bestehende Literatur, indem sie wichtige kommunikative Kompetenzen sowohl im nonverbalen als auch im verbalen Bereich identifizieren, die bei autistischen Erwachsenen unbeeinträchtigt zu sein scheinen. Insbesondere verdeutlichen sie, dass das Verhalten hinsichtlich der Perspektivübernahme und der Zuschreibung mentaler Zustände einer anderen Person bei autistischen Erwachsenen in zwei verschiedenen Bereichen der Sprachwahrnehmung ähnlich wie bei nicht-autistischen Teilnehmenden ausfällt: (i) nonverbale Sprache, was sich hier zum einen auf Körpersprache – speziell auf Blickverhalten – bezieht, sowie auf Intonation (Studien 1a–c), und (ii) freie indirekte Rede (auch: erlebte Rede) in schriftlich präsentierten Kurzgeschichten (Studie 2).

Das für die Studien 1a–c entwickelte Paradigma zeigte robuste und replizierbare Ergebnisse: Die zentralen Erkenntnisse der Online-Studie (Studie 1a) konnten in einer kontrollierten Laborumgebung (Studien 1b und 1c) sowohl mit autistischen als auch mit nicht-autistischen Teilnehmenden erfolgreich repliziert werden. Zusammengenommen zeigen die Studien 1a–c, dass autistische und nicht-autistische Studienteilnehmende Prominenz-Hinweisreize – also Signale, die ein Element wie eine Silbe, ein Objekt oder eine*n Protagonist*in hervorheben – gekonnt nutzten, um Intonation und Blickdauer im Hinblick auf den mentalen Zustand einer anderen Person zu interpretieren. Die Bedeutung eines Objekts für eine virtuelle Figur wurde von beiden Diagnosegruppen höher eingeschätzt, wenn der Blick der virtuellen Figur länger auf dem Objekt verweilte oder der Tonakzent der jeweiligen Äußerung, die das Objekt bezeichnete, höher war. Über beide Diagnosegruppen hinweg konnten drei verhaltensspezifische Untergruppen identifiziert werden: „Lookers“ (Schauende), die primär die Blickdauer der virtuellen Figur für ihre Bewertungen heranzogen, „Listeners“ (Hörende), die sich vor allem auf die Intonation fokussierten, und „Neithers“ (weder/noch), deren Bewertungen von keinem der beiden Faktoren signifikant beeinflusst wurden. Interessanterweise legte die autistische Gruppe im Vergleich zur nicht-autistischen Gruppe mehr Gewicht auf die Blickdauer als auf die Intonation.

Die Nutzung nonverbaler Informationen für das Auslesen mentaler Zustände anderer Personen stellt für autistische Menschen häufig eine Herausforderung dar. Das vorliegende Studiendesign zeigte keine derartigen Schwierigkeiten in der autistischen Gruppe. Die externe Validität der Ergebnisse ist jedoch begrenzt. Darüber hinaus lässt das stark strukturierte und reduktionistische Aufgabendesign unterschiedliche Lösungsstrategien zu (oder eine Kombination aus diesen): Die Teilnehmenden können sich entweder auf intuitive Eindrücke verlassen oder gezielt Muster erkennen, um diese zur Bewältigung der Aufgabe zu nutzen. Die Beobachtung, dass autistische Teilnehmende eher die Blickdauer als die Tonhöhe berücksichtigen, könnte das Ergebnis einer solchen systematischen Aufgabenlösungsstrategie sein. Um in zukünftigen Studien eine höhere externe Validität zu erreichen, sollten naturalistischere und weniger stark strukturierte Szenarien einbezogen werden.

In Studie 2 wurde die Wahrnehmung von freier indirekter Rede in schriftlich präsentierten Kurzgeschichten untersucht, die mit impliziter Perspektivübernahme einhergehen soll. Konkret bewerteten die Teilnehmenden, wie natürlich sie den letzten Satz einer Kurzgeschichte empfanden. Autistische und nicht-autistische Lesende nahmen die entsprechenden Sätze in ähnlicher Weise wahr: (i) Sätze mit freier indirekter Rede wurden im Vergleich zu Sätzen ohne freie indirekte Rede als weniger natürlich empfunden, und (ii) Sätze mit freier indirekter Rede, die sich auf den/die prominentere/n von zwei ProtagonistInnen bezogen, wurden als natürlicher wahrgenommen als solche, die sich auf den/die weniger prominente/n ProtagonistIn bezogen.

Die Ergebnisse weisen darauf hin, dass die Verarbeitung von freier indirekter Rede bei autistischen und nicht-autistischen Erwachsenen ähnlich verläuft. Die Rolle der Perspektivübernahme bei der Verarbeitung von freier indirekter Rede ist jedoch, obwohl sie bereits in der Fachliteratur thematisiert wurde, bislang weder vollständig verstanden noch umfassend untersucht worden. Darüber hinaus konzentrieren sich die Schlussfolgerungen aus diesem Experiment auf die Bewertung der Natürlichkeit, die lediglich einen Aspekt der Verarbeitung freier indirekter Rede abbildet. Dennoch stellen die Ergebnisse eine wichtige Grundlage dar, um mögliche Unterschiede zwischen autistischen und nicht-autistischen Lesenden von freier indirekter Rede auf anderen Ebenen – sei es verhaltensbezogen oder physiologisch – weiter zu erforschen.

Die in dieser Arbeit vorgestellten Studien erweitern die bestehende Literatur zur Mentalisierungsfähigkeit bzw. zur Perspektivübernahme bei Erwachsenen mit Autismus sowie zur Prominenz in der Sprachwahrnehmung und in der Kommunikation im Allgemeinen. Die Hauptergebnisse aus den beiden untersuchten Bereichen – nonverbale Sprache (hier: Blick und

Intonation) und die Wahrnehmung von freier indirekter Rede – sprechen nicht dafür, dass autistische Teilnehmende in diesen Studien Schwierigkeiten mit der Perspektivübernahme hatten. Da beide Aufgaben unterschiedliche Lösungsstrategien ermöglichen, sind weitere Untersuchungen erforderlich, um (i) die Rolle von Mentalisierungsprozessen und Perspektivübernahme in diesen Aufgaben besser zu verstehen und (ii) die Gemeinsamkeiten und Unterschiede im Verhalten von autistischen und nicht-autistischen Teilnehmenden in diesen Aufgaben systematisch zu vergleichen.

9 Appendix

9.1 Supplementary Material for Studies 1b and 1c

The supplementary material comprises details on the auditory stimulus pretest, the complete list of object names alongside their English translations as well as the complete video and audio material. As the video and audio material is not included in this thesis, the reader is referred to the supplementary material online: https://www.frontiersin.org/articles/10.3389/fcomm.2024.1483135/full#supplementary-material.

9.1.1 Perceptual Pretest of Auditory Stimuli

Auditory stimuli were submitted to a perceptual pretest for "similarity", "naturalness" and accent type by six trained phoneticians. More precisely, they rated (i) if the original utterances sounded the same as the stimuli produced for the "low" condition, (ii) if the stimuli produced for the "low" and "high" condition sounded like natural utterances, and (iii) if the stimuli exhibited an H*-accent, signaling broad focus, or if the stimuli exhibited an L+H*-accent, signaling contrastive focus, or neither.

Stimuli that entered the "low" condition and the respective original utterances were rated as sounding the same in 86.76 % of cases. No stimulus was rated "the same" by less than 50 % of the raters. Stimuli produced for the "low" and "high" condition were rated as sounding natural in 92.39 % and 75.00 % of cases, respectively. No stimulus was rated "natural" by less than 50 % of the raters. Stimuli produced for the "low" and "high" condition were rated as H* and L+H*, respectively, in 83.70 % and 78.99 % of cases. No stimulus was rated to be of the intended accent-type by less than 50 % of the raters.

9.1.2 List of Object Names and their English Translations

Trials	Object Name	English Translation
Practice Trials	Eule	Owl
	Hocker	Stool
	Puppe	Doll
	Schlüssel	Key
Experimental Trials	Affe	Monkey
	Ampel	Traffic light
	Anker	Anchor
	Apfel	Apple
	Auto	Car
	Besen	Broom
	Biene	Bee
	Birne	Pear
	Bleistift	Pencil
	Blume	Flower
	Bluse	Blouse
	Brille	Glasses

Trials	Object Name	English Translation
Experimental Trials	Brunnen	Well
	Bürste	Brush
	Drachen	Kite
	Eisbär	Polar bear
	Ente	Duck
	Erdnuss	Peanut
	Esel	Donkey
	Fahne	Flag
	Fahrrad	Bicycle
	Fenster	Window
	Flasche	Bottle
	Fliege	Fly
	Flugzeug	Airplane
	Gabel	Fork
	Geige	Violin
	Glocke	Bell
	Gürtel	Belt
	Hammer	Hammer
	Handschuh	Glove
	Harfe	Harp
	Hase	Rabbit
	Hose	Pants
	Käfer	Beetle
	Katze	Cat
	Kerze	Candle
	Kette	Necklace
	Kirche	Church
	Koffer	Suitcase
	Kreisel	Spinning top
	Krone	Crown
	Kühlschrank	Refrigerator
	Kürbis	Pumpkin
	Lampe	Lamp
	Leiter	Ladder
	Löffel	Spoon
	Löwe	Lion
	Mantel	Coat
	Meißel	Chisel
	Messer	Knife
	Mütze	Cap
	Nadel	Needle
	Nagel	Nail
	Nashorn	Rhinoceros
	Pfanne	Pan
	Pfeife	Pipe
	Pfirsich	Peach
	Pinsel	Paintbrush
	Raupe	Caterpillar

Trials	Object Name	English Translation
Experimental Trials	Rollschuh	Roller skate
	Säge	Saw
	Sandwich	Sandwich
	Schachtel	Box
	Schere	Scissors
	Schlange	Snake
	Schleife	Bow
	Schneemann	Snowman
	Schraube	Screw
	Schreibtisch	Desk
	Schüssel	Bowl
	Socke	Sock
	Sofa	Sofa
	Sonne	Sun
	Spargel	Asparagus
	Spinnrad	Spinning wheel
	Stiefel	Boot
	Tasse	Cup
	Tiger	Tiger
	Toaster	Toaster
	Torte	Cake
	Trommel	Drum
	Türgriff	Doorknob
	Vase	Vase
	Vogel	Bird
	Waschbär	Raccoon
	Weste	Vest
	Wolke	Cloud
	Zange	Pliers
	Zebra	Zebra
	Ziege	Goat
	Zwiebel	Onion

9.2 Supplementary Material for Study 2

The supplementary material comprises the complete list of items used in study 2.

This list can also be found online: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2021.675633/full#supplementary-material

Item/Filler	Item/Filler Number	Condition	Sentence Number	Sentence
Item	1	A	1	Vor zwei Jahren buchte Georg Flugtickets für den Urlaub nach Griechenland.
			2	Er hatte seiner Freundin schon lange einen schönen Sommerurlaub versprochen.
			3	Schon in wenigen Wochen würde er mit seiner Süßen in der Sonne liegen.
Item	1	B	1	Vor zwei Jahren buchte Leonie Flugtickets für den Urlaub nach Griechenland.
			2	Sie hatte ihrem Freund schon lange einen schönen Sommerurlaub versprochen.
			3	Schon in wenigen Wochen würde er mit seiner Süßen in der Sonne liegen.
Item	1	C	1	Vor zwei Jahren buchte Leonie Flugtickets für den Urlaub nach Griechenland.
			2	Sie hatte ihrem Freund schon lange einen schönen Sommerurlaub versprochen.
			3	Wenige Wochen später lag er mit ihr in der Sonne.
Item	1	D	1	Vor zwei Jahren buchte Leonie Flugtickets für den Urlaub nach Griechenland.
			2	Sie hatte ihrem Freund schon lange einen schönen Sommerurlaub versprochen.
			3	Wenige Wochen später lag sie mit ihm in der Sonne.
Item	2	A	1	Nach dem Bestehen der Meisterprüfung bekam Nina endlich ein höheres Gehalt.
			2	Deswegen konnte sie nach langem Warten mit ihrem Freund in eine eigene Wohnung ziehen.
			3	Noch heute würde sie sich mit ihrem Liebsten auf Wohnungssuche begeben.
Item	2	B	1	Nach dem Bestehen der Meisterprüfung bekam Mirko endlich ein höheres Gehalt.
			2	Deswegen konnte er nach langem Warten mit seiner Freundin in eine eigene Wohnung ziehen.
			3	Noch heute würde sie sich mit ihrem Liebsten auf Wohnungssuche begeben.
Item	2	C	1	Nach dem Bestehen der Meisterprüfung bekam Mirko endlich ein höheres Gehalt.
			2	Deswegen konnte er nach langem Warten mit seiner Freundin in eine eigene Wohnung ziehen.
			3	Sie wollte sich zeitnah mit ihm auf Wohnungssuche begeben.
Item	2	D	1	Nach dem Bestehen der Meisterprüfung bekam Mirko endlich ein höheres Gehalt.
			2	Deswegen konnte er nach langem Warten mit seiner Freundin in eine eigene Wohnung ziehen.
			3	Er wollte sich zeitnah mit ihr auf Wohnungssuche begeben.
Item	3	A	1	Letzten Winter fuhr Miro zum ersten Mal in den Skiurlaub.
			2	Er wohnte mit seiner Frau in einer richtig gemütlichen Hütte.
			3	Morgen früh schon würde er mit seiner Liebsten die Piste unsicher machen.
Item	3	B	1	Letzten Winter fuhr Olga zum ersten Mal in den Skiurlaub.
			2	Sie wohnte mit ihrem Mann in einer richtig gemütlichen Hütte.
			3	Morgen früh schon würde er mit seiner Liebsten die Piste unsicher machen.
Item	3	C	1	Letzten Winter fuhr Olga zum ersten Mal in den Skiurlaub.
			2	Sie wohnte mit ihrem Mann in einer richtig gemütlichen Hütte.
			3	Bereits am ersten Morgen machte er mit ihr die Piste unsicher.
Item	3	D	1	Letzten Winter fuhr Olga zum ersten Mal in den Skiurlaub.
			2	Sie wohnte mit ihrem Mann in einer richtig gemütlichen Hütte.
			3	Bereits am ersten Morgen machte sie mit ihm die Piste unsicher.

Item/Filler	Item/Filler Number	Condition	Sentence Number	Sentence
Item	4	A	1	Letzte Woche geriet Moritz auf einem Schulausflug in Rage.
			2	Während der gesamten Busfahrt stritt er mit seiner Sitznachbarin.
			3	Jetzt gleich würde er dieser Wahnsinnigen eine scheuern.
Item	4	B	1	Letzte Woche geriet Lotta auf einem Schulausflug in Rage.
			2	Während der gesamten Busfahrt stritt sie mit ihrem Sitznachbarn.
			3	Jetzt gleich würde er dieser Wahnsinnigen eine scheuern.
Item	4	C	1	Letzte Woche geriet Lotta auf einem Schulausflug in Rage.
			2	Während der gesamten Busfahrt stritt sie mit ihrem Sitznachbarn.
			3	Irgendwann scheuerte er ihr eine.
Item	4	D	1	Letzte Woche geriet Lotta auf einem Schulausflug in Rage.
			2	Während der gesamten Busfahrt stritt sie mit ihrem Sitznachbarn.
			3	Irgendwann scheuerte sie ihm eine.
Item	5	A	1	In der letzten Deutschklausur wollte Hanna unbedingt Extrapunkte ergattern.
			2	Deshalb bat sie ihren Klassenkameraden um Hilfe.
			3	Jetzt gleich müsste sie sich mit diesem Streber treffen.
Item	5	B	1	In der letzten Deutschklausur wollte Sebastian unbedingt Extrapunkte ergattern.
			2	Deshalb bat er seine Klassenkameradin um Hilfe.
			3	Jetzt gleich müsste sie sich mit diesem Streber treffen.
Item	5	C	1	In der letzten Deutschklausur wollte Sebastian unbedingt Extrapunkte ergattern.
			2	Deshalb bat er seine Klassenkameradin um Hilfe.
			3	Sie traf sich mit ihm.
Item	5	D	1	In der letzten Deutschklausur wollte Sebastian unbedingt Extrapunkte ergattern.
			2	Deshalb bat er seine Klassenkameradin um Hilfe.
			3	Er traf sich mit ihr.
Item	6	A	1	Als 2009 ein Börsencrash drohte, legte Judith noch schnell veruntreute Firmengelder in einer Steueroase an.
			2	Anschließend verließ sie mit ihrem Geliebten das Land.
			3	Schon morgen würde sie mit ihrem Schatz in Panama ein neues Leben beginnen.
Item	6	B	1	Als 2009 ein Börsencrash drohte, legte Jan noch schnell veruntreute Firmengelder in einer Steueroase an.
			2	Anschließend verließ er mit seiner Geliebten das Land.
			3	Schon morgen würde sie mit ihrem Schatz in Panama ein neues Leben beginnen.
Item	6	C	1	Als 2009 ein Börsencrash drohte, legte Jan noch schnell veruntreute Firmengelder in einer Steueroase an.
			2	Anschließend verließ er mit seiner Geliebten das Land.
			3	Am nächsten Tag begann sie mit ihm in Panama ein neues Leben.
Item	6	D	1	Als 2009 ein Börsencrash drohte, legte Jan noch schnell veruntreute Firmengelder in einer Steueroase an.
			2	Anschließend verließ er mit seiner Geliebten das Land.
			3	Am nächsten Tag begann er mit ihr in Panama ein neues Leben.

Item/Filler	Item/Filler Number	Condition	Sentence Number	Sentence
Item	7	A	1	Nach jahrelangem Spielen gewann Tanja vor zwei Monaten endlich im Lotto.
			2	Sie teilte die gute Nachricht sofort ihrem Mann mit.
			3	Heute Abend noch würde sie mit ihrem Liebsten darauf anstoßen.
Item	7	B	1	Nach jahrelangem Spielen gewann Werner vor zwei Monaten endlich im Lotto.
			2	Er teilte die gute Nachricht sofort seiner Frau mit.
			3	Heute Abend noch würde sie mit ihrem Liebsten darauf anstoßen.
Item	7	C	1	Nach jahrelangem Spielen gewann Werner vor zwei Monaten endlich im Lotto.
			2	Er teilte die gute Nachricht sofort seiner Frau mit.
			3	Am Abend wollte sie mit ihm darauf anstoßen.
Item	7	D	1	Nach jahrelangem Spielen gewann Werner vor zwei Monaten endlich im Lotto.
			2	Er teilte die gute Nachricht sofort seiner Frau mit.
			3	Am Abend wollte er mit ihr darauf anstoßen.
Item	8	A	1	Vor einem Jahr durfte Lisa zum ersten Mal eine richtig große Rolle in einem Kinofilm spielen.
			2	Am Tag der Premiere stritt sie mit ihrem Manager über die Beteiligung an der Gage.
			3	Jetzt gleich würde sie die Angelegenheit mit diesem gierigen Idioten klären.
Item	8	B	1	Vor einem Jahr durfte Jimmy zum ersten Mal eine richtig große Rolle in einem Kinofilm spielen.
			2	Am Tag der Premiere stritt er mit seiner Managerin über die Beteiligung an der Gage.
			3	Jetzt gleich würde sie die Angelegenheit mit diesem gierigen Idioten klären.
Item	8	C	1	Vor einem Jahr durfte Jimmy zum ersten Mal eine richtig große Rolle in einem Kinofilm spielen.
			2	Am Tag der Premiere stritt er mit seiner Managerin über die Beteiligung an der Gage.
			3	Sie wollte die Angelegenheit umgehend mit ihm klären.
Item	8	D	1	Vor einem Jahr durfte Jimmy zum ersten Mal eine richtig große Rolle in einem Kinofilm spielen.
			2	Am Tag der Premiere stritt er mit seiner Managerin über die Beteiligung an der Gage.
			3	Er wollte die Angelegenheit umgehend mit ihr klären.
Item	9	A	1	Im Jahr 2010 erreichte Tom das Finale des Buchstabier-Wettbewerbs.
			2	Dort traf er auf seine Erzrivalin.
			3	Jetzt würde er es dieser Klugscheißerin zeigen.
Item	9	B	1	Im Jahr 2010 erreichte Julia das Finale des Buchstabier-Wettbewerbs.
			2	Dort traf sie auf ihren Erzrivalen.
			3	Jetzt würde er es dieser Klugscheißerin zeigen.
Item	9	C	1	Im Jahr 2010 erreichte Julia das Finale des Buchstabier-Wettbewerbs.
			2	Dort traf sie auf ihren Erzrivalen.
			3	Dieses Mal wollte er es ihr zeigen.
Item	9	D	1	Im Jahr 2010 erreichte Julia das Finale des Buchstabier-Wettbewerbs.
			2	Dort traf sie auf ihren Erzrivalen.
			3	Dieses Mal wollte sie es ihm zeigen.

Item/Filler	Item/Filler Number	Condition	Sentence Number	Sentence
Item	10	A	1	Vor drei Monaten machte sich Henning auf den Weg zu einem Vorstellungsgespräch.
			2	Dort musste er sich gegen seine ehemalige Kollegin behaupten.
			3	Heute würde er diese Schleimerin locker ausstechen.
Item	10	B	1	Vor drei Monaten machte sich Alexandra auf den Weg zu einem Vorstellungsgespräch.
			2	Dort musste sie sich gegen ihren ehemaligen Kollegen behaupten.
			3	Heute würde er diese Schleimerin locker ausstechen.
Item	10	C	1	Vor drei Monaten machte sich Alexandra auf den Weg zu einem Vorstellungsgespräch.
			2	Dort musste sie sich gegen ihren ehemaligen Kollegen behaupten.
			3	Dieses Mal konnte er sie ausstechen.
Item	10	D	1	Vor drei Monaten machte sich Alexandra auf den Weg zu einem Vorstellungsgespräch.
			2	Dort musste sie sich gegen ihren ehemaligen Kollegen behaupten.
			3	Dieses Mal konnte sie ihn ausstechen.
Item	11	A	1	Vor zwei Monaten meldete sich Kai auf einer Internetplattform zur Partnersuche an.
			2	Bereits nach wenigen Tagen schwärmte er für seine Chatpartnerin.
			3	Noch heute Abend würde er endlich seine Traumfrau treffen.
Item	11	B	1	Vor zwei Monaten meldete sich Christina auf einer Internetplattform zur Partnersuche an.
			2	Bereits nach wenigen Tagen schwärmte sie für ihren Chatpartner.
			3	Noch heute Abend würde er endlich seine Traumfrau treffen.
Item	11	C	1	Vor zwei Monaten meldete sich Christina auf einer Internetplattform zur Partnersuche an.
			2	Bereits nach wenigen Tagen schwärmte sie für ihren Chatpartner.
			3	Sie traf ihn noch am selben Abend.
Item	11	D	1	Vor zwei Monaten meldete sich Christina auf einer Internetplattform zur Partnersuche an.
			2	Bereits nach wenigen Tagen schwärmte sie für ihren Chatpartner.
			3	Er traf sie noch am selben Abend.
Item	12	A	1	Vor einigen Jahren kandidierte Petra bei den Kommunalwahlen.
			2	Am Tag der Stimmenauszählung traf sie auf ihren Kontrahenten.
			3	Heute würde sie diesen Heuchler bloßstellen.
Item	12	B	1	Vor einigen Jahren kandidierte Thorsten bei den Kommunalwahlen.
			2	Am Tag der Stimmenauszählung traf er auf seine Kontrahentin.
			3	Heute würde sie diesen Heuchler bloßstellen.
Item	12	C	1	Vor einigen Jahren kandidierte Thorsten bei den Kommunalwahlen.
			2	Am Tag der Stimmenauszählung traf er auf seine Kontrahentin.
			3	Sie stellte ihn bei der Diskussion bloß.
Item	12	D	1	Vor einigen Jahren kandidierte Thorsten bei den Kommunalwahlen.
			2	Am Tag der Stimmenauszählung traf er auf seine Kontrahentin.
			3	Er stellte sie bei der Diskussion bloß.

Item/Filler	Item/Filler Number	Condition	Sentence Number	Sentence
Item	13	A	1	Vor zwei Jahren wechselte Hannah den Job.
			2	Schon nach wenigen Tagen diskutierte sie mit ihrem neuen Arbeitskollegen wegen der Aufgabenverteilung.
			3	Jetzt gleich würde sie diesem Faulpelz die Meinung geigen.
Item	13	B	1	Vor zwei Jahren wechselte Steffen den Job.
			2	Schon nach wenigen Tagen diskutierte er mit seiner neuen Arbeitskollegin wegen der Aufgabenverteilung.
			3	Jetzt gleich würde sie diesem Faulpelz die Meinung geigen.
Item	13	C	1	Vor zwei Jahren wechselte Steffen den Job.
			2	Schon nach wenigen Tagen diskutierte er mit seiner neuen Arbeitskollegin wegen der Aufgabenverteilung.
			3	Sie wies ihn daraufhin zurecht.
Item	13	D	1	Vor zwei Jahren wechselte Steffen den Job.
			2	Schon nach wenigen Tagen diskutierte er mit seiner neuen Arbeitskollegin wegen der Aufgabenverteilung.
			3	Er wies sie daraufhin zurecht.
Item	14	A	1	Vor einigen Jahren musste Maria unbedingt abnehmen.
			2	Sie überredete ihren Freund zum Mitmachen.
			3	Nächstes Wochenende wollte sie mit diesem Stubenhocker ein Fitnessstudio besuchen.
Item	14	B	1	Vor einigen Jahren musste Niklas unbedingt abnehmen.
			2	Er überredete seine Freundin zum Mitmachen.
			3	Nächstes Wochenende wollte sie mit diesem Stubenhocker ein Fitnessstudio besuchen.
Item	14	C	1	Vor einigen Jahren musste Niklas unbedingt abnehmen.
			2	Er überredete seine Freundin zum Mitmachen.
			3	Am folgenden Wochenende wollte sie mit ihm ein Fitnessstudio besuchen.
Item	14	D	1	Vor einigen Jahren musste Niklas unbedingt abnehmen.
			2	Er überredete seine Freundin zum Mitmachen.
			3	Am folgenden Wochenende wollte er mit ihr ein Fitnessstudio besuchen.
Item	15	A	1	Zu Weihnachten bekam Simon ein neues Wissensspiel geschenkt.
			2	Er lud seine Kollegin zu einer Partie ein.
			3	Jetzt würde er diese Besserwisserin haushoch besiegen.
Item	15	B	1	Zu Weihnachten bekam Tabea ein neues Wissensspiel geschenkt.
			2	Sie lud ihren Kollegen zu einer Partie ein.
			3	Jetzt würde er diese Besserwisserin haushoch besiegen.
Item	15	C	1	Zu Weihnachten bekam Tabea ein neues Wissensspiel geschenkt.
			2	Sie lud ihren Kollegen zu einer Partie ein.
			3	Sie besiegte ihn haushoch.
Item	15	D	1	Zu Weihnachten bekam Tabea ein neues Wissensspiel geschenkt.
			2	Sie lud ihren Kollegen zu einer Partie ein.
			3	Er besiegte sie haushoch.

Item/Filler	Item/Filler Number	Condition	Sentence Number	Sentence
Item	16	A	1	Im Jahr 2012 geriet Manuel in einen Rechtsstreit.
			2	Er kämpfte mit seiner Nachbarin um die korrekte Festlegung der Grundstücksgrenze.
			3	Jetzt würde er diese Furie vor Gericht fertigmachen.
Item	16	B	1	Im Jahr 2012 geriet Paula in einen Rechtsstreit.
			2	Sie kämpfte mit ihrem Nachbarn um die korrekte Festlegung der Grundstücksgrenze.
			3	Jetzt würde er diese Furie vor Gericht fertigmachen.
Item	16	C	1	Im Jahr 2012 geriet Paula in einen Rechtsstreit.
			2	Sie kämpfte mit ihrem Nachbarn um die korrekte Festlegung der Grundstücksgrenze.
			3	Er schlug sie vor Gericht.
Item	16	D	1	Im Jahr 2012 geriet Paula in einen Rechtsstreit.
			2	Sie kämpfte mit ihrem Nachbarn um die korrekte Festlegung der Grundstücksgrenze.
			3	Sie schlug ihn vor Gericht.
Item	17	A	1	Im Jahr 2015 nahm Elias an einem Gesangswettbewerb teil.
			2	Am Finaltag geriet er mit seiner Duettpartnerin aneinander.
			3	Jetzt würde er diese Diva um Verzeihung bitten müssen.
Item	17	B	1	Im Jahr 2015 nahm Nadine an einem Gesangswettbewerb teil.
			2	Am Finaltag geriet sie mit ihrem Duettpartner aneinander.
			3	Jetzt würde er diese Diva um Verzeihung bitten müssen.
Item	17	C	1	Im Jahr 2015 nahm Nadine an einem Gesangswettbewerb teil.
			2	Am Finaltag geriet sie mit ihrem Duettpartner aneinander.
			3	Er bat sie bald darauf um Verzeihung.
Item	17	D	1	Im Jahr 2015 nahm Nadine an einem Gesangswettbewerb teil.
			2	Am Finaltag geriet sie mit ihrem Duettpartner aneinander.
			3	Sie bat ihn bald darauf um Verzeihung.
Item	18	A	1	Im Frühjahr bekam Henning die Nebenkostenabrechnung.
			2	Er diskutierte mit seiner Freundin über die hohen Heizkosten.
			3	Jetzt müsste er dieser Frostbeule sagen, dass es so nicht weitergeht.
Item	18	B	1	Im Frühjahr bekam Eva die Nebenkostenabrechnung.
			2	Sie diskutierte mit ihrem Freund über die hohen Heizkosten.
			3	Jetzt müsste er dieser Frostbeule sagen, dass es so nicht weitergeht.
Item	18	C	1	Im Frühjahr bekam Eva die Nebenkostenabrechnung.
			2	Sie diskutierte mit ihrem Freund über die hohen Heizkosten.
			3	Er sagte ihr, dass es so nicht weitergeht.
Item	18	D	1	Im Frühjahr bekam Eva die Nebenkostenabrechnung.
			2	Sie diskutierte mit ihrem Freund über die hohen Heizkosten.
			3	Sie sagte ihm, dass es so nicht weitergeht.

Item/Filler	Item/Filler Number	Condition	Sentence Number	Sentence
Item	19	A	1	Im Sommer besuchte Birte eine Kirmes.
			2	Sie stand mit ihrem Freund in der Schlange vor der Achterbahn.
			3	Nun müsste sie diesem Schisser Mut zureden.
Item	19	B	1	Im Sommer besuchte Ben eine Kirmes.
			2	Er stand mit seiner Freundin in der Schlange vor der Achterbahn.
			3	Nun müsste sie diesem Schisser Mut zureden.
Item	19	C	1	Im Sommer besuchte Ben eine Kirmes.
			2	Er stand mit seiner Freundin in der Schlange vor der Achterbahn.
			3	Sie musste ihm Mut zureden.
Item	19	D	1	Im Sommer besuchte Ben eine Kirmes.
			2	Er stand mit seiner Freundin in der Schlange vor der Achterbahn.
			3	Er musste ihr Mut zureden.
Item	20	A	1	In den Sommerferien war Kira auf dem Weg in den Urlaub.
			2	Auf der langen Autofahrt geriet sie immer wieder mit ihrem Bruder aneinander.
			3	Jetzt gleich würde sie diesen Zappelphilipp bei den Eltern verpetzen.
Item	20	B	1	In den Sommerferien war Lukas auf dem Weg in den Urlaub.
			2	Auf der langen Autofahrt geriet er immer wieder mit seiner Schwester aneinander.
			3	Jetzt gleich würde sie diesen Zappelphilipp bei den Eltern verpetzen.
Item	20	C	1	In den Sommerferien war Lukas auf dem Weg in den Urlaub.
			2	Auf der langen Autofahrt geriet er immer wieder mit seiner Schwester aneinander.
			3	Sie verpetzte ihn bei den Eltern.
Item	20	D	1	In den Sommerferien war Lukas auf dem Weg in den Urlaub.
			2	Auf der langen Autofahrt geriet er immer wieder mit seiner Schwester aneinander.
			3	Er verpetzte sie bei den Eltern.
Item	21	A	1	Am Montag Morgen lief Jaqueline gehetzt zum Klassenraum.
			2	Auf dem Flur stieß sie heftig mit ihrem Mitschüler zusammen.
			3	Jetzt müsste sie mit diesem Tollpatsch zur Schulkrankenschwester.
Item	21	B	1	Am Montag Morgen lief Arne gehetzt zum Klassenraum.
			2	Auf dem Flur stieß er heftig mit seiner Mitschülerin zusammen.
			3	Jetzt müsste sie mit diesem Tollpatsch zur Schulkrankenschwester.
Item	21	C	1	Am Montag Morgen lief Arne gehetzt zum Klassenraum.
			2	Auf dem Flur stieß er heftig mit seiner Mitschülerin zusammen.
			3	Sie ging mit ihm zur Schulkrankenschwester.
Item	21	D	1	Am Montag Morgen lief Arne gehetzt zum Klassenraum.
			2	Auf dem Flur stieß er heftig mit seiner Mitschülerin zusammen.
			3	Er ging mit ihr zur Schulkrankenschwester.

Item/Filler	Item/Filler Number	Condition	Sentence Number	Sentence
Item	22	A	1	Letzte Woche fuhr Maren mit ihrem Auto zum Supermarkt.
			2	Auf dem Weg stieß sie mit einem anderen Autofahrer zusammen.
			3	Gleich würde sie diesen Irren zur Rede stellen.
Item	22	B	1	Letzte Woche fuhr Tim mit dem Auto zum Supermarkt.
			2	Auf dem Weg stieß er mit einer anderen Autofahrerin zusammen.
			3	Gleich würde sie diesen Irren zur Rede stellen.
Item	22	C	1	Letzte Woche fuhr Tim mit dem Auto zum Supermarkt.
			2	Auf dem Weg stieß er mit einer anderen Autofahrerin zusammen.
			3	Sie wollte ihn daraufhin zur Rede stellen.
Item	22	D	1	Letzte Woche fuhr Tim mit dem Auto zum Supermarkt.
			2	Auf dem Weg stieß er mit einer anderen Autofahrerin zusammen.
			3	Er wollte sie daraufhin zur Rede stellen.
Item	23	A	1	Letzte Woche hatte Janina einen schlimmen Streit.
			2	Daraufhin schmiss sie ihren Freund aus der Wohnung raus.
			3	Jetzt würde sie diesen Idioten hoffentlich nie wieder sehen.
Item	23	B	1	Letzte Woche hatte Lars einen schlimmen Streit.
			2	Daraufhin schmiss er seine Freundin aus der Wohnung raus.
			3	Jetzt würde sie diesen Idioten hoffentlich nie wieder sehen.
Item	23	C	1	Letzte Woche hatte Lars einen schlimmen Streit.
			2	Daraufhin schmiss er seine Freundin aus der Wohnung raus.
			3	Sie hoffte, ihn nie wieder sehen zu müssen.
Item	23	D	1	Letzte Woche hatte Lars einen schlimmen Streit.
			2	Daraufhin schmiss er seine Freundin aus der Wohnung raus.
			3	Er hoffte, sie nie wieder sehen zu müssen.
Item	24	A	1	Letzten Sonntag suchte Annika nach Raritäten auf dem Flohmarkt.
			2	An einem Stand verhandelte sie sehr lange mit einem Herren über einen fairen Preis für eine alte Teekanne.
			3	Jetzt würde sie diesem Geizhals aber keine weiteren Zugeständnisse mehr machen.
Item	24	B	1	Letzten Sonntag suchte Anton nach Raritäten auf dem Flohmarkt.
			2	An einem Stand verhandelte er sehr lange mit einer Dame über einen fairen Preis für eine alte Teekanne.
			3	Jetzt würde sie diesem Geizhals aber keine weiteren Zugeständnisse mehr machen.
Item	24	C	1	Letzten Sonntag suchte Anton nach Raritäten auf dem Flohmarkt.
			2	An einem Stand verhandelte er sehr lange mit einer Dame über einen fairen Preis für eine alte Teekanne.
			3	Sie wollte ihm keine weiteren Zugeständnisse machen.
Item	24	D	1	Letzten Sonntag suchte Anton nach Raritäten auf dem Flohmarkt.
			2	An einem Stand verhandelte er sehr lange mit einer Dame über einen fairen Preis für eine alte Teekanne.
			3	Er wollte ihr keine weiteren Zugeständnisse machen.

Item/Filler	Item/Filler Number	Condition	Sentence Number	Sentence
Filler	1	ambiguous/funny	1	Während im Wohnzimmer der Fernseher lief, kochte Frank in der Küche einen leckeren Auflauf.
			2	Er hatte einen guten Freund zum Abendessen eingeladen.
			3	Er war mit Käse überbacken.
Filler	2	ambiguous/funny	1	Als Karl seinen Geburtstag feierte, besuchte ihn seine ganze Familie.
			2	Er setzte sich mit seinen Gästen an den Tisch und schnitt den Kuchen an.
			3	Er war sogar mit Marzipan ummantelt und mit Krokant-Streuseln verziert.
Filler	3	ambiguous/funny	1	Als das erste Gehalt auf das Konto überwiesen wurde, wollte Tobias das Geld in sein Hobby investieren.
			2	Er kaufte sich endlich den langersehnten Comic.
			3	Er war eine Sammleredition.
Filler	4	ambiguous/funny	1	Als die Wirtschaftskrise überwunden war, bekam Andrea eine Gehaltserhöhung.
			2	Endlich konnte sie in eine größere Wohnung ziehen.
			3	Sie lag mitten in der Stadt.
Filler	5	ambiguous/funny	1	Als Gerhard endlich die Fahrprüfung bestanden hatte, war er nicht mehr aufzuhalten.
			2	Er wollte sofort mit dem Oldtimer seiner Eltern eine Runde drehen.
			3	Er stand den ganzen Sommer in der Garage.
Filler	6	ambiguous/funny	1	Als alle Abiturientinnen in einer feierlichen Zeremonie verabschiedet wurden, gebührte Lina eine besondere Ehre.
			2	Als beste Schülerin ihres Jahrgangs widmete die Direktorin ihr eine Ehrentafel.
			3	Sie stand jetzt ein Jahr in der Vitrine der Schule
Filler	7	ambiguous/funny	1	Sonntags war Peter immer ganz erschöpft von der Woche.
			2	Er verbrachte am liebsten den ganzen Tag in seinem Sessel.
			3	Weil er so alt war, drängte seine Frau darauf, einen neuen zu kaufen.
Filler	8	ambiguous/funny	1	Am Tag der offenen Tür zeigte Christoph seinem Sohn stolz seinen Arbeitsplatz.
			2	Sein Sohn interessierte sich besonders für den Roboter, der die Autos zusammenbaut.
			3	Er konnte sehr präzise schweißen.
Filler	9	ambiguous/funny	1	Als nur noch zwei Spieler am Tisch saßen, war Oliver klar, dass er kurz davor war zu gewinnen.
			2	Als nächstes deckte der Dealer einen König auf.
			3	Er lag da nun direkt neben der Dame und dem Ass.
Filler	10	ambiguous/funny	1	In der Woche vor Silvester war Florian total gestresst im Supermarkt.
			2	Er ging versehentlich zum langsamsten Kassierer.
			3	Er ließ sich alle Zeit der Welt.

Item/Filler	Item/Filler Number	Condition	Sentence Number	Sentence
Filler	11	rather unnatural	1	Nachdem Sabine ihren Schulabschluss erhalten hatte, wollte sie eine Ausbildung beginnen.
			2	Sie bewarb sich dafür bei einer anerkannten Firma.
			3	Sie wurde kurz nach dem Krieg gegründet.
Filler	12	rather unnatural	1	In den Osterferien fuhr Lisa immer nach Italien.
			2	Mit ihrer Mutter teilte sie ihre Liebe für italienische Pizza.
			3	Sie war aus besonders dünnem Teig.
Filler	13	rather unnatural	1	An seinem Namenstag feiert die ganze Familie mit Julius ein Fest.
			2	Seinen Geburtstag kann er nämlich nur alle 4 Jahre feiern.
			3	Er ist nämlich am 29. Februar.
Filler	14	rather unnatural	1	Als Elena ihre Erstkommunion feierte, wurde sie reich beschenkt.
			2	Von ihrer Oma bekam sie eine Schultasche.
			3	Sie ist aus feinstem Leder.
Filler	15	rather unnatural	1	Nachdem Micha heimlich einen Cocktailkurs besucht hatte, wollte er seine Freundin mit leckeren Drinks überraschen.
			2	Stolz mixte er ihr seine Eigenkreation.
			3	Er bestand jedoch hauptsächlich aus Zucker.
Filler	16	rather unnatural	1	Am Tag des großen Wettkampfes war Niklas furchtbar aufgeregt.
			2	Sein Trainer hatte ihm einen Glücksstein zugesteckt.
			3	Er hüpfte in seiner Tasche umher.
Filler	17	rather unnatural	1	Als Sara zum ersten Mal Gehalt bekam, lud sie ihre Familie zum Essen ein.
			2	In einer kurzen Ansprache bedankte sie sich besonders bei ihrer Mutter für die Unterstützung in den letzten Jahren.
			3	Sie dauerte nur ein paar Minuten, aber sie rührte alle zu Tränen.
Filler	18	rather unnatural	1	An ihrem Geburtstag schmiss Leonie eine riesige Party.
			2	Sie bat ihre Schwester, Alkohol zu besorgen.
			3	Diese war nämlich noch minderjährig.
Filler	19	rather unnatural	1	Als Kind bekam Erik einen Welpen von seinem Großvater.
			2	Er liebte seinen zotteligen Freund sehr.
			3	Er war ein Labrador.
Filler	20	rather unnatural	1	Als Richard in seine erste eigene Wohnung zog, lud er seine Familie zum Abendessen ein.
			2	Sein Onkel brachte ihm einen besonders guten Rotwein mit.
			3	Er war aus seiner eigenen Kellerei.

Item/Filler	Item/Filler Number	Condition	Sentence Number	Sentence
Filler	21	rather unnatural	1	Als Tarzan Jane zum ersten Mal sah, verliebte er sich sofort.
			2	Er nahm sie mit zu einer wunderschönen Lichtung.
			3	Sie lag mitten im Dschungel und war sonnig.
Filler	22	rather unnatural	1	Als Cordula ein Musikinstrument lernen wollte, kaufte ihre Mutter ihr eine Blockflöte.
			2	Sie übte jeden Tag.
			3	Sie war aus Holz.
Filler	23	rather unnatural	1	In den Sommerferien fuhr Holger mit seiner Familie nach Spanien.
			2	Im Stau war er von seinem quengelnden Sohn genervt.
			3	Er war 10 Kilometer lang.
Filler	24	rather unnatural	1	Nach Feierabend musste Tim in die Werkstatt.
			2	Er hatte ein Problem mit seinem Wagen.
			3	Er war total verrostet.
Filler	25	rather unnatural	1	Letztes Jahr brauchte Elisa eine Veränderung.
			2	Sie ließ sich eine neue Frisur schneiden.
			3	Sie war kurz und fransig.
Filler	26	rather unnatural	1	In seiner Jugend hatte Udo häufig Unfug im Sinn.
			2	Einmal stahl er einen Gameboy.
			3	Er war ein neues Modell.
Filler	27	rather unnatural	1	Nach dem Essen ging Freddy mit seiner neuen Freundin noch die Promenade entlang.
			2	Er kaufte ihr einen Eisbecher.
			3	Er war unheimlich lecker.
Filler	28	rather unnatural	1	Als der Frühling anbrach, ging Maja häufig joggen.
			2	Sie kaufte sich extra eine neue Fitnessuhr.
			3	Sie war digital.
Filler	29	rather unnatural	1	Als das letzte Semester begann, war Melina hochmotiviert.
			2	Sie fieberte der Abschlussklausur entgegen.
			3	Sie war dreistündig.
Filler	30	rather unnatural	1	Als Tristan seinen ersten großen Auftritt hatte, wollte er besonders schick aussehen.
			2	Er kaufte sich einen neuen Anzug.
			3	Er war aus Seide.

Item/Filler	Item/Filler Number	Condition	Sentence Number	Sentence
Filler	31	rather natural	1	Als die ersten Elektroautos auf den Markt kamen, begab sich Josefine zum nächsten Autohaus.
			2	Sie sprach dort mit einem der zuständigen Händler.
			3	Der konnte sie wirklich gut beraten.
Filler	32	rather natural	1	Als alle Klausuren geschrieben waren, konnte Leon endlich aufatmen.
			2	Er hatte die letzten Wochen zusammen mit einer Kommilitonin sehr viel gelernt.
			3	Die war ihm eine große Hilfe.
Filler	33	rather natural	1	In ihrem ersten Jahr bei der Feuerwehr wurde Lea auf die Probe gestellt.
			2	Besonders ihre männlichen Kollegen machten ihr das Leben schwer.
			3	Aber sie hielt sich wacker und schloss die Ausbildung mit Bestnote ab.
Filler	34	rather natural	1	In der Oberstufe war Jenny eine sehr gute Schülerin.
			2	Dabei zog sie jedes Wochenende mit ihrer besten Freundin um die Häuser.
			3	Dass sie so eine Streberin war, hätte man ihr dabei kaum zugetraut.
Filler	35	rather natural	1	Als der Elternsprechtag näher rückte, war Ingo verzweifelt.
			2	Er wusste, dass seine Lehrerin seinem Vater von der schlechten Note erzählen würde.
			3	Vielleicht sollte er es ihm doch lieber selbst sagen.
Filler	36	rather natural	1	Wenn der FC spielt, ist Max immer im Stadion.
			2	Seine Tante hat ihm eine Dauerkarte geschenkt.
			3	Dafür ist er ihr unendlich dankbar.
Filler	37	rather natural	1	Als Achim eine Diät machte, verzichtete er weitestgehend auf Kohlenhydrate.
			2	Seine Freundin machte ihm jeden Abend einen großen Salat.
			3	Sie freute sich, dass sie ihren Liebsten beim Abnehmen unterstützen konnte.
Filler	38	rather natural	1	Als die Wohnung endlich fertig eingerichtet war, wollte Linda sich entspannt zurücklehnen.
			2	Sie ließ sich in ihren Lieblingssessel fallen.
			3	Sie war wirklich glücklich über ihre neue Einrichtung.
Filler	39	rather natural	1	Als Michelle von einem Nena-Konzert in ihrer Nähe erfuhr, holte sie ohne zu zögern zwei Karten.
			2	Sie wollte ihren Bruder mitnehmen.
			3	Sie würde mit ihm alle beliebten Lieder mitsingen.
Filler	40	rather natural	1	Am Wochenende war Luis in der Stadt unterwegs.
			2	Weil der Heimweg so weit war, übernachtete er bei seiner Tante.
			3	Die freute sich immer, wenn ihr Neffe zu Besuch kam.

10 References

Achim, A. M., Achim, A., & Fossard, M. (2017). Knowledge likely held by others affects speakers' choices of referential expressions at different stages of discourse. *Language, Cognition and Neuroscience*, *32*(1), 21–36. https://doi.org/10.1080/23273798.2016.1234059

Adil, S., Lacoste-Badie, S., & Droulers, O. (2018). Face presence and gaze direction in print advertisements: How they influence consumer responses – An eye-tracking study. *Journal of Advertising Research*, *58*(4), 443–455. https://doi.org/10.2501/JAR-2018-004

Al Moubayed, S., & Beskow, J. (2009). Effects of visual prominence cues on speech intelligibility. In *Proceedings of Auditory-Visual Speech Processing* (pp. 43–46). International Speech Communication Association.

Alexanderson, S., House, D., & Beskow, J. (2013). Aspects of co-occurring syllables and head nods in spontaneous dialogue. In *Proceedings of Auditory-Visual Speech Processing* (pp. 169–172). International Speech Communication Association.

Al Moubayed, S., Beskow, J., & Granström, B. (2009). Auditory visual prominence. *Journal on Multimodal User Interfaces*, *3*(4), 299–309. https://doi.org/10.1007/s12193-010-0054-0

Ambrazaitis, G., Frid, J., & House, D. (2020). Word prominence ratings in Swedish television news readings – effects of pitch accents and head movements. *10th International Conference on Speech Prosody*, 314–318. https://doi.org/10.21437/SpeechProsody.2020-64

Ambrazaitis, G., & House, D. (2017). Multimodal prominences: Exploring the patterning and usage of focal pitch accents, head beats and eyebrow beats in Swedish television news readings. *Speech Communication*, *95*, 100–113. https://doi.org/10.1016/j.specom.2017.08.008

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders (DSM-5; Fifth Edition).* American Psychiatric Publishing.

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, *116*(4), 953–970. https://doi.org/10.1037/a0016923

Ariel, M. (1990). *Accessing Noun-Phrase Antecedents*. Routledge.

Arnold, D., Wagner, P., & Baayen, H. (2013). Using generalized additive models and random forests to model prosodic prominence in German. *Interspeech 2013*, 272–276.

Arnold, J. E. (2010). How speakers refer: The role of accessibility. *Language and Linguistics Compass*, *4*(4), 187–203. https://doi.org/10.1111/j.1749-818X.2010.00193.x

Arnold, J. E., Bennetto, L., & Diehl, J. J. (2009). Reference production in young speakers with and without autism: Effects of discourse status and processing constraints. *Cognition*, *110*(2), 131–146. https://doi.org/10.1016/j.cognition.2008.10.016

Aster, M., Neubauer, A., & Horn, R. (2006). *Hamburg-Wechsler-Intelligenz-Test für Erwachsene III.* Huber.

Auyeung, B., Lombardo, M. V., Heinrichs, M., Chakrabarti, B., Sule, A., Deakin, J. B., Bethlehem, R. a. I., Dickens, L., Mooney, N., Sipple, J. a. N., Thiemann, P., & Baron-Cohen, S. (2015). Oxytocin increases eye contact during a real-time, naturalistic social interaction in males with and without autism. *Translational Psychiatry*, *5*(2), Article 2. https://doi.org/10.1038/tp.2014.146

Banfield, A. (1982). *Unspeakable Sentences: Narration and representation in the language of fiction.* Routledge.

Baron-Cohen, S. (1989). The autistic child's theory of mind: A case of specific developmental delay. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *30*(2), 285–297. https://doi.org/10.1111/j.1469-7610.1989.tb00241.x

Baron-Cohen, S., Campbell, R., Karmiloff-Smith, A., Grant, J., & Walker, J. (1995). Are children with autism blind to the mentalistic significance of the eyes? *British Journal of Developmental Psychology*, *13*(4), 379–398. https://doi.org/10.1111/j.2044-835X.1995.tb00687.x

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, *21*(1), 37–46.

Baron-Cohen, S., O'Riordan, M., Stone, V., Jones, R., & Plaisted, K. (1999). Recognition of faux pas by normally developing children and children with Asperger syndrome or high-functioning autism. *Journal of Autism and Developmental Disorders*, *29*(5), 407–418. https://doi.org/10.1023/a:1023035012436

Baron-Cohen, S., Richler, J., Bisarya, D., Gurunathan, N., & Wheelwright, S. (2003). The systemizing quotient: An investigation of adults with Asperger syndrome or high-functioning autism, and normal sex differences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *358*(1430), 361–374. https://doi.org/10.1098/rstb.2002.1206

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: An investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, *34*(2), 163–175.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *42*(2), 241–251.

Baron-Cohen, S., Wheelwright, S., & Jolliffe, T. (1997). Is there a "language of the eyes"? Evidence from normal adults, and adults with autism or Asperger Syndrome. *Visual Cognition*, *4*(3), 311–331. https://doi.org/10.1080/713756761

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Males and Females, Scientists and Mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5–17. https://doi.org/10.1023/A:1005653411471

Baumann, S., & Winter, B. (2018). What makes a word prominent? Predicting untrained German listeners' perceptual judgments. *Journal of Phonetics*, *70*, 20–38. https://doi.org/10.1016/j.wocn.2018.05.004

Bayliss, A., Di Pellegrino, G., & Tipper, S. (2005). Sex differences in eye gaze and symbolic cueing of attention. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *58*, 631–650. https://doi.org/10.1080/02724980443000124

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory-II*. Psychological Corporation.

Begeer, S., De Rosnay, M., Lunenburg, P., Stegge, H., & Terwogt, M. M. (2014). Understanding of emotions based on counterfactual reasoning in children with autism spectrum disorders. *Autism: The International Journal of Research and Practice*, *18*(3), 301–310. https://doi.org/10.1177/1362361312468798

Ben-David, B. M., Ben-Itzchak, E., Zukerman, G., Yahav, G., & Icht, M. (2020). The Perception of Emotions in Spoken Language in Undergraduates with High Functioning Autism Spectrum Disorder: A Preserved Social Skill. *Journal of Autism and Developmental Disorders*, *50*(3), 741–756. https://doi.org/10.1007/s10803-019-04297-2

Bennett, T. A., Szatmari, P., Bryson, S., Duku, E., Vaccarella, L., & Tuff, L. (2013). Theory of Mind, Language and Adaptive Functioning in ASD: A Neuroconstructivist Perspective. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, *22*(1), 13–19. PMCID: PMC3565710

Beveridge, M. E. L., & Pickering, M. J. (2013). Perspective taking in language: Integrating the spatial and action domains. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00577

Biau, E., & Soto-Faraco, S. (2013). Beat gestures modulate auditory integration in speech perception. *Brain and Language*, *124*(2), 143–152. https://doi.org/10.1016/j.bandl.2012.10.008

Bimpikou, S. (2023). *Inside characters' minds: The role of reports in narrative perspective taking* [Doctoral dissertation, University of Groningen]. https://doi.org/10.33612/diss.735608404

Bishop, J. (2016). Focus projection and prenuclear accents: Evidence from lexical processing. *Language, Cognition and Neuroscience*, *32*(2), 236–253. https://doi.org/10.1080/23273798.2016.1246745

Bishop, J., Kuo, G., & Kim, B. (2020). Phonology, phonetics, and signal-extrinsic factors in the perception of prosodic prominence: Evidence from Rapid Prosody Transcription. *Journal of Phonetics*, *82*, Article 100977. https://doi.org/10.1016/j.wocn.2020.100977

Bobak, A. K., & Langton, S. R. H. (2015). Working memory load disrupts gaze-cued orienting of attention. *Frontiers in Psychology*, *6*, Article 1258. https://doi.org/10.3389/fpsyg.2015.01258

Boucher, J., Cowell, P., Howard, M., Broks, P., Farrant, A., Roberts, N., & Mayes, A. (2005). A combined clinical, neuropsychological, and neuroanatomical study of adults with high functioning autism. *Cognitive Neuropsychiatry*, *10*(3), 165–213. https://doi.org/10.1080/13546800444000038

Bowler, D. M. (1992). "Theory of mind" in Asperger's syndrome. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *33*(5), 877–893. https://doi.org/10.1111/j.1469-7610.1992.tb01962.x

Bowler, D. M., Gardiner, J. M., & Grice, S. J. (2000). Episodic memory and remembering in adults with Asperger syndrome. *Journal of Autism and Developmental Disorders*, *30*(4), 295–304. https://doi.org/10.1023/A:1005575216176

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. In *Proceedings of the National Academy of Sciences*, *105*(38), (pp. 14325–14329). https://doi.org/10.1073/pnas.0803390105

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes*, *25*(7–9), 1044–1098. https://doi.org/10.1080/01690965.2010.504378

Brewer, N., Young, R. L., & Barnett, E. (2017). Measuring Theory of Mind in Adults with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *47*(7), 1927–1941. https://doi.org/10.1007/s10803-017-3080-x

Brickenkamp, R. (2002). *Test d2: Aufmerksamkeits-Belastungs-Test* (9th ed., revised and newly standardized). Hogrefe.

Brône, G., Oben, B., Jehoul, A., Vranjes, J., & Feyaerts, K. (2017). Eye gaze and viewpoint in multimodal interaction management. *Cognitive Linguistics*, *28*(3), 449–483. https://doi.org/10.1515/cog-2016-0119

Brosnan, M., & Ashwin, C. (2023). Thinking, fast and slow on the autism spectrum. *Autism*, *27*(5), 1245–1255. https://doi.org/10.1177/13623613221132437

Brosnan, M., Ashwin, C., & Lewton, M. (2017). Brief Report: Intuitive and Reflective Reasoning in Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *47*(8), 2595–2601. https://doi.org/10.1007/s10803-017-3131-3

Brosnan, M., Lewton, M., & Ashwin, C. (2016). Reasoning on the Autism Spectrum: A Dual Process Theory Account. *Journal of Autism and Developmental Disorders*, *46*(6), 2115–2125. https://doi.org/10.1007/s10803-016-2742-4

Brown-Schmidt, S. (2009). The role of executive function in perspective taking during online language comprehension. *Psychonomic Bulletin & Review*, *16*(5), 893–900. https://doi.org/10.3758/PBR.16.5.893

Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V., & Rizzolatti, G. (2005). Listening to action-related sentences modulates the activity of the motor system: A combined TMS and behavioral study. *Cognitive Brain Research*, *24*(3), 355–363. https://doi.org/10.1016/j.cogbrainres.2005.02.020

Burke, D. M., & Shafto, M. A. (2008). Language and Aging. In F. I. M. Craik & T. A. Salthouse (Eds.), *The Handbook of Aging and Cognition* (3rd ed., pp. 373–443). Psychology Press.

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C., & Vuorre, M. (2019). Ordinal Regression Models in Psychology: A Tutorial. *Advances in Methods and Practices in Psychological Science, 2*(1), 77–101. https://doi.org/10.1177/2515245918823199

Buswell, G. T. (1935). *How People Look at Pictures: A Study of the Psychology of Perception in Art.* University of Chicago Press.

Callenmark, B., Kjellin, L., Rönnqvist, L., & Bölte, S. (2014). Explicit versus implicit social cognition testing in autism spectrum disorder. *Autism*, *18*(6), 684–693. https://doi.org/10.1177/1362361313492393

Cañadas, E., & Lupiáñez, J. (2012). Spatial interference between gaze direction and gaze location: A study on the eye contact effect. *Quarterly Journal of Experimental Psychology (2006)*, *65*(8), 1586–1598. https://doi.org/10.1080/17470218.2012.659190

Cantalini, G., & Moneglia, M. (2020). The annotation of gesture and gesture / prosody synchronization in multimodal speech corpora. *Journal of Speech Sciences*, *9, 7–30*. https://doi.org/10.20396/joss.v9i00.14956

Capage, L., & Watson, A. C. (2001). Individual Differences in Theory of Mind, Aggressive Behavior, and Social Skills in Young Children. *Early Education and Development*, *12*(4), 613–628. https://doi.org/10.1207/s15566935eed1204_7

Carstensen, L., Turan, B., Scheibe, S., Ram, N., Hershfield, H., Samanez-Larkin, G., Brooks, K., & Nesselroade, J. (2010). Emotional experience improves with age: Evidence based on over 10 years of experience sampling. *Psychology and Aging*, *26*, 21–33. https://doi.org/10.1037/a0021285

Caruana, N., Stieglitz Ham, H., Brock, J., Woolgar, A., Kloth, N., Palermo, R., & McArthur, G. (2018). Joint attention difficulties in autistic adults: An interactive eye-tracking study. *Autism*, *22*(4), 502–512. https://doi.org/10.1177/1362361316676204

Castelhano, M., Wieth, M., & Henderson, J. (2007). I see what you see: Eye movements in real-world scenes are affected by perceived direction of gaze. In M. H. K. Poel, B. Taatgen, & M. van Rijn (Eds.), *Attention in cognitive systems: Theories and systems from an interdisciplinary viewpoint: 4th International Workshop on Attention in Cognitive Systems* (pp. 251–262). Springer. https://doi.org/10.1007/978-3-540-77343-6_16

Chafe, W. L. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. N. Li (Ed.), *Subject and topic* (pp. 27–55). Academic Press.

Chan, K. K. L., & To, C. K. S. (2016). Do Individuals with High-Functioning Autism Who Speak a Tone Language Show Intonation Deficits? *Journal of Autism and Developmental Disorders*, *46*(5), 1784–1792. https://doi.org/10.1007/s10803-016-2709-5

Chapple, M., Davis, P., Billington, J., & Corcoran, R. (2023). Exploring the different cognitive, emotional and imaginative experiences of autistic and non-autistic adult readers when contemplating serious literature as compared to non-fiction. *Frontiers in Psychology*, *14*, Article 1001268. https://doi.org/10.3389/fpsyg.2023.1001268

Chen, Z., McCrackin, S. D., Morgan, A., & Itier, R. J. (2021). The Gaze Cueing Effect and its Enhancement by Facial Expressions are Impacted by Task Demands: Direct Comparison of Target Localization and Discrimination Tasks. *Frontiers in Psychology*, *12*, Article 618606. https://doi.org/10.3389/fpsyg.2021.618606

Cheng, S. T. T., Lam, G. Y. H., & To, C. K. S. (2017). Pitch Perception in Tone Language-Speaking Adults with and without Autism Spectrum Disorders. *I-Perception*, *8*(3), Article 2041669517711200. https://doi.org/10.1177/2041669517711200

Chita-Tegmark, M. (2016). Social attention in ASD: A review and meta-analysis of eye-tracking studies. *Research in Developmental Disabilities*, *48*, 79–93. https://doi.org/10.1016/j.ridd.2015.10.011

Chuk, T., Chan, A. B., Shimojo, S., & Hsiao, J. H. (2016). Mind reading: Discovering individual preferences from eye movements using switching hidden Markov models. In *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 182–187).

Chung, Y. S., Barch, D., & Strube, M. (2014). A meta-analysis of mentalizing impairments in adults with schizophrenia and Autism Spectrum Disorder. *Schizophrenia Bulletin*, *40*(3), 602–616. https://doi.org/10.1093/schbul/sbt048

Cole, J., Mo, Y., & Hasegawa-Johnson, M. (2010). Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology*, *110*, 425–452. https://doi.org/10.1515/labphon.2010.022

Colle, L., Baron-Cohen, S., Wheelwright, S., & van der Lely, H. K. J. (2008). Narrative discourse in adults with high-functioning autism or Asperger syndrome. *Journal of Autism and Developmental Disorders*, *38*(1), 28–40. https://doi.org/10.1007/s10803-007-0357-5

Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., & Lepore, F. (2008). Audio-visual integration of emotion expression. *Brain Research*, *1242*, 126–135. https://doi.org/10.1016/j.brainres.2008.04.023

Conson, M., Mazzarella, E., Esposito, D., Grossi, D., Marino, N., Massagli, A., & Frolli, A. (2015). "Put myself into your place": Embodied simulation and perspective taking in autism spectrum disorders. *Autism Research: Official Journal of the International Society for Autism Research*, *8*(4), 454–466. https://doi.org/10.1002/aur.1460

Criss, A., & Shiffrin, R. (2004). Interactions between study task, study time, and the low-frequency hit rate advantage in recognition memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *30*, 778–786. https://doi.org/10.1037/0278-7393.30.4.778

Crocker, M., & Keller, F. (2005). Probabilistic Grammars as Models of Gradience in Language Processing. In G. Fanselow, C. Féry, M. Schlesewsky, & R. Vogel (Eds.), *Gradience in Grammar: Generative Perspectives* (pp. 227–245). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199274796.003.0012

Cvejic, E., Kim, J., Davis, C., & Gibert, G. (2010). Prosody for the eyes: Quantifying visual prosody using guided principal component analysis. In *Proceedings of the 11th Annual Conference of the International Speech Communication Association* (pp. 1433–1436). https://doi.org/10.21437/Interspeech.2010-434

Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*, *47*(2), 292–314. https://doi.org/10.1016/S0749-596X(02)00001-3

Dalton, K. M., Nacewicz, B. M., Johnstone, T., Schaefer, H. S., Gernsbacher, M. A., Goldsmith, H. H., Alexander, A. L., & Davidson, R. J. (2005). Gaze fixation and the neural circuitry of face processing in autism. *Nature Neuroscience*, *8*(4), 519–526. https://doi.org/10.1038/nn1421

David, N., Aumann, C., Bewernick, B. H., Santos, N. S., Lehnhardt, F.-G., & Vogeley, K. (2010). Investigation of mentalizing and visuospatial perspective taking for self and other in Asperger syndrome. *Journal of Autism and Developmental Disorders*, *40*(3), 290–299. https://doi.org/10.1007/s10803-009-0867-4

de Gelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition and Emotion*, *14*(3), 289–311. https://doi.org/10.1080/026999300378824

DeAngelus, M. A., & Pelz, J. (2009). Top-down control of eye movements: Yarbus revisited. *Visual Cognition*, *6–7*(17), 790–811. https://doi.org/10.1080/13506280902793843

Diehl, J. J., Watson, D., Bennetto, L., Mcdonough, J., & Gunlogson, C. (2009). An acoustic analysis of prosody in high-functioning autism. *Applied Psycholinguistics*, *30*(3), 385–404. https://doi.org/10.1017/S0142716409090201

Dodd, M. D., Weiss, N., McDonnell, G. P., Sarwal, A., & Kingstone, A. (2012). Gaze cues influence memory…but not for long. *Acta Psychologica*, *141*(2), 270–275. https://doi.org/10.1016/j.actpsy.2012.06.003

Dohen, M., & Lœvenbruck, H. (2009). Interaction of Audition and Vision for the Perception of Prosodic Contrastive Focus. *Language and Speech*, *52*(2–3), 177–206. https://doi.org/10.1177/0023830909103166

Dohen, M., Loevenbruck, H., & Hill, H. (2006, May 2). Visual Correlates of Prosodic Contrastive Focus in French: Description and Inter-Speaker Variability. In *Proceedings of Speech Prosody* (Article 118). https://doi.org/10.21437/SpeechProsody.2006-210

Driver, J., Davis, G., Kidd, P., Maxwell, E., Ricciardelli, P., & Baron-Cohen, S. (1999). Gaze Perception Triggers Reflexive Visuospatial Orienting. *Visual Cognition*, *6*(5), 509–540. https://doi.org/10.1080/135062899394920

Droll, J. A., & Eckstein, M. P. (2009). Gaze control and memory for objects while walking in a real world environment. *Visual Cognition*, *17*(6–7), 1159–1184. https://doi.org/10.1080/13506280902797125

Dziobek, I., Fleck, S., Kalbe, E., Rogers, K., Hassenstab, J., Brand, M., Kessler, J., Woike, J. K., Wolf, O. T., & Convit, A. (2006). Introducing MASC: A movie for the assessment of social cognition. *Journal of Autism and Developmental Disorders*, *36*(5), 623–636. https://doi.org/10.1007/s10803-006-0107-0

Einav, S., & Hood, B. M. (2006). Children's use of the temporal dimension of gaze for inferring preference. *Developmental Psychology*, *42*(1), 142–152. https://doi.org/10.1037/0012-1649.42.1.142

Esteve-Gibert, N., Borràs-Comes, J., Asor, E., Swerts, M., & Prieto, P. (2017). The timing of head movements: The role of prosodic heads and edges. *The Journal of the Acoustical Society of America*, *141*(6), 4727–4739. https://doi.org/10.1121/1.4986649

Esteve-Gibert, N., & Prieto, P. (2013). Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *Journal of Speech, Language, and Hearing Research*, *56*(3), 850–864. https://doi.org/10.1044/1092-4388(2012/12-0049)

Fedor, J., Lynn, A., Foran, W., DiCicco-Bloom, J., Luna, B., & O'Hearn, K. (2018). Patterns of fixation during face recognition: Differences in autism across age. *Autism*, *22*(7), 866–880. https://doi.org/10.1177/1362361317714989

Ferreira, F., Apel, J., & Henderson, J. M. (2008). Taking a new look at looking at nothing. *Trends in Cognitive Sciences*, *12*(11), 405–410. https://doi.org/10.1016/j.tics.2008.07.007

Féry, C., & Kügler, F. (2008). Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics*, *36*(4), 680–703. https://doi.org/10.1016/j.wocn.2008.05.001

Flavell, J. H. (1977). The development of knowledge about visual perception. *Nebraska Symposium on Motivation, 25*, 43–76.

Fletcher-Watson, S., Leekam, S. R., Benson, V., Frank, M. C., & Findlay, J. M. (2009). Eye-movements reveal attention to social information in autism spectrum disorder. *Neuropsychologia*, *47*(1), 248–257. https://doi.org/10.1016/j.neuropsychologia.2008.07.016

Föcker, J., Gondan, M., & Röder, B. (2011). Preattentive processing of audio-visual emotional signals. *Acta Psychologica*, *137*(1), 36–47. https://doi.org/10.1016/j.actpsy.2011.02.004

Fonagy, P., Gergely, G., Jurist, E., & Target, M. (2004). *Affektregulierung, Mentalisierung und die Entwicklung des Selbst.* Klett-Cotta.

Forgeot d'Arc, B., Delorme, R., Zalla, T., Lefebvre, A., Amsellem, F., Moukawane, S., Letellier, L., Leboyer, M., Mouren, M.-C., & Ramus, F. (2017). Gaze direction detection in autism spectrum disorder. *Autism*, *21*(1), 100–107. https://doi.org/10.1177/1362361316630880

Fougnie, D., Cockhren, J., & Marois, R. (2018). A common source of attention for auditory and visual tracking. *Attention, Perception, & Psychophysics*, *80*(6), 1571–1583. https://doi.org/10.3758/s13414-018-1524-9

Fraundorf, S. H., Watson, D. G., & Benjamin, A. S. (2010). Recognition memory reveals just how CONTRASTIVE contrastive accenting really is. *Journal of Memory and Language*, *63*(3), 367–386. https://doi.org/10.1016/j.jml.2010.06.004

Fraundorf, S. H., Watson, D. G., & Benjamin, A. S. (2012). The effects of age on the strategic use of pitch accents in memory for discourse: A processing-resource account. *Psychology and Aging*, *27*(1), 88–98. https://doi.org/10.1037/a0024138

Freeth, M., Chapman, P., Ropar, D., & Mitchell, P. (2010). Do gaze cues in complex scenes capture and direct the attention of high functioning adolescents with ASD? Evidence from eye-tracking. *Journal of Autism and Developmental Disorders*, *40*(5), 534–547. https://doi.org/10.1007/s10803-009-0893-2

Freire, A., Eskritt, M., & Lee, K. (2004). Are eyes windows to a deceiver's soul? Children's use of another's eye gaze cues in a deceptive situation. *Developmental Psychology*, *40*(6), 1093–1104. https://doi.org/10.1037/0012-1649.40.6.1093

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review*, *5*(3), 490–495. https://doi.org/10.3758/BF03208827

Friesen, C. K., Ristic, J., & Kingstone, A. (2004). Attentional Effects of Counterpredictive Gaze and Arrow Cues. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(2), 319–329. https://doi.org/10.1037/0096-1523.30.2.319

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze Cueing of Attention. *Psychological Bulletin*, *133*(4), 694–724. https://doi.org/10.1037/0033-2909.133.4.694

Frischen, A., & Tipper, S. P. (2006). Long-term gaze cueing effects: Evidence for retrieval of prior states of attention from memory. *Visual Cognition*, *14*(3), 351–364. https://doi.org/10.1080/13506280544000192

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, *50*(4), 531–534. https://doi.org/10.1016/j.neuron.2006.05.001

Frith, U., & de Vignemont, F. (2005). Egocentrism, allocentrism, and Asperger syndrome. *Consciousness and Cognition*, *14*(4), 719–738. https://doi.org/10.1016/j.concog.2005.04.006

Frith, U., Morton, J., & Leslie, A. M. (1991). The cognitive basis of a biological disorder: Autism. *Trends in Neurosciences*, *14*(10), 433–438. https://doi.org/10.1016/0166-2236(91)90041-R

Gernsbacher, M. A., & Yergeau, M. (2019). Empirical Failures of the Claim That Autistic People Lack a Theory of Mind. *Archives of Scientific Psychology*, *7*(1), 102–118. https://doi.org/10.1037/arc0000067

Gianelli, C., Farnè, A., Salemme, R., Jeannerod, M., & Roy, A. C. (2011). The agent is right: When motor embodied cognition is space-dependent. *PloS One*, *6*(9), Article e25036. https://doi.org/10.1371/journal.pone.0025036

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition*, *13*(1), 8–20. https://doi.org/10.3758/BF03198438

Globerson, E., Amir, N., Kishon-Rabin, L., & Golan, O. (2015). Prosody recognition in adults with high-functioning autism spectrum disorders: From psychoacoustics to cognition. *Autism Research: Official Journal of the International Society for Autism Research*, *8*(2), 153–163. https://doi.org/10.1002/aur.1432

Golan, O., Baron-Cohen, S., Hill, J. J., & Rutherford, M. D. (2007). The "Reading the Mind in the Voice" test-revised: A study of complex emotion recognition in adults with and without autism spectrum conditions. *Journal of Autism and Developmental Disorders*, *37*(6), 1096–1106. https://doi.org/10.1007/s10803-006-0252-5

Gregory, S. E. A., & Jackson, M. C. (2017). Joint attention enhances visual working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(2), 237–249. https://doi.org/10.1037/xlm0000294

Gregory, S. E. A., & Kessler, K. (2022). Investigating age differences in the influence of joint attention on working memory. *Psychology and Aging*, *37*(6), 731–741. https://doi.org/10.1037/pag0000694

Grice, M., & Baumann, S. (2007). An introduction to intonation – functions and models. In J. Trouvain & U. Gut (Eds.), *Non-Native Prosody: Phonetic Description and Teaching Practice* (pp. 25–52). De Gruyter Mouton. https://doi.org/10.1515/9783110198751.1.25

Grice, M., Krüger, M., & Vogeley, K. (2016). Adults with Asperger syndrome are less sensitive to intonation than control persons when listening to speech. *Culture and Brain*, *4*(1), 38–50. https://doi.org/10.1007/s40167-016-0035-6

Grice, M., Wehrle, S., Krüger, M., Spaniol, M., Cangemi, F., & Vogeley, K. (2023). Linguistic prosody in autism spectrum disorder – An overview. *Language and Linguistics Compass*, *17*(5), Article e12498. https://doi.org/10.1111/lnc3.12498

Gronau, Q. F., Singmann, H., & Wagenmakers, E. (2020). Bridgesampling: An R Package for Estimating Normalizing Constants. *Journal of Statistical Software*, *92*(10), 1–29. https://doi.org/10.18637/jss.v092.i10

Grossman, R. B., Zane, E., Mertens, J., & Mitchell, T. (2019). Facetime vs. Screentime: Gaze Patterns to Live and Video Social Stimuli in Adolescents with ASD. *Scientific Reports*, *9*(1), Article 1. https://doi.org/10.1038/s41598-019-49039-7

Guillon, Q., Hadjikhani, N., Baduel, S., & Rogé, B. (2014). Visual social attention in autism spectrum disorder: Insights from eye tracking studies. *Neuroscience and Biobehavioral Reviews*, *42*, 279–297. https://doi.org/10.1016/j.neubiorev.2014.03.013

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive Status and the Form of Referring Expressions in Discourse. *Language*, *69*(2), 274–307. https://doi.org/10.2307/416535

Habib, A., Harris, L., Pollick, F., & Melville, C. (2019). A meta-analysis of working memory in individuals with autism spectrum disorders. *PLoS ONE*, *14*(4), Article e0216198. https://doi.org/10.1371/journal.pone.0216198

Halliday, M. A. K. (1967). Notes on Transitivity and Theme in English: Part 2. *Journal of Linguistics*, *3*(2), 199–244. https://doi.org/10.1017/S0022226700016613

Hammerschmidt, K., & Jürgens, U. (2007). Acoustical Correlates of Affective Prosody. *Journal of Voice*, *21*(5), 531–540. https://doi.org/10.1016/j.jvoice.2006.03.002

Happé, F. G. E. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders*, *24*(2), 129–154. https://doi.org/10.1007/BF02172093

Hartung, F., Hagoort, P., & Willems, R. M. (2017). Readers select a comprehension mode independent of pronoun: Evidence from fMRI during narrative comprehension. *Brain and Language*, *170*, 29–38. https://doi.org/10.1016/j.bandl.2017.03.007

Heavey, L., Phillips, W., Baron-Cohen, S., & Rutter, M. (2000). The Awkward Moments Test: A Naturalistic Measure of Social Understanding in Autism. *Journal of Autism and Developmental Disorders*, *30*(3), 225–236. https://doi.org/10.1023/A:1005544518785

Henderson, J. M., Williams, C. C., & Falk, R. J. (2005). Eye movements are functional during face learning. *Memory & Cognition*, *33*(1), 98–106. https://doi.org/10.3758/BF03195300

Hendriks, P., Koster, C., & Hoeks, J. C. J. (2014). Referential choice across the lifespan: Why children and elderly adults produce ambiguous pronouns. *Language and Cognitive Processes*, *29*(4), 391–407. https://doi.org/10.1080/01690965.2013.766356

Henry, J. D., Phillips, L. H., Ruffman, T., & Bailey, P. E. (2013). A meta-analytic review of age differences in theory of mind. *Psychology and Aging*, *28*(3), 826–839. https://doi.org/10.1037/a0030677

Hernandez, N., Metzger, A., Magné, R., Bonnet-Brilhault, F., Roux, S., Barthelemy, C., & Martineau, J. (2009). Exploration of core features of a human face by healthy and autistic adults analyzed by visual scanning. *Neuropsychologia*, *47*(4), 1004–1012. https://doi.org/10.1016/j.neuropsychologia.2008.10.023

Himmelmann, N. P., & Primus, B. (2015). Prominence Beyond Prosody – A First Approximation. In A. De Dominicis (Ed.), *Prominences in Linguistics. Proceedings of the pS-prominenceS International Conference* (pp. 38–58). Disucom Press. URN: urn:nbn:de:hbz:38-249356

Hinterwimmer, S. (2019). Prominent protagonists. *Journal of Pragmatics*, *154*, 79–91. https://doi.org/10.1016/j.pragma.2017.12.003

Hinterwimmer, S., & Meuser, S. (2019). Erlebte Rede und Protagonistenprominenz. In S. Engelberg, C. Fortmann, & I. Rapp (Eds.), *Rede- und Gedankenwiedergabe in narrativen Strukturen – Ambiguitäten und Varianz* (pp. 177–200). Buske.

Hirshman, E., & Arndt, J. (1997). Discriminating alternative conceptions of false recognition: The cases of word concreteness and word frequency. *Learning, Memory, and Cognition*, *6*(23), 1306–1323. https://doi.org/10.1037/0278-7393.23.6.1306

Hobson, R. P., Ouston, J., & Lee, A. (1988). What's in a face? The case of autism. *British Journal of Psychology*, *79*(4), 441–453. https://doi.org/10.1111/j.2044-8295.1988.tb02745.x

House, D., Beskow, J., & Granström, B. (2001). Timing and interaction of visual cues for prominence in audiovisual speech perception. In *Proceedings of the 7th European Conference on Speech Communication and Technology (pp.* 387–390). https://doi.org/10.21437/Eurospeech.2001-61

Howlin, P., & Magiati, I. (2017). Autism spectrum disorder: Outcomes in adulthood. *Current Opinion in Psychiatry*, *30*(2), 69–76. https://doi.org/10.1097/YCO.0000000000000308

Hurley, R., & Bishop, J. (2016). Prosodic and individual influences on the interpretation of Only. *Speech Prosody*, *8*, 193–197. https://doi.org/10.21437/SpeechProsody.2016-40

Hutchins, T. L., Prelock, P. A., & Bonazinga, L. (2012). Psychometric evaluation of the Theory of Mind Inventory (ToMI): A study of typically developing children and children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *42*(3), 327–341. https://doi.org/10.1007/s10803-011-1244-7

Hutton, S. B., & Nolte, S. (2011). The effect of gaze cues on attention to print advertisements. *Applied Cognitive Psychology*, *25*(6), 887–892. https://doi.org/10.1002/acp.1763

Itier, R. J., & Batty, M. (2009). Neural bases of eye and gaze processing: The core of social cognition. *Neuroscience and Biobehavioral Reviews*, *33*(6), 843–863. https://doi.org/10.1016/j.neubiorev.2009.02.004

Itier, R. J., Villate, C., & Ryan, J. D. (2007). Eyes always attract attention but gaze orienting is task-dependent: Evidence from eye movement monitoring. *Neuropsychologia*, *45*(5), 1019–1028. https://doi.org/10.1016/j.neuropsychologia.2006.09.004

Ito, K., Kryszak, E., & Ibanez, T. (2022). Effect of Prosodic Emphasis on the Processing of Joint-Attention Cues in Children with ASD. In *Proceedings of the International Conference on Speech Prosody* (pp. 110–114). https://doi.org/10.21437/SpeechProsody.2022-23

Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language*, *58*(2), 541–573. https://doi.org/10.1016/j.jml.2007.06.013

Jasinskaja, K., Chiriacescu, S. I., Donazzan, M., Heusinger, K. von, & Hinterwimmer, S. (2015). Prominence in discourse. In A. De Dominicis (Ed.), *Prominences in Linguistics. Proceedings of the pS-prominenceS International Conference* (pp. 134–153). Disucom Press. URN: urn:nbn:de:hbz:38-98937

Jasmin, K., Dick, F., Holt, L., & Tierney, A. (2019). Tailored perception: Individuals' speech and music perception strategies fit their perceptual abilities. *Journal of Experimental Psychology: General*, *149*(5), 914–934. https://doi.org/10.1037/xge0000688

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, *40*(1), 1–19. https://doi.org/10.1016/0010-0277(91)90045-6

Joordens, S., & Hockley, W. E. (2000). Recollection and familiarity through the looking glass: When old does not mirror new. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *26*(6), 1534–1555. https://doi.org/10.1037//0278-7393.26.6.1534

Jording, M., Engemann, D., Eckert, H., Bente, G., & Vogeley, K. (2019). Distinguishing Social from Private Intentions Through the Passive Observation of Gaze Cues. *Frontiers in Human Neuroscience*, *13,* Article 442. https://doi.org/10.3389/fnhum.2019.00442

Jording, M., Hartz, A., Bente, G., Schulte-Rüther, M., & Vogeley, K. (2019). Inferring Interactivity from Gaze Patterns During Triadic Person-Object-Agent Interactions. *Frontiers in Psychology*, *10,* Article 1913. https://doi.org/10.3389/fpsyg.2019.01913

Kaiser, E., & Cohen, A. (2012, September). *In someone else's shoes: A psycholinguistic investigation of free indirect discourse and perspective-taking* [Conference presentation]. Quotation: Perspectives from Philosophy and Linguistics. http://www.ruhr-uni-bochum.de/phil-lang/quotation/mam/pdf/kaisercohen.pdf

Kaland, C., Krahmer, E., & Swerts, M. (2014). White Bear Effects in Language Production: Evidence from the Prosodic Realization of Adjectives. *Language and Speech*, *57*(4), 470–486. https://doi.org/10.1177/0023830913513710

Kargas, N., Lopez, B., Morris, P., & Reddy, V. (2016). Relations Among Detection of Syllable Stress, Speech Abnormalities, and Communicative Ability in Adults with Autism Spectrum Disorders. *Journal of Speech Language and Hearing Research*, *59*(1), 206–215. https://doi.org/10.1044/2015_JSLHR-S-14-0237

Kember, H., Choi, J., Yu, J., & Cutler, A. (2021). The Processing of Linguistic Prominence. *Language and Speech*, *64*(2), 413–436. https://doi.org/10.1177/0023830919880217

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychologica*, *26*, 22–63. https://doi.org/10.1016/0001-6918(67)90005-4

Kendon, A. (1972). Some relationships between body motion and speech. In A. Seigman & B. Pope (Eds.), *Studies in Dyadic Communication* (pp. 177–216). Pergamon Press.

Kendrick, K. H., Holler, J., & Levinson, S. C. (2023). Turn-taking in human face-to-face interaction is multimodal: Gaze direction and manual gestures aid the coordination of turn transitions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *378*(1875), Article 20210473. https://doi.org/10.1098/rstb.2021.0473

Klami, A. (2010). Inferring task-relevant image regions from gaze data. *2010 IEEE International Workshop on Machine Learning for Signal Processing*, 101–106. https://doi.org/10.1109/MLSP.2010.5589230

Klami, A., Saunders, C., de Campos, T. E., & Kaski, S. (2008). Can relevance of images be inferred from eye movements? In *Proceedings of the 1st ACM International Conference on Multimedia Information Retrieval* (pp. 134–140). https://doi.org/10.1145/1460096.1460120

Kleinman, J., Marciano, P. L., & Ault, R. L. (2001). Advanced theory of mind in high-functioning adults with autism. *Journal of Autism and Developmental Disorders*, *31*(1), 29–36. https://doi.org/10.1023/a:1005657512379

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Conceptual Distinctiveness Supports Detailed Visual Long-Term Memory for Real-World Objects. *Journal of Experimental Psychology. General*, *139*(3), 558–578. https://doi.org/10.1037/a0019165

Krahmer, E., Ruttkay, Z., Swerts, M., & Wesselink, W. (2002a). Perceptual evaluation of audiovisual cues for prominence. In *Proceedings of the 7th International Conference on Spoken Language Processing* (pp. 1933–1936). https://doi.org/10.21437/ICSLP.2002-436

Krahmer, E., Ruttkay, Z., Swerts, M., & Wesselink, W. (2002b). Pitch, Eyebrows and the Perception of Focus. In *Proceedings of Speech Prosody* (pp. 443–446). https://doi.org/10.21437/SpeechProsody.2002-96

Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language*, *57*(3), 396–414. https://doi.org/10.1016/j.jml.2007.06.005

Kristensen, L. B., Wang, L., Petersson, K. M., & Hagoort, P. (2013). The Interface Between Language and Attention: Prosodic Focus Marking Recruits a General Attention Network in Spoken Language Comprehension. *Cerebral Cortex*, *23*(8), 1836–1848. https://doi.org/10.1093/cercor/bhs164

Kuhn, G., Benson, V., Fletcher-Watson, S., Kovshoff, H., McCormick, C. A., Kirkby, J., & Leekam, S. R. (2010). Eye movements affirm: Automatic overt gaze and arrow cueing for typical adults and adults with autism spectrum disorder. *Experimental Brain Research*, *201*(2), 155–165. https://doi.org/10.1007/s00221-009-2019-7

Kurumada, C., Brown, M., Bibyk, S., Pontillo, D. F., & Tanenhaus, M. K. (2014). Is it or isn't it: Listeners make rapid use of prosody to infer speaker meanings. *Cognition*, *133*(2), 335–342. https://doi.org/10.1016/j.cognition.2014.05.017

Kushch, O., Igualada, A., & Prieto, P. (2018). Prominence in speech and gesture favour second language novel word learning. *Language, Cognition and Neuroscience*, *33*(8), 992–1004. https://doi.org/10.1080/23273798.2018.1435894

Kuzmanovic, B., Schilbach, L., Lehnhardt, F.-G., Bente, G., & Vogeley, K. (2011). A matter of words: Impact of verbal and nonverbal information on impression formation in high-functioning autism. *Research in Autism Spectrum Disorders*, *5*(1), 604–613. https://doi.org/10.1016/j.rasd.2010.07.005

Lawrence, E. J., Shaw, P., Baker, D., Baron-Cohen, S., & David, A. S. (2004). Measuring empathy: Reliability and validity of the Empathy Quotient. *Psychological Medicine*, *34*(5), 911–919. https://doi.org/10.1017/s0033291703001624

Lee, K., Eskritt, M., Symons, L. A., & Muir, D. (1998). Children's use of triadic eye gaze information for "mind reading." *Developmental Psychology*, *34*(3), 525–539. https://doi.org/10.1037//0012-1649.34.3.525

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian Cognitive Modeling: A Practical Course*. Cambridge University Press.

Lenth, R. V., Bürkner, P., Herve, M., Love, J., Riebl, H., & Singmann, H. (2021). *emmeans: Estimated marginal means, aka least-squares means.* (R package version 1.7.0) [Computer software]. https://cran.r-project.org/web/packages/emmeans/index.html

Leslie, A. M., & Thaiss, L. (1992). Domain specificity in conceptual development: Neuropsychological evidence from autism. *Cognition*, *43*(3), 225–251. https://doi.org/10.1016/0010-0277(92)90013-8

Lever, A. G., & Geurts, H. M. (2016). Age-related differences in cognition across the adult lifespan in autism spectrum disorder. *Autism Research*, *9*(6), 666–676. https://doi.org/10.1002/aur.1545

Lever, A. G., Werkle-Bergner, M., Brandmaier, A., Ridderinkhof, K., & Geurts, H. (2015). Atypical Working Memory Decline Across the Adult Lifespan in Autism Spectrum Disorder? *Journal of Abnormal Psychology*, *124*(4), 1014–1026. https://doi.org/10.1037/abn0000108

Lewis, D. (1970). General Semantics. *Synthese*, *22*(1/2), 18–67.

Light, L. L., & Capps, J. L. (1986). Comprehension of pronouns in young and older adults. *Developmental Psychology*, *22*(4), 580–585. https://doi.org/10.1037/0012-1649.22.4.580

Lin, S., Keysar, B., & Epley, N. (2010). Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology*, *46*(3), 551–556. https://doi.org/10.1016/j.jesp.2009.12.019

Livingston, L. A., & Happé, F. (2017). Conceptualising compensation in neurodevelopmental disorders: Reflections from autism spectrum disorder. *Neuroscience & Biobehavioral Reviews*, *80*, 729–742. https://doi.org/10.1016/j.neubiorev.2017.06.005

Lockwood, H. T., & Macaulay, M. (2012). Prominence Hierarchies. *Language and Linguistics Compass*, *6*(7), 431–446. https://doi.org/10.1002/lnc3.345

Loukusa, S., Mäkinen, L., Kuusikko-Gauffin, S., Ebeling, H., & Moilanen, I. (2014). Theory of mind and emotion recognition skills in children with specific language impairment, autism spectrum disorder and typical development: Group differences and connection to knowledge of grammatical morphology, word-finding abilities and verbal working memory. *International Journal of Language & Communication Disorders*, *49*(4), 498–507. https://doi.org/10.1111/1460-6984.12091

Macdonald, R. G., & Tatler, B. W. (2013). Do as eye say: Gaze cueing and language in a real-world social interaction. *Journal of Vision*, *13*(4), Article 6. https://doi.org/10.1167/13.4.6

MacWhinney, B. (2000). Perspective taking and grammar. *Japanese Society for the Language Sciences*, *1*, 1–25.

Makowski, D., Ben-Shachar, M. S., Chen, S. H. A., & Lüdecke, D. (2019). Indices of Effect Existence and Significance in the Bayesian Framework. *Frontiers in Psychology*, *10*, Article 2767. https://doi.org/10.3389/fpsyg.2019.02767

Malmberg, K. J., & Nelson, T. O. (2003). The word frequency effect for recognition memory and the elevated-attention hypothesis. *Memory & Cognition*, *31*(1), 35–43. https://doi.org/10.3758/BF03196080

Martini, P., & Maljkovic, V. (2009). Short-term memory for pictures seen once or twice. *Vision Research*, *49*(13), 1657–1667. https://doi.org/10.1016/j.visres.2009.04.007

Martino, B. D., Harrison, N. A., Knafo, S., Bird, G., & Dolan, R. J. (2008). Explaining Enhanced Logical Consistency during Decision Making in Autism. *Journal of Neuroscience*, *28*(42), 10746–10750. https://doi.org/10.1523/JNEUROSCI.2895-08.2008

Marvin, R. S., Greenberg, M. T., & Mossler, D. G. (1976). The Early Development of Conceptual Perspective Taking: Distinguishing among Multiple Perspectives. *Child Development*, *47*(2), 511–514. https://doi.org/10.2307/1128810

Mayrand, F., Capozzi, F., & Ristic, J. (2024). Gaze communicates both cue direction and agent mental states. *Frontiers in Psychology*, *15*, Article 1472538. https://doi.org/10.3389/fpsyg.2024.1472538

Mazzarella, E., Hamilton, A., Trojano, L., Mastromauro, B., & Conson, M. (2012). Observation of another's action but not eye gaze triggers allocentric visual perspective. *Quarterly Journal of Experimental Psychology*, *65*(12), 2447–2460. https://doi.org/10.1080/17470218.2012.697905

McGurk, H., & Macdonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748. https://doi.org/10.1038/264746a0

Melcher, D. (2001). Persistence of visual memory for scenes. *Nature*, *412*(6845), Article 6845. https://doi.org/10.1038/35086646

Melcher, D. (2006). Accumulation and persistence of memory for natural scenes. *Journal of Vision*, *6*(1), Article 2. https://doi.org/10.1167/6.1.2

Mixdorff, H., Hönemann, A., & Fagel, S. (2013). Integration of Acoustic and Visual Cues in Prominence Perception. In *Proceedings of Auditory-Visual Speech Processing* (pp. 230–234). International Speech Communication Association. https://doi.org/10.21437/Interspeech.2013-73

Mizuno, A., Liu, Y., Williams, D. L., Keller, T. A., Minshew, N. J., & Just, M. A. (2011). The neural basis of deictic shifting in linguistic perspective-taking in high-functioning autism. *Brain: A Journal of Neurology*, *134*(8), 2422–2435. https://doi.org/10.1093/brain/awr151

Morett, L. M., & Fraundorf, S. H. (2019). Listeners consider alternative speaker productions in discourse comprehension and memory: Evidence from beat gesture and pitch accenting. *Memory & Cognition*, *47*(8), 1515–1530. https://doi.org/10.3758/s13421-019-00945-1

Müllner, D. (2013). fastcluster: Fast Hierarchical, Agglomerative Clustering Routines for R and Python. *Journal of Statistical Software*, *53*(1), 1–18. https://doi.org/10.18637/jss.v053.i09

Munhall, K. G., Jones, J., Callan, D., Kuratate, T., & Vatikiotis-Bateson, E. (2004). Visual Prosody and Speech Intelligibility Head Movement Improves Auditory Speech Perception. *Psychological Science*, *15*, 133–137. https://doi.org/10.1111/j.0963-7214.2004.01502010.x

Murray, K., Johnston, K., Cunnane, H., Kerr, C., Spain, D., Gillan, N., Hammond, N., Murphy, D., & Happé, F. (2017). A new test of advanced theory of mind: The "Strange Stories Film Task" captures social processing differences in adults with autism spectrum disorders. *Autism Research*, *10*(6), 1120–1132. https://doi.org/10.1002/aur.1744

Norbury, C. Frazier. (2005). The relationship between theory of mind and metaphor: Evidence from children with language impairment and autistic spectrum disorder. *British Journal of Developmental Psychology*, *23*(3), 383–399. https://doi.org/10.1348/026151005X26732

O'Connor, J. D., & Arnold, G. F. (1973). *Intonation of Colloquial English* (2nd ed.). Longman.

O'Connor, K. (2012). Auditory processing in autism spectrum disorder: A review. *Neuroscience & Biobehavioral Reviews*, *36*(2), 836–854. https://doi.org/10.1016/j.neubiorev.2011.11.008

O'Hearn, K., Tanaka, J., Lynn, A., Fedor, J., Minshew, N., & Luna, B. (2014). Developmental plateau in visual object processing from adolescence to adulthood in autism. *Brain and Cognition*, *90*, 124–134. https://doi.org/10.1016/j.bandc.2014.06.004

Ozonoff, S., Pennington, B. F., & Rogers, S. J. (1991). Executive function deficits in high-functioning autistic individuals: Relationship to theory of mind. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *32*(7), 1081–1105. https://doi.org/10.1111/j.1469-7610.1991.tb00351.x

Pantelis, P. C., & Kennedy, D. P. (2017). Deconstructing atypical eye gaze perception in autism spectrum disorder. *Scientific Reports*, *7*(1), Article 14990. https://doi.org/10.1038/s41598-017-14919-3

Paul, R., Augustyn, A., Klin, A., & Volkmar, F. R. (2005). Perception and Production of Prosody by Speakers with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *35*(2), 205–220. https://doi.org/10.1007/s10803-004-1999-1

Pearson, A., Ropar, D., & Hamilton, A. F. de C. (2013). A review of visual perspective taking in autism spectrum disorder. *Frontiers in Human Neuroscience*, *7,* Article 652. https://doi.org/10.3389/fnhum.2013.00652

Peñuelas-Calvo, I., Sareen, A., Sevilla-Llewellyn-Jones, J., & Fernández-Berrocal, P. (2019). The "Reading the Mind in the Eyes" Test in Autism-Spectrum Disorders Comparison with Healthy Controls: A Systematic Review and Meta-analysis. *Journal of Autism and Developmental Disorders*, *49*(3), 1048–1061. https://doi.org/10.1007/s10803-018-3814-4

Perner, J., & Wimmer, H. (1985). "John thinks that Mary thinks that…" attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology*, *39*(3), 437–471. https://doi.org/10.1016/0022-0965(85)90051-7

Peterson, C. C., Wellman, H. M., & Slaughter, V. (2012). The mind behind the message: Advancing theory-of-mind scales for typically developing children, and those with deafness, autism, or Asperger syndrome. *Child Development*, *83*(2), 469–485. https://doi.org/10.1111/j.1467-8624.2011.01728.x

Ponnet, K. S., Roeyers, H., Buysse, A., De Clercq, A., & Van der Heyden, E. (2004). Advanced mind-reading in adults with Asperger syndrome. *Autism: The International Journal of Research and Practice*, *8*(3), 249–266. https://doi.org/10.1177/1362361304045214

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(4), 515–526. https://doi.org/10.1017/S0140525X00076512

Prieto, P., Puglesi, C., Borràs-Comes, J., Arroyo, E., & Blat, J. (2015). Exploring the contribution of prosody and gesture to the perception of focus using an animated agent. *Journal of Phonetics*, *49*, 41–54. https://doi.org/10.1016/j.wocn.2014.10.005

Quadflieg, S., Mason, M. F., & Macrae, C. N. (2004). The owl and the pussycat: Gaze cues and visuospatial orienting. *Psychonomic Bulletin & Review*, *11*(5), Article 826. https://doi.org/10.3758/BF03196708

Qureshi, A. W., Apperly, I. A., & Samson, D. (2010). Executive function is necessary for perspective selection, not Level-1 visual perspective calculation: Evidence from a dual-task study of adults. *Cognition*, *117*(2), 230–236. https://doi.org/10.1016/j.cognition.2010.08.003

R Core Team (2019). *R: A language and environment for statistical computing* (Version 3.6.0) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org

Remington, A., & Fairnie, J. (2017). A sound advantage: Increased auditory capacity in autism. *Cognition*, *166*, 459–465. https://doi.org/10.1016/j.cognition.2017.04.002

Ren, Y., Li, H., Li, Y., Xu, Z., Luo, R., Ping, H., Ni, X., Yang, J., & Yang, W. (2023). Sustained visual attentional load modulates audiovisual integration in older and younger adults. *I-Perception*, *14*(1), eLocator 20416695231157348. https://doi.org/10.1177/20416695231157348

Riby, D. M., Hancock, P. J., Jones, N., & Hanley, M. (2013). Spontaneous and cued gaze-following in autism and Williams syndrome. *Journal of Neurodevelopmental Disorders*, *5*(1), Article 13. https://doi.org/10.1186/1866-1955-5-13

Ricciardelli, P., Bricolo, E., Aglioti, S. M., & Chelazzi, L. (2002). My eyes want to look where your eyes are looking: Exploring the tendency to imitate another individual's gaze. *Neuroreport*, *13*(17), 2259–2264. https://doi.org/10.1097/01.wnr.0000044227.79663.2e

Ring, M., Gaigg, S. B., & Bowler, D. M. (2015). Object-location memory in adults with autism spectrum disorder. *Autism Research*, *8*(5), 609–619. https://doi.org/10.1002/aur.1478

Ristic, J., Mottron, L., Friesen, C. K., Iarocci, G., Burack, J. A., & Kingstone, A. (2005). Eyes are special but not for everyone: The case of autism. *Cognitive Brain Research*, *24*(3), 715–718. https://doi.org/10.1016/j.cogbrainres.2005.02.007

Roettger, T., Mahrt, T., & Cole, J. (2019). Mapping prosody onto meaning-the case of information structure in American English. *Language, Cognition and Neuroscience, 34*(7), 841–860. https://doi.org/10.1080/23273798.2019.1587482

Roettger, T., Turner, D., & Cole, J. (2020, October 14). Intonational processing is incremental and holistic. *PsyArXiv*. https://doi.org/10.31234/osf.io/nhbgs

Rohrer, P. L., Delais-Roussarie, E., & Prieto, P. (2023). Visualizing prosodic structure: Manual gestures as highlighters of prosodic heads and edges in English academic discourses. *Lingua*, *293*, Article 103583. https://doi.org/10.1016/j.lingua.2023.103583

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics*, *1*(1), 75–116. https://doi.org/10.1007/BF02342617

Rosenblau, G., Kliemann, D., Dziobek, I., & Heekeren, H. R. (2017). Emotional prosody processing in autism spectrum disorder. *Social Cognitive and Affective Neuroscience*, *2*(12), 224–239. https://doi.org/10.1093/scan/nsw118

Roy, J., Cole, J., & Mahrt, T. (2017). Individual differences and patterns of convergence in prosody perception. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *8*(1), Article 22. https://doi.org/10.5334/labphon.108

RStudio Team (2016). *RStudio: Integrated Development for R* (Version 1.1.423) [Computer software]. RStudio, Inc. http://www.rstudio.com

Rutherford, M. D., Baron-Cohen, S., & Wheelwright, S. (2002). Reading the mind in the voice: A study with normal adults and adults with Asperger syndrome and high functioning autism. *Journal of Autism and Developmental Disorders*, *32*(3), 189–194. https://doi.org/10.1023/a:1015497629971

Sajjacholapunt, P., & Ball, L. (2014). The influence of banner advertisements on attention and memory: Human faces with averted gaze can enhance advertising effectiveness. *Frontiers in Psychology*, *5,* Article 166. https://doi.org/10.3389/fpsyg.2014.00166

Salem, S., Weskott, T., & Holler, A. (2017). Does narrative perspective influence readers' perspective-taking? An empirical study on free indirect discourse, psycho-narration and first-person

narration. *Glossa: A Journal of General Linguistics*, *2*(1), Article 1. https://doi.org/10.5334/gjgl.225

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(5), 1255–1266. https://doi.org/10.1037/a0018729

Scarborough, R., Keating, P., Mattys, S., Cho, T., & Alwan, A. (2009). Optical Phonetics and Visual Perception of Lexical and Phrasal Stress in English. *Language and Speech*, *52*, 135–175. https://doi.org/10.1177/0023830909103165

Scheeren, A. M., Rosnay, M. de, & Koot, H. M. (2013). Rethinking theory of mind in high-functioning autism spectrum disorder. *Journal of Child Psychology and Psychiatry*, *54*(6), 628–635. https://doi.org/10.1111/jcpp.12007

Schelinski, S., Roswandowitz, C., & von Kriegstein, K. (2017). Voice identity processing in autism spectrum disorder. *Autism Research: Official Journal of the International Society for Autism Research*, *10*(1), 155–168. https://doi.org/10.1002/aur.1639

Schelinski, S., & von Kriegstein, K. (2019). The Relation Between Vocal Pitch and Vocal Emotion Recognition Abilities in People with Autism Spectrum Disorder and Typical Development. *Journal of Autism and Developmental Disorders*, *49*(1), 68–82. https://doi.org/10.1007/s10803-018-3681-z

Schneider, D., Slaughter, V. P., Bayliss, A. P., & Dux, P. E. (2013). A temporally sustained implicit theory of mind deficit in autism spectrum disorders. *Cognition*, *129*(2), 410–417. https://doi.org/10.1016/j.cognition.2013.08.004

Schuh, J., Eigsti, I.-M., & Mirman, D. (2016). Discourse comprehension in autism spectrum disorder: Effects of working memory load and common ground. *Autism Research: Official Journal of the International Society for Autism Research*, *9*(12), 1340–1352. https://doi.org/10.1002/aur.1632

Schulman, A. I. (1967). Word length and rarity in recognition memory. *Psychonomic Science*, *9*(4), 211–212. https://doi.org/10.3758/BF03330834

Schuwerk, T., Vuori, M., & Sodian, B. (2015). Implicit and explicit Theory of Mind reasoning in autism spectrum disorders: The impact of experience. *Autism: The International Journal of Research and Practice*, *19*(4), 459–468. https://doi.org/10.1177/1362361314526004

Schwarzkopf, S., Schilbach, L., Vogeley, K., & Timmermans, B. (2014). "Making it explicit" makes a difference: Evidence for a dissociation of spontaneous and intentional level 1 perspective taking in high-functioning autism. *Cognition*, *131*(3), 345–354. https://doi.org/10.1016/j.cognition.2014.02.003

Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind Eyes: An Absence of Spontaneous Theory of Mind in Asperger Syndrome. *Science*, *325*(5942), 883–885. https://doi.org/10.1126/science.1176170

Setien-Ramos, I., Lugo-Marín, J., Gisbert-Gustemps, L., Díez-Villoria, E., Magán-Maganto, M., Canal-Bedia, R., & Ramos-Quiroga, J. A. (2022). Eye-Tracking Studies in Adults with Autism Spectrum Disorder: A Systematic Review and Meta-analysis. *Journal of Autism and Developmental Disorders,* 53, 1–14. https://doi.org/10.1007/s10803-022-05524-z

Shaked, M., Gamliel, I., & Yirmiya, N. (2006). Theory of mind abilities in young siblings of children with autism. *Autism: The International Journal of Research and Practice*, *10*(2), 173–187. https://doi.org/10.1177/1362361306062023

Sharda, M., Subhadra, T. P., Sahay, S., Nagaraja, C., Singh, L., Mishra, R., Sen, A., Singhal, N., Erickson, D., & Singh, N. C. (2010). Sounds of melody – Pitch patterns of speech in autism. *Neuroscience Letters*, *478*(1), 42–45. https://doi.org/10.1016/j.neulet.2010.04.066

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, *6*(1), 156–163. https://doi.org/10.1016/S0022-5371(67)80067-7

Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature Neuroscience*, *6*(12), 1317–1322. https://doi.org/10.1038/nn1150

Spaniol, M., Janz, A., Wehrle, S., Vogeley, K., & Grice, M. (2023, April 28). Multimodal signalling: The interplay of oral and visual feedback in conversation. In *Proceedings of the International Congress of Phonetic Sciences* (pp. 4110- 4114).

Srinivasan, R. J., & Massaro, D. W. (2003). Perceiving Prosody from the Face and Voice: Distinguishing Statements from Echoic Questions in English. *Language and Speech*, *46*(1), 1–22. https://doi.org/10.1177/00238309030460010201

Stalnaker, R. (2002). Common Ground. *Linguistics and Philosophy*, *25*(5), 701–721. https://doi.org/10.1023/A:1020867916902

Steele, S., Joseph, R. M., & Tager-Flusberg, H. (2003). Brief Report: Developmental Change in Theory of Mind Abilities in Children with Autism. *Journal of Autism and Developmental Disorders*, *33*(4), 461–467. https://doi.org/10.1023/A:1025075115100

Stewart, M. E., McAdam, C., Ota, M., Peppé, S., & Cleland, J. (2013). Emotional recognition in autism spectrum conditions from voices and faces. *Autism*, *17*(1), 6–14. https://doi.org/10.1177/1362361311424572

Swerts, M., & Krahmer, E. (2008). Facial expression and prosodic prominence: Effects of modality and facial area. *Journal of Phonetics*, *36*(2), 219–238. https://doi.org/10.1016/j.wocn.2007.05.001

Swerts, M., & Krahmer, E. (2010). Visual prosody of newsreaders: Effects of information structure, emotional content and intended audience on facial expressions. *Journal of Phonetics*, *38*(2), 197–206. https://doi.org/10.1016/j.wocn.2009.10.002

Swettenham, J. G. (1996). What's Inside Someone's Head? Conceiving of the Mind as a Camera Helps Children with Autism Acquire an Alternative to a Theory of Mind. *Cognitive Neuropsychiatry*, *1*(1), 73–88. https://doi.org/10.1080/135468096396712

Taylor, L. J., Lev-Ari, S., & Zwaan, R. A. (2008). Inferences about action engage action systems. *Brain and Language*, *107*(1), 62–67. https://doi.org/10.1016/j.bandl.2007.08.004

Terken, J. (1991). Fundamental frequency and perceived prominence of accented syllables. *The Journal of the Acoustical Society of America*, *89*, 1768–1776. https://doi.org/10.1121/1.401019

Terrizzi, B. F., & Beier, J. S. (2016). Automatic cueing of covert spatial attention by a novel agent in preschoolers and adults. *Cognitive Development*, *40*, 111–119. https://doi.org/10.1016/j.cogdev.2016.08.001

Theeuwes, J., Belopolsky, A., & Olivers, C. N. L. (2009). Interactions between working memory, attention and eye movements. *Acta Psychologica*, *132*(2), 106–114. https://doi.org/10.1016/j.actpsy.2009.01.005

Theuring, C., Gredebäck, G., & Hauf, P. (2007). Object processing during a joint gaze following task. *European Journal of Developmental Psychology*, *4*(1), 65–79. https://doi.org/10.1080/17405620601051246

Thézé, R., Gadiri, M. A., Albert, L., Provost, A., Giraud, A.-L., & Mégevand, P. (2020). Animated virtual characters to explore audio-visual speech in controlled and naturalistic environments. *Scientific Reports*, *10*(1), Article 1. https://doi.org/10.1038/s41598-020-72375-y

Tomasino, B., Werner, C. J., Weiss, P. H., & Fink, G. R. (2007). Stimulus properties matter more than perspective: An fMRI study of mental imagery and silent reading of action phrases. *NeuroImage*, *36*, T128–T141. https://doi.org/10.1016/j.neuroimage.2007.03.035

Tottenham, N., Hertzig, M. E., Gillespie-Lynch, K., Gilhooly, T., Millner, A. J., & Casey, B. J. (2014). Elevated amygdala response to faces and gaze aversion in autism spectrum disorder. *Social Cognitive and Affective Neuroscience*, *9*(1), 106–117. https://doi.org/10.1093/scan/nst050

van Tiel, B., Deliens, G., Geelhand, P., Murillo Oosterwijk, A., & Kissine, M. (2021). Strategic Deception in Adults with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *51*(1), 255–266. https://doi.org/10.1007/s10803-020-04525-0

Velloso, R. de L., Duarte, C. P., & Schwartzman, J. S. (2013). Evaluation of the theory of mind in autism spectrum disorders with the Strange Stories test. *Arquivos De Neuro-Psiquiatria*, *71*(11), 871–876. https://doi.org/10.1590/0004-282X20130171

Vlamings, P. H. J. M., Stauder, J. E. A., van Son, I. A. M., & Mottron, L. (2005). Atypical visual orienting to gaze- and arrow-cues in adults with high functioning autism. *Journal of Autism and Developmental Disorders*, *35*(3), 267–277. https://doi.org/10.1007/s10803-005-3289-y

Vogels, J., Bimpikou, S., Kapelle, O., & Maier, E. (2024, June 17). *Taking the perspective of narrative characters: A mouse-tracking study on the processing of ambiguous referring expressions in narrative discourse*. https://doi.org/10.17605/OSF.IO/3R7BQ

Vroomen, J., Driver, J., & de Gelder, B. (2001). Is cross-modal integration of emotional expressions independent of attentional resources? *Cognitive, Affective & Behavioral Neuroscience*, *1*(4), 382–387. https://doi.org/10.3758/cabn.1.4.382

Wagner, P., Ćwiek, A., & Samlowski, B. (2019). Exploiting the speech-gesture link to capture fine-grained prosodic prominence impressions and listening strategies. *Journal of Phonetics, 76,* Article 100911. https://doi.org/10.1016/j.wocn.2019.07.001

Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction: An overview. *Speech Communication*, *57*, 209–232. https://doi.org/10.1016/j.specom.2013.09.008

Wahl, S., Marinović, V., & Träuble, B. (2019). Gaze cues of isolated eyes facilitate the encoding and further processing of objects in 4-month-old infants. *Developmental Cognitive Neuroscience*, *36*, Article 100621. https://doi.org/10.1016/j.dcn.2019.100621

Wang, L., Beaman, C. P., Jiang, C., & Liu, F. (2022). Perception and Production of Statement-Question Intonation in Autism Spectrum Disorder: A Developmental Investigation. *Journal of Autism and Developmental Disorders*, *52*(8), 3456–3472. https://doi.org/10.1007/s10803-021-05220-4

Wang, L., & Chu, M. (2013). The role of beat gesture and pitch accent in semantic processing: An ERP study. *Neuropsychologia*, *51*(13), 2847–2855. https://doi.org/10.1016/j.neuropsychologia.2013.09.027

Wang, S., Jiang, M., Duchesne, X. M., Laugeson, E. A., Kennedy, D. P., Adolphs, R., & Zhao, Q. (2015). Atypical Visual Saliency in Autism Spectrum Disorder Quantified through Model-Based Eye Tracking. *Neuron*, *88*(3), 604–616. https://doi.org/10.1016/j.neuron.2015.09.042

Wardlow, L. (2013). Individual Differences in Speaker's Perspective Taking: The Roles of Executive Control and Working Memory. *Psychonomic Bulletin & Review*, *20*(4), 766–772. https://doi.org/10.3758/s13423-013-0396-1

Watson, D. G., Tanenhaus, M. K., & Gunlogson, C. A. (2008). Interpreting Pitch Accents in Online Comprehension: H* vs. L+H*. *Cognitive Science*, *32*(7), 1232–1244. https://doi.org/10.1080/03640210802138755

Weber, A., Braun, B., & Crocker, M. W. (2006). Finding Referents in Time: Eye-Tracking Evidence for the Role of Contrastive Accents. *Language and Speech*, *49*(3), 367–392. https://doi.org/10.1177/00238309060490030301

Wehrle, S., Cangemi, F., Vogeley, K., & Grice, M. (2022). New evidence for melodic speech in Autism Spectrum Disorder. In *Proceedings of Speech Prosody* (pp. 37–41). https://doi.org/10.21437/SpeechProsody.2022-8

White, S. J., Coniston, D., Rogers, R., & Frith, U. (2011). Developing the Frith-Happé animations: A quick and objective test of Theory of Mind for adults with autism. *Autism Research: Official Journal of the International Society for Autism Research*, *4*(2), 149–154. https://doi.org/10.1002/aur.174

Wilson, C. E., Happé, F., Wheelwright, S. J., Ecker, C., Lombardo, M. V., Johnston, P., Daly, E., Murphy, C. M., Spain, D., Lai, M.-C., Chakrabarti, B., Sauter, D. A., Baron-Cohen, S., & Murphy, D. G. M. (2014). The Neuropsychology of Male Adults with High-Functioning Autism or Asperger Syndrome. *Autism Research*, *7*(5), 568–581. https://doi.org/10.1002/aur.1394

World Health Organization (2019). *ICD-11: International Classification of Diseases, 11th Revision. The global standard for diagnostic health information*. https://icd.who.int/

World Medical Association (2013). World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. *Journal of the American Medical Association, 310*(20), 2191–2194. https://doi.org/10.1001/jama.2013.281053

Yarbus, A. L. (1967). Eye Movements During Perception of Complex Objects. In A. L. Yarbus (Ed.), *Eye Movements and Vision* (pp. 171–211). Springer US. https://doi.org/10.1007/978-1-4899-5379-7_8

Zaidenberg, H. (2015). *Prosodic Deficit in the Perception of Focus: Evidence from Hebrew Speaking Individuals with Asperger Syndrome* [Unpublished Master's Thesis]. Tel Aviv University.

Zarcone, A., van Schijndel, M., Vogels, J., & Demberg, V. (2016). Salience and Attention in Surprisal-Based Accounts of Language Processing. *Frontiers in Psychology*, *7*, Article 844. https://doi.org/10.3389/fpsyg.2016.00844

Zeidan, J., Fombonne, E., Scorah, J., Ibrahim, A., Durkin, M. S., Saxena, S., Yusuf, A., Shih, A., & Elsabbagh, M. (2022). Global prevalence of autism: A systematic review update. *Autism Research*, *15*(5), 778–790. https://doi.org/10.1002/aur.2696

Zeman, S. (2017). Confronting perspectives: Modeling perspectival complexity in language and cognition. *Glossa: A Journal of General Linguistics*, *2*(1), Article 6. https://doi.org/10.5334/gjgl.213

Zhang, M., Xu, S., Chen, Y., Lin, Y., Ding, H., & Zhang, Y. (2022). Recognition of affective prosody in autism spectrum conditions: A systematic review and meta-analysis. *Autism*, *26*(4), 798–813. https://doi.org/10.1177/1362361321995725

Zhang, Y., Xiang, Y., Guo, Y., & Zhang, L. (2018). Beauty-related perceptual bias: Who captures the mind of the beholder? *Brain and Behavior*, *8*(5), Article e00945. https://doi.org/10.1002/brb3.945

Zimmermann, J. T., Ellison, T. M., Cangemi, F., Wehrle, S., Vogeley, K., & Grice, M. (2024). Lookers and listeners on the autism spectrum: The roles of gaze duration and pitch height in inferring mental states. *Frontiers in Communication*, *9,* Article 1483135. https://doi.org/10.3389/fcomm.2024.1483135

Zimmermann, J. T., Meuser, S., Hinterwimmer, S., & Vogeley, K. (2021). Preserved perspective taking in free indirect discourse in autism spectrum disorder. *Frontiers in Psychology*, *12,* Article 675633. https://doi.org/10.3389/fpsyg.2021.675633

Zimmermann, J. T., Wehrle, S., Cangemi, F., Grice, M., & Vogeley, K. (2020). Listeners and Lookers: Using pitch height and gaze duration for inferring mental states. In *Proceedings of the 10th International Conference on Speech Prosody* (pp. 290–294). https://doi.org/10.21437/SpeechProsody.2020-59

Zıvralı Yarar, E., Howlin, P., Charlton, R., & Happé, F. (2021). Age-related effects on social cognition in adults with Autism Spectrum Disorder: A possible protective effect on theory of mind. *Autism Research: Official Journal of the International Society for Autism Research*, *14*(5), 911–920. https://doi.org/10.1002/aur.2410

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*(2), 162–185. https://doi.org/10.1037/0033-2909.123.2.162

Zwaan, R. A., & Taylor, L. J. (2006). Seeing, acting, understanding: Motor resonance in language comprehension. *Journal of Experimental Psychology: General*, *135*(1), 1–11. https://doi.org/10.1037/0096-3445.135.1.1

Zwickel, J., White, S. J., Coniston, D., Senju, A., & Frith, U. (2011). Exploring the building blocks of social cognition: Spontaneous agency perception and visual perspective taking in autism. *Social Cognitive and Affective Neuroscience*, *6*(5), 564–571. https://doi.org/10.1093/scan/nsq088

11 Acknowledgments

I would like to thank my supervisors Professor Dr Dr Kai Vogeley and Professor Dr Martine Grice who gave me the opportunity to work in the scope of their departments' cooperation and thus provided me with an inspiring interdisciplinary research environment. Thank you very much, Kai, for your consult, your support and your invaluable experience, but also for giving me a lot of freedom. Martine, I am deeply grateful for the opportunity to not only enjoy your phonetic expertise, but also your constructive, compassionate mindset and your motivating guidance.

I want to express my gratitude to the IPHS, in particular to my tutors Prof Dr Elke Kalbe and Prof Dr Manfred Döpfner for accompanying me throughout the PhD program "Health Sciences", and to Prof Dr Peter Weiss-Blankenhorn for reviewing my thesis.

Further, I would like to thank the CRC "Prominence in Language" for giving me the opportunity to pursue my studies and immerse into a whole new world. I would like to thank the CRC's members for being a knowledgeable and constant source of feedback, support and inspiration. I am especially grateful to my co-authors who devoted their time and energy to these projects.

A big "Thank you!" goes to all the (former) members and associates of the research group "Social Cognition / Autism" of the Department of Psychiatry at the University Hospital Cologne and of the Phonetic Institute at the University of Cologne. Thank you for your support, advice and ideas throughout the different phases of this project! I would like to especially thank Carola, Simon and Francesco.

I would like to also thank my spirit animals, my family and my friends near and far. Thank you, Kae, Milena and Susi for your brains! Alex, Cansu and Jana, I am very grateful not only for your cognitive investment, but for you being my everyday (PhD) support group, reliable patronuses and an endless well of energy.

Without the participation of the participants taking part in these experiments, this thesis would not have been possible. Many thanks to everyone for investing their time and effort!

12 Eidesstattliche Erklärung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle Stellen – einschließlich Tabellen, Karten und Abbildungen –, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, sind in jedem Einzelfall als Entlehnung kenntlich gemacht. Ich versichere an Eides statt, dass diese Dissertationsschrift noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie – abgesehen von unten angegebenen Teilpublikationen – noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss der Promotion nicht ohne Genehmigung der / des Vorsitzenden des IPHS-Promotionsausschusses vornehmen werde. Die Bestimmungen dieser Ordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Dr. Kai Vogeley und Prof. Dr. Martine Grice betreut worden.

Darüber hinaus erkläre ich hiermit, dass ich die Ordnung zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten der Universität zu Köln gelesen und sie bei der Durchführung der Dissertation beachtet habe und verpflichte mich hiermit, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten zu beachten und umzusetzen.

12.1 Veröffentlichte Publikationen und Eigenanteil

Die dieser kumulativen Dissertation zugrunde liegenden veröffentlichten Studien sind von mir mit Unterstützung von Prof. Dr. Dr. Kai Vogeley und Prof. Dr. Martine Grice durchgeführt worden.

Zimmermann, J. T., Wehrle, S., Cangemi, F., Grice, M., & Vogeley, K. (2020). Listeners and Lookers: Using pitch height and gaze duration for inferring mental states. In *Proceedings of the 10th International Conference on Speech Prosody* (pp. 290–294). DOI: 10.21437/SpeechProsody. 2020-59. Es handelt sich um eine Publikation zu einer empirischen Untersuchung, die in einem Peer-Review-Verfahren von vier Reviewerinnen bzw. Reviewern begutachtet wurde. Diese stammen aus einem Pool an externen Reviewerinnen und Reviewern sowie Mitgliedern des Konferenz-Komitees. Die Arbeiten an diesem Projekt sind in enger Zusammenarbeit mit dem IfL(Institute for Language) – Phonetik der Universität zu Köln entstanden. Die Promotionskandidatin war mit Prof. Dr. Dr. Kai Vogeley und Prof. Dr. Martine Grice verantwortlich für die Konzeption des Projekts. Die Erstellung des Audiomaterials erfolgte in Kooperation mit Dr. Simon Wehrle und Dr. Francesco Cangemi. Für die sonstige Vorbereitung der Studie, u.a. die Programmierung des Experiments, war maßgeblich die

Promotionskandidatin verantwortlich, ebenso wie für die Leitung der Studie, die Datenerhebung und -auswertung sowie die Manuskripterstellung.

Zimmermann, J. T., Ellison, T. M., Cangemi, F., Wehrle, S., Vogeley, K. & Grice, M. (2024). Lookers and listeners on the autism spectrum: The roles of gaze duration and pitch height in inferring mental states. *Frontiers in Communication, 9,* Article 1483135. DOI: 10.3389/fcomm.2024.1483135. Die Arbeiten an diesem Projekt sind in enger Zusammenarbeit mit dem IfL(Institute for Language) – Phonetik der Universität zu Köln entstanden. Die Promotionskandidatin war mit Prof. Dr. Dr. Kai Vogeley und Prof. Dr. Martine Grice verantwortlich für die Konzeption des Projekts. Die Erstellung des Audiomaterials erfolgte in Kooperation mit Dr. Francesco Cangemi und Dr. Simon Wehrle. Für die sonstige Vorbereitung der Studie, u.a. die Programmierung des Experiments, war maßgeblich die Promotionskandidatin verantwortlich, ebenso wie für die Leitung der Studie, die Datenerhebung sowie die Manuskripterstellung. Sie war zusammen mit T. Mark Ellison verantwortlich für die Datenauswertung.

Zimmermann, J. T., Meuser, S., Hinterwimmer, S., & Vogeley, K. (2021). Preserved perspective taking in free indirect discourse in autism spectrum disorder. *Frontiers in Psychology*, *12*, Article 675633. DOI: 10.3389/fpsyg.2021.675633. Bei diesem Projekt handelt es sich um eine Kooperation mit dem Institut für deutsche Sprache und Literatur I der Universität zu Köln. Die Promotions-kandidatin war gemeinsam mit den anderen Autorinnen bzw. Autoren verantwortlich für die Konzeption der Studie. Sie war zusammen mit Sara Meuser verantwortlich für die Vorbereitung der Online-Studie, die Studienleitung sowie die Datenerhebung. Die Promotionskandidatin war darüber hinaus maßgeblich verantwortlich für die Datenauswertung sowie die Manuskripterstellung.

Ich versichere, dass ich alle Angaben wahrheitsgemäß nach bestem Wissen und Gewissen gemacht habe und verpflichte mich, jedmögliche, die obigen Angaben betreffenden Veränderungen, dem IPHS-Promotionsausschuss unverzüglich mitzuteilen.

.....

Datum Unterschrift