Tracing the Neurocognitive Sources of Communicative Challenges in Autism Spectrum Disorder

Kexin Cai[1]

Supervisors: Dr. Arjen Stolk[1,2], Dr. Jana Bašnáková[1], Dr. Saskia Koch[1], Dr. Ivan Toni[1]

[1] *Donders Institute for Brain, Cognition, and Behavior, Radboud University, Nijmegen, The Netherlands*

[2] *Psychological and Brain Sciences, Dartmouth College, Hanover, NH, USA*

Autism Spectrum Disorder (ASD) is diagnosed on the basis of communicative challenges experienced in everyday interactions, yet the neurocognitive sources of those challenges remain largely unknown. A popular suggestion holds that autistic individuals struggle to predict other people's behaviors, either due to a bottleneck in bottom-up sensory processing or an attenuated sensitivity to top-down priors when perceiving communicative behaviors. However, studies of predictive processing have produced mixed results in the perceptual domain, and few studies to date have investigated ASD predictive capabilities in the context of live social interaction. This dissertation takes the first steps to bridge this gap in assessing bottom-up and top-down contributions to interpreting genuinely interactive communicative behaviors in autistic and neurotypical individuals. I report the construction and validation of a theoretically grounded methodological framework that allows the simultaneous recording of eye gaze and electroencephalographical activity from pairs of participants engaged in dynamically unfolding communicative interactions. Predictive capabilities are assessed while participants solve a series of non-verbal coordination problems in the two-player Tacit Communication Game. A benefit of this novel communicative medium is that it offers the possibility to manipulate access to top-down priors during communicative interpretation by introducing coordination problems that are more easily solved in the light of previous interactions. Moreover, the computer-controlled interactions allow for temporally precise measurement of communicative behavior to determine any bottlenecks in bottom-up sensory processing. I demonstrate how eye tracking can be employed to assess bottom-up and top-down processing in this live social context through the use of qualitative and quantitative methods and potential methodologies that can be used to further investigate the neurocognitive sources of autistic communicative challenges.

Keywords: social interaction, conceptual alignment, mutual understanding, eye-tracking, autism

Corresponding author: Kexin Cai; E-mail: kexincai626@gmail.com

Interactive context construction in human communication

Everyday interaction, though smooth as it typically appears to be, is full of ambiguity, metaphor, implicature, and the occasional misdirection. We travel with the word's uses through "a complicated network of similarities overlapping and criss-crossing" (Wittgenstein, 2010). How human minds manage to remain connected while they traverse these often unique paths is a key question at the roots of human sociality. In a conversation, interlocutors with idiosyncratic conceptual worlds appear to cast their respective mental representations to each other, but it remains a mystery how one could catch the fleeting meaning casted by the other at the time of the delivery and how this *cast-catch* game lands the players on a "common ground" eventually—not necessarily agree on but at least converge on one issue. Plain sailing as daily communication is, it's easy to overlook the complexity of the underlying problem, leading to previous attempts to carve the nature of human communication by reducing it to a signal transmission process.

Signal-centered frameworks. Starting from Shannon's (1948) mathematical theory of communication, psychologists (e.g., Liberman & Mattingly, 1985) have comparably defined human communication as a system that consists of a sender and a receiver using coding and decoding mechanisms, respectively; and a physical medium for transferring signals. Recent empirical neuroscientific studies also attribute the retrieval and selection of the meaning to the signal itself, e.g., neural synchrony (Hasson et al., 2012), and shared sensorimotor associations (Pickering & Garrod, 2013; Pulvermüller et al., 2014) are the common mechanisms to implement information transfer. Admittedly the signal transmission view constitutes a consequential aspect of the essence of human communication. However, it cannot exclusively account for the nature of human communication; nor where its true complexity and distinctiveness reside.

Signal-centered frameworks implicitly or explicitly assume that communicators share the same set of pre-defined coding-decoding rules, e.g., a common language (Akmajian, Farmer, Bickmore, et al., 2017), which is, however, neither a necessity nor sufficient condition for mutual understanding. The view is falsified by two contrasting types of observations in daily life, namely *tourists' successes* and *robots' failures*. The former elucidates the unnecessity of pre-defined coding-decoding rules for successful communication. Imagine a foreign tourist shopping in a supermarket in the Netherlands. At the counter, the cashier waves the receipt to her and asks, "Wilt u het bonnetje?". The tourist cannot decode the speech signal without prior knowledge of Dutch, but drawing on inferences from her previous shopping experiences, and the current context—it happens after paying the bill, and the cashier is waving the receipt; the tourist grasp the gist of the question quickly and responds by waving her hands. This resolution of ambiguity emerges as early as when we learn a common language as an infant, who is essentially a tourist fresh off the boat. *Robots' failures*, on the other hand, depicts a situation where sharing existing coding-decoding rules is not a sufficient condition for mutual understanding. Consider you are interacting with an artificial agent such as Apple's Siri, "Hey Siri, I would like to go fishing, so where is the nearest bank?" Siri would recommend banks that offer financial services based on the statistical regularities rather than river banks in line with the current communicative context. Artificial agents have long built on the same guiding principles underlying signal-centered frameworks, namely that signals have stereotyped and fully predictable consequences (*fixed-code fallacy*, see Stolk et al., 2020). This approach neglects the interactive construction of context by virtue of seeking and confirming the evidence of mutual understanding. Moreover, Bayesian methods that rely on the prior cannot account for the fast catch of the interpretation of a signal without a clearly defined prior. And the

deficiency in artificial agents cannot be ameliorated by more advanced engineering techniques without shifting away from Shannon-like signal-centered frameworks (Stolk et al., 2016).

Conceptual alignment framework. Mutual understanding, according to the conceptual alignment framework, is achieved by interaction members continuously interweaving and converging on shared context. The neglected process of interactive context construction in signal-centered accounts, or *joint epistemic engineering* (Stolk et al., 2020), refers to a process that members of the interaction negotiate and create a dynamic shared conceptual space allowing for rapid resolution of ambiguities of multi-layered communicative signals (Stolk et al., 2016; Toni & Stolk, 2019). Construction of the shared conceptual space kicks in even before the conversation begins, as when communicators need to conceive of a context that allows their opening signal to be better understood than otherwise by a specific addressee (Stolk, Noordzij, Volman, et al., 2014), e.g., information on where the conversation takes place, assumption about the others' knowledge; and keeps lively updated throughout (Stolk, Noordzij, Verhagen, et al., 2014). Here, the meaning of a word does not reside in or stably attached to the signal, but assimilates into the mutually inferred space—to create, delete or modify the elements; where the signal is merely the behavioral vehicle for probing and shaping that fleeting conceptual space defined by the ongoing interaction. By virtue of the interface for references, interlocutors' inferred thoughts can be aligned to the current context and one another as well, enabling their quick resolution of ambiguities, converge on the same page and even invention of new temporary usage, leading to the final mutual understanding. The interactive and dynamic conceptual alignment account built on Marr's computational view better explains smooth and effortless communication despite the ambiguous nature of meanings—the true vitality, complexity, and distinctiveness of human communication.

Neural evidence from a magnetoencephalography (MEG) study (Stolk et al., 2013) provided initial neuroscientific evidence for the conceptual alignment framework, where the meaning of a signal is hatched by a fleeting conceptual space that is determined by the ongoing interaction. Three findings underpin the investigation. First, the observed neural activities were in the brain regions necessary for processing conceptual knowledge—right temporal lobe (TL) and ventromedial prefrontal cortex (vmPFC). Second, communicative comprehension and production share the same patterns of neural activity, which directs to the same conversational context. These findings indicate that the observed neural activities were related to task features instead of different sensorimotor events. Third, the timing of the shared patterns preceded any communicative signal generated and interpreted as it is defined by the interaction instead of the signal itself. An ensuing dual-fMRI study (Stolk, Noordzij, Verhagen, et al., 2014) recorded two communicators' brain activities simultaneously and found that the superior temporal gyrus (STG) became increasingly involved over time, and the synchronized inter-cerebral dynamics between two communicators was independent of signal occurrences, indicating the communicators' adjustments of their shared conceptual space. These four findings empirically prove that a fleeting conceptual space is jointly established, continuously updated, and overall defined by the dynamic interaction, in favor of the conceptual alignment framework.

In sum, the theoretical and empirical considerations above suggest that human communication is better interpreted as a conceptual alignment challenge than a signal transmission problem. Here I consider the possibility to extend the conceptual alignment framework into the domain of autism—a disorder that is typified by communicative challenges during social interaction.

Communicative challenges in Autism Spectrum Disorder

Autism Spectrum Disorder (ASD) is diagnosed based on communicative deficits in daily social interaction, and the core symptoms consist of poor social interaction quality as well as repetitive and restricted patterns of behavior (American Psychiatric Association, 2013). The deficits have been attributed to a core impairment in social cognition—representing mental states, also known as "theory of mind" or "mentalizing" (Baron-Cohen, Leslie, & Frith, 1985; Happé, 1993) and are most evident when intended and literal meanings diverge—irony or sarcasm (Zalla et al., 2014) or when it requires going beyond explicit statement—metaphor (Kalandadze, Norbury, Nærland, et al., 2016). An extensive amount of empirical evidence supports an alteration in theory of mind; concomitantly, quite a few fail to prove the universal impairment of mentalizing in ASD or even replicate seminal theory-of-mind findings (for a review, see Gernsbacher & Yergeau, 2019). Besides the classic atypical social cognition explanation, other main accounts include:

1. *Social motivation and attention*. The social motivation theory (Chevallier, Kohls, Troiani, et al., 2012) argues that motivation deficits can exert downstream effects on the development of social cognition, with the rationale that diminished social motivation and/or attention regardless of diagnosis starting in early childhood would hinder the child from receiving crucial social inputs, thus atypical social cognition is a consequence rather than a cause of disrupted social interest. A meta-analysis of fMRI evidence (Clements et al., 2018) found domain-general reward system dysfunction in ASD in favor of the social motivation hypothesis. Another meta-analysis of eye-tracking studies indicated that individuals with ASD have overall reduced social attention compared to those with typical development (TD) (Chita-Tegmark, 2016).

2. *Executive function*. Communicative challenges, especially pragmatic deficits in ASD are shown to arise from a deficit in executive function (for a meta-analysis, see Demetriou, Lampit, Quintana, et al., 2018). The atypical executive dysfunction hypothesis (Hill, 2004; Pennington & Ozonoff, 1996) includes difficulties in set-shifting, response inhibition, and working memory. Recent empirical evidence suggests that autistic use of referring expressions can be impacted by cognitive flexibility and working memory simultaneously (Malkin & Abbot-Smith, 2021).

3. *Motor function*. The motor deficits account is based on the notion that early motor abilities can set the stage for the development of social interaction and communicative skills (Bhat et al., 2012; Mody, Shui, Nowinski, et al., 2017). Since motor dysfunction is an early emerging component of the ASD phenotype (Thomas, Davis, Karmiloff-Smith, et al., 2016), vulnerability in motor function early in development (LeBarton & Landa, 2019) can lead to atypical language and social behaviors.

4. *Communication cues*. Autistic individuals possess intact socially relevant action perception, indicating they are perceptually capable of distinguishing between subtle nonverbal cues, though they often fail to exploit them during real-life social interaction (Cusack et al., 2015, but see Bolis & Schilbach, 2018 for argument of an imparied action prediction). Yet, impaired genetically-driven social visual engagement—the active seeking of social information in autistic children (Constantino, Kennon-McGill, Weichselbaum, et al., 2017) and atypical integration of verbal and non-verbal communication cues (Jung, Tu, Lang, et al., 2019) can account for communicative challenges in ASD.

5. *Perceptual function*. The atypical perceptual function is not part of the diagnostic criteria yet and neither the nature nor the connection with core communicative symptoms is clear. Classic theories suggest that autistic individuals exhibit impaired global processing (Booth & Happé, 2018), enhanced local processing (Mottron et al.,

2006), and a preference for detail (Koldewyn et al., 2013). Recent perceptual inference accounts emphasize autistic individuals' attenuated priors in predictive processing (for a review, see Stark, Stacey, Mandy, et al., 2021). Empirical evidence (Chambon, Farrer, Pacherie, et al., 2017) indicated an atypical interplay between top-down priors and bottom-up sensory processing when predicting and understanding others' actions in autism, which can be an explanation for autistic challenges in inferring mental states. However, recent evidence (Feigin, Shalom-Sperber, Zachor, et al., 2021) showed that autistic individuals showed an increased influence of prior choices instead of prior stimuli, thus challenging the "hypo-priors" view.

Other novel accounts include a non-deficient-based perspective (Casartelli, Federici, Fumagalli, et al., 2020; Davis & Crompton, 2021), arguing that the mismatch of interaction styles between autistic and neurotypical interlocutors lead to bidirectional communicative challenges. These multifaceted accounts and mixed empirical results highlight the current situation that a systematic and deep understanding of autistic communicative challenges is still lacking, either from a perspective of cause or intervention. Yet, understanding communicative challenges in autism may ultimately uncover new targets for interventions aimed at improving autistic individuals' experiences within social interaction. Prior to introducing the methodological framework developed as part of this thesis project, here I propose a novel neurocognitive account for autistic communicative deficits by extending considerations from the conceptual alignment framework to the domain of autism.

Extending conceptual alignment as a novel neurocognitive account. As theoretically and empirically suggested by the conceptual alignment framework (Stolk et al., 2016), neurotypical individuals rely on a jointly-built conceptual space to resolve inherently ambiguous communicative symbols. We hypothesize that autistic individuals' communicative challenges can possibly be explained within this conceptual alignment framework—their difficulties creating, utilizing, and/or updating the fleeting conceptual space result in their inability to resolve the pervasive ambiguities in communicative signals and vulnerability in everyday interactions. Wadge et al. (2019) explored the possibility initially by embedding autistic individuals in an interactive and dynamic social setting (Stolk, Verhagen, et al., 2013) that precludes the use of pre-existing shared representations (e.g., linguistic and gestural emblems, facial expressions). The behavioral study suggested that autistic individuals' communicative challenges cannot be simply traced to altered sensory processing, motor performance, interaction memory, social motivation or attention, or cognitive perseveration, which exclude most aforementioned main accounts for communicative challenges in ASD. Rather, autistic individuals struggled to align their conceptualizations of communicative behaviors with their conversation partner when resolving the ambiguity of those behaviors required referencing their recent communicative history. However, the *altered perceptual function* account has not been disentangled from the intended interpretation in this behavioral analysis, thus the cognitive sources of the communicative challenges in ASD have not been fully elucidated. It remains unknown whether the observed autistic deficiency originates from a difficulty with the top-down process of utilizing the shared conceptual space or merely an altered bottom-up perceptual function.

Zooming into the *altered perceptual function* account, previous studies framed in the context of predictive processing are based on the theory that reciprocal top-down and bottom-up message implements *hierarchical probabilistic inference* on the source of sensory stimulation (Clark, 2013; Friston 2005). The ramification swept across a range of hypotheses about various brain functions, including those implicated in autism, e.g., perceptual or sensorimotor, social and cognitive or learning domain (for a review, see Palmer, Lawson, & Hohwy, 2017). Recent models of Bayesian inference incorporate *estimates of environmental volatility*, e.g., the change of causes

of sensory input over time (Lawson Mathys & Rees, 2017; Mathys et al., 2014) with the rationale that the finer mechanism underlying context-dependent adjustment of sensory weightings can be a key to understanding social-cognitive processes in autism where inferring others' mental states is required (Palmer, Seth, et al., 2015). Empirical results suggest that autistic individuals struggle to predict others' behaviors due to higher-weighted incoming sensory signals and attenuated prior expectations; Pellicano & Burr (2012) suggest a *chronically reduced precision of prior beliefs* underlying autistic sensory symptoms; similarly, Van de Cruys et al. (2014) argue that an *increased and inflexible weighting of prediction errors* underpin social or nonsocial autistic traits. In contrast, one study (Tewolde et al., 2018) found no difference between neurotypical and autistic children regarding anticipating changes in dynamic objects, suggesting no pervasively different predictive mechanism in autism; autistic individuals do not show attenuated prediction compared with neurotypical individuals either during processing linguistic constructs (Brennan et al., 2019; Zhou et al., 2019); recent evidence further challenges the "hypo-priors" view by separating the influence of prior choices and prior stimuli (Feigin, Shalom-Sperber, Zachor, et al., 2021). To complicate the picture further, even the nature of predictive processing is still under debate, e.g., Friston et al. (2013) argue that the communicative deficits are not due to a failure of prediction but a failure of metacognition to instantiate top-down predictions during perceptual synthesis. The mixed results thus require a closer examination of perceptual inferences in autism and a clearer disentanglement between putative perceptual effects and intended interpretation on communicative challenges in autism from a perspective of conceptual alignment.

It should be noted that the top-down process in this project refers to the downstream effect of the ability to utilize the shared conceptual space on communicative (mis)alignment while the bottom-up process represents the effect of altered perceptual function including both priors and likelihood precision on the communicative challenges. A previous study (Wadge et al., 2019) explored the interpretive compatibility of the conceptual alignment framework for autistic communicative challenges but has not excluded the possibility of altered perceptual function. The cognitive source of observed deficient conceptual alignment in autism can come from bottom-up perceptual processing—e.g., atypical interaction between priors and likelihood precision, instead of intended top-down processing suggested by the conceptual alignment framework. Therefore, aiming to test the reliability of the proposed conceptual alignment account underpinning the autistic communicative deficits, the bottom-up and top-down contributions are assessed through the use of dual eye trackers in the current study by embedding pairs of participants with ASD or TD in a live setting that affords genuinely interactive real-time communicative behaviors.

Methodological framework

Social interaction pervades all aspects and all stages of our daily life, from learning babbling from birth to sharing ideas and experiences with others later in our life. Paradoxically, most previous studies aimed at understanding human communication appear to minimize demands on achieving mutual understanding through the medium of scripted utterances and isolated sentences in a non-interactive lab environment (Schilbach, Timmermans, Reddy, et al., 2013; Wheatley et al., 2019), limiting the inferences that can be drawn about the mechanisms underpinning human communication. Empirical evidence further shows that neural and cognitive processes differ between a second-person context—truly interactive setting, and a third-person setup—traditional observation of social

stimuli (Redcay & Schilbach, 2019), necessitating a shift from social observation to a dual-brain approach. A multi-person interactive setting that adapts a simultaneous dual-brain approach can open the possibility to investigate how interlocutors reciprocally negotiate and cooperate as well as adapt and align their beliefs to each other; furthermore, it can shed light on social impairment as when behavioral performance patterns and/or underlying neural mechanisms diverge between individuals with social disorder and typical development (Wadge et al., 2019; Quiñones-Camacho, Fishburn, Belardi, et al., 2021). Here, by embedding pairs of participants in an experimentally controlled and computer programmed novel platform, pairs of autistic individuals' emergent dynamics in real time are measured by dual eye trackers to examine the cognitive sources of communicative (mis)alignment.

Measuring live communication in the lab. Capturing emergent dynamics during human communication in the lab poses a challenge to the establishment of an experimentally effective communicative platform, which is supposed to preserve an intended key component in natural conversation—dynamic conversational context while controlling and/or excluding the other irrelevant components simultaneously, e.g., communication channel, pre-existing conventions, spontaneous turn-taking, etc. A recent effective approach (Selten & Warglien, 2007; Stolk et al., 2013) involves participants in the *Tacit Communication Game* (TCG, Figure 1A), a platform that is computer programmed and experimentally well-controlled. Compared to natural dialogue, TCG serves as a novel medium that (1) adapts novel communicative signals that minimize participants' access to pre-existing conventions commonly used in daily communication (e.g., a common idiom, body emblems, facial expressions), thus providing control over communicators' shared cognitive history; (2) contains only one single communication channel—moving geometric shapes on the screen, thus providing control over communicators' interactive environment; (3) assigns experimentally-controlled roles, *communicator* or *addressee*, to each of the two players alternatively, allowing the assessment of shared and unique processes across production and comprehension. However, as TCG preserves the intended key component that is shared with natural dialogue—dynamic conversational context, it enables the investigation into shared conceptual space where fleeting idiosyncrasies of an ongoing interaction are mutually negotiated and defined.

The game allows manipulating access to top-down priors by virtue of so-called *Known* and *Novel* trial types. In previously trained known trials, representations of specific goals or strategies have been established as the *prior*. In contrast, for novel trials that haven't been presented before, two players are supposed to invent and converge on idiosyncratic strategies to solve the communication problem together without any reference to pre-existing conventions, thus the emergent alignment of newly-created strategies cannot follow from a Bayesian inference process but rather from a top-down process that reflects a fleeting conceptual alignment. Therefore, by contrasting conceptual alignment between known and novel trials as well as groups of participants with ASD and TD, the contribution of *altered perceptual function* account and *conceptual alignment* account can be quantified and disentangled.

Figure 1. Live eye tracking in the two-player Tacit communication Game (TCG). **A.** Sequence of events during an example communicative interaction. The joint goal of the players is to reproduce a target configuration of their two assigned geometric shapes) shown to one of the players only, namely the "*communicator*" of that interaction (phase 1). This requires the communicator to move his shape in a way that conveys to the other player, the "*addressee*", the target location and orientation of her shape (phase 2). The communicator has 10 seconds to

achieve this by virtue of otherwise unrestricted translations and rotations of his shape on the game board, while also ensuring his final position matches the position indicated by the target configuration. For example, he might visit the addressee's target location (number 1 action), step out in the direction of her shape's target orientation (2), and proceed to his own target position (3). Given that the addressee cannot see the target configuration, she needs to infer her target position and orientation from the communicator's movements. Note that the shown solution is only one among several possible solutions, and that rapidly extracting the gist of the shape movements, like words in everyday interaction, might rely on shared dynamic priors. After the addressee has moved her shape into her inferred target position (phase 3), feedback is presented to both players on their joint success (phase 4). TCG was programmed and delivered via MATLAB software. Adapted from Stolk et al. 2016. **B.** Eye scanpath during an example communicative interaction (pair 1, trial 33). It can be seen from the scanpath in the top panel that the communicator's fixation leads both his own movements (phase 2) and the movements made by the addressee (phase 3). By contrast, as shown in the bottom panel, the addressee's scanpath trails the communicator's movements (phase 2). Fixation locations are indicated with circles containing a plus sign, with circle size corresponding to the duration of the fixation. Saccadic movements are indicated with colored traces, with trace color corresponding to the temporal order of the saccade. The numbers outside the gameboard correspond to the phases of the game in panel A.

Using eye tracking to dissociate top-down and bottom-up processes in communication. Eye-mind link (Just & Carpenter, 1980; Rayner & Reingold, 2015) indicates that the structure of the eye determines its high acuity vision limited to a small portion of the visual field—fovea, leading to our tendency to direct the fovea to the currently processed stimulus. Thus, the measurement of pupil dilation and gaze position is a noninvasive indicator of spontaneous responses concurrent with stimulus presentation without requirement of overt responses (Tamietto et al., 2009). Furthermore, since eye movements are controlled by an extensive and distributed system (Chita- Tegmark, 2016; Wang et al., 2015), they can be perturbed by altered neural processing (Carter & Luke, 2018). Pupillometry studies, measuring pupil size and reactivity, is based on the theory that autonomic nervous system alterations was suggested to be present in ASD (Hellmer & Nyström, 2017) and predictive of communicative deficits (Porges, 2005; for a meta-analysis, see Vries, Fouquaet, Boets, et al., 2020). In parallel, gaze direction as a rapid and automatic process triggers reflexive shifts in visual attention (Watanabe, Miki, & Kakigi, 2002), thus atypical gaze direction patterns in autism can affect capturing of attention and place a barrier in childhood development (Jones et al., 2008). For example, children with TD prefer to look faces in the eyes while autistic children preferentially orient their eye gaze on surroundings that are less relevant to social interaction (Zwaigenbaum et al., 2005). Through the use of eye trackers, the *altered perceptual function* account for autistic communicative challenges can be directly examined by comparing *addressee's* gaze movement spatiotemporal patterns during *pre-target movement* intervals, where contextual effects of top-down priors are expected to remain constant.. In contrast, the *conceptual alignment* account can be tested through the comparison of gaze patterns during *post-target movement* intervals, where top-down priors are likely to alter the course of gaze behavior through context-mediated identification of communicatively relevant portions in movement. Specifically, we predict that the *Novel* and *Known* conditions differentially affect gaze behavior during post-target movement intervals but not during pre-target intervals.

Previous studies involving analysis on gaze position usually adopt indicators such as area of first fixation (Schneider et al. 2013), fixation duration to the areas of interest (AOIs) (Amso et al., 2014), etc. Yet, these indicators are not compatible with current real-time TCG where AOIs—geometric shapes occupy relatively small areas and are moving around all the time. More importantly, they are unable to capture the emergence of mutual understanding during dynamically unfolding communicative interaction. Here, I explore the possibility that eye fixations can shed light on the cognitive processes underlying emergent dynamics of communicative behaviors by drawing inferences from their continuous variation of spatiotemporal patterns. The attempt can break fresh

ground in preliminarily qualifying and quantifying emergent mutual understanding using eye-tracking methods as well as diving into the distinguished patterns that autistic individuals present, thus enabling tracing the neurocognitive sources of communicative challenges in autism. Besides, a preprocessing pipeline was established based on MATLAB toolbox FieldTrip which enables removing artifacts and outliers (e.g., blink and eyelid signal, drift, linear trend, etc.) as well as converting raw eye tracking data into more readily interpretable fixation time series.

Current project

The current project is mainly dedicated to open up the possibility to investigate neurocognitive sources of autistic communicative challenges both technically and theoretically. Technically speaking:

(1) An eye-tracking preprocessing pipeline was built for cleaning both pupil size and gaze position data in terms of various artifacts and/or outliers, e.g., blink, drift, linear trend, etc. The pipeline is based on FieldTrip, an open-source MATLAB toolbox dedicated to the analysis of electrophysiological time series data.

(2) Preliminary qualitative and quantitative methods were built and tested for gaze position datasets while one promising advanced quantitative method was proposed.

(3) TCG platform was constructed and validated to allow for capturing real-time dynamically unfolding communicative interaction from simultaneous recording of both dual eye gaze and dual-EEG data.

Theoretically, we closely examine and disentangle whether autistic communicative challenges originate from their difficulty aligning meaning with others when referencing recent communicative history is required (top-down process, *conceptual alignment* account) or merely an altered perception function (bottom-up process, *altered perceptual function* account).

Eye-tracking Methods

Datasets

The current eye tracking dataset consists of recordings from 22 TD and 18 ASD individuals, obtained from an ongoing project combining dual-fMRI and eye tracking. Eye tracking data was collected simultaneously using an MRI-compatible Eyelink 1000 plus equipment (SR-research) from the left eye with a sampling frequency of 1000 Hz. The physical width and height of the MRI-compatible screen are 698.4 mm and 392.9 mm and the eye-screen distance is 1340 mm since participants viewed the screen via an IR-reflective mirror in the MRI scanner. A five-point calibration and validation was performed once before the start of TCG (see arguments below that multiple calibration during the TCG is not preferred). Luminance was matched throughout TCG. Pupil size was recorded in terms of area with the unit being pixels and recorded resolution being 0.1% of the diameter. Eye positions were recorded in terms of gaze positions, which are actual positions on the display screen (1920*1080 pixels) and fully compensated for head position and distance from the display. The original .EDF format data was converted to .asc files using a EDF2ASC software provided by SR-research company for further processing in MATLAB.

Preprocessing

Raw data recorded by eye trackers contains plenty of undesirable features, especially in the current combined fMRI-eye tracking study. The special difficulties of MRI-dependent eye tracking acquisition arise since the preparation of scanning—participant positioning within the head coil for the optimal camera visibility. The positioning is more complicated compared to a normal fMRI study as the optimal configuration is only accessible while participants are pushed into the scanner, where lighting and camera visibility of the eye can then be confirmed. Beside, a compromise between screen visibility for the participant and eye visibility for the camera should always be determined. The situation can be more complicated if the participants are using MR-compatible glasses, as the reflection of IR spotlight on the lens surface should be avoided by carefully adjusting the angle of glasses. The participants may need to slightly move the head in the head coil during preparation to satisfy the quality of eye tracking while the in-scanner adjustment is however undesired in terms of the pursuit of immobility for MR image quality. Thus the initial positioning can take a longer time than usual MRI preparation and sometimes eye tracking quality is sacrificed for optimal MR image quality. Besides, the difficulty also presents in the form of multiple calibration failures especially when the initial positioning is suboptimal. Reflected in the data quality, the suboptimal positioning or calibration can introduce more tracklose samples, reduced spatial precision and inaccurate automated detection of eye movement events, as the latter is highly-dependent on eye movement velocity, thus artificial velocity change resulted from positioning can disturb the detection (Hanke, Mathôt, Ort et al., 2019).

Based on aforementioned difficulties in acquiring eye tracking data with optimal quality and the obtained suboptimal eye tracking data compared with MRI-independent acquisition, *preprocessing* is highly required to compensate and to extract intended components buried by concurrent noise or artifacts for further analysis. Yet the definition of artifacts is highly dependent on the context—what constitutes undesirable features in one paradigm can be the signal of interest in another study. For example, eye blinks are usually considered detrimental to recorded signal since pupil dilation or gaze position can not be measured during a blink, and blinks are often followed by a prolonged pupil constriction (Knapen et al., 2016), but blink rate or duration can also be an indicator of fatigue that is of interest in another context (Stern, Boyer, & Schroeder, 1994; Marandi et al., 2018). In other words, preprocessing steps aimed at eliminating effects of artifacts and outliers can vary contingent upon experimental paradigms.

In contrast with other eye-tracking studies that usually perform calibration at the beginning of every block, such as static picture viewing (Wang et al., 2015), the TCG is dedicated to capturing continuously unfolding communicative interaction throughout, thus interrupting the task when substantial drift is detected is not an option. Yet the pursuit of an intact interactive process brought up an apparent drift and a linear trend artifact with the degrading of the eye tracker calibration over time, possibly due to participants' fatigue or subtle head movements, which unfortunately cannot be controlled perfectly in practice. Reflected in the density of positions (Figure 2B top row, heatmap of original v.s. filtered data and detrended data), the drift blurred the original edge of the TCG game board, presenting a cloud-like and less-sharp structure compared with the filtered data excluding the drifts; the linear trend diverts the whole TCG game board towards the left bottom. Quantitatively, the 10s-moving average of x eye position decreases over time while the 10s-moving standard deviation (*SD*) of both x and y eye

positions increase with time (Figure 2B bottom row, original data), indicating the degradation of the eye tracker calibration and gradual drift of eye positions. Therefore, a high-pass filtering method is required for the removal of the gradual change of positions over time and the detrending method for the linear trend. Besides, other components such as blink and eyelid signal, spurious pupil size, positions with abnormal velocity are all considered as artifacts in TCG, which, along with outlier positions, will be excluded by the established preprocessing pipeline below.

Current commercial eye tracker companies provide proprietary software for preprocessing eye data (e.g., Tobii Pro Lab by Tobii, Data Viewer by SR-Research, etc.), yet the tools include only basic processing of data neglecting some functions crucial for our paradigm, e.g., blinks and eyelid signal correction or a time-course visualization. Other guidelines (Kret & Sjak-Shie, 2018; Mathôt et al., 2018) or open-source toolbox (e.g., CHAP based on MATLAB for pupillometry data, see Hershman et al., 2019; ETRAN based on R for eye movement data, see Zhegallo & Marmalyuk, 2015; GazeParser based on PsychoPy, see Sogo, 2013, etc.) do not offer tools that allow correcting some types of artifacts such as drifts or linear trends in TCG, which are virtually unavoidable during unconstrained communicative interactions. In the remainder of the thesis, a preprocessing pipeline (Figure 2A) will be introduced, which is specially developed for the data at hand—yet it can also be applied to other datasets troubled by similar artifacts. It builds on functionalities contained in the open-source FieldTrip toolbox (Oostenveld et al., 2011), which was designed for the analysis of brain time-series data (e.g., electro-/magnetoencephalographic signals), and constitutes a platform that can be readily extended to deal with pupil and eye gaze time-series. Further analysis will also be carried out in the platform, thus it can smooth the transition from data preprocessing to analysis and visualization. The pipeline entails the aforementioned artifacts and outliers rejection and four additional optional modules that allow offline position recalibration, fixation extraction, heatmap visualization, and time-course visualization, respectively. Note that the construction of subject m-files that contain subject-specific details (Figure 2A, pink area with solid frame) is required initially for batch processing which is convenient to systematically examine the outcome of analysis (for details, see Oostenveld et al., 2011). The below demonstration of the effect of preprocessing on data cleaning is based on one actual neurotypical dataset with suboptimal quality.

Figure 2. Preprocessing pipeline for eye-tracking data. **A.** Preprocessing flowchart. Dotted frames denote one-time steps and dark orange blocks represent interactive cleaning sessions storing reusable parameters in subject-specific matlab annotation files. Blue squares and green circles indicate the type of information used in cleaning, being either pupil size or gaze positions. **B.** Depiction of effects of each preprocessing step on the data (one neurotypical participant). The top row is the overall heatmap—density of positions in the whole TCG. From left to right: heatmap of original, de-blinked, filtered, detrended, and fully preprocessed data. The superimposed 3-by-3 grid represents the TCG game board. The color bar indicates fixation count. Note the origin of all heatmaps shown in this thesis is in the left top corner, consistent with the convention of the Eyelink acquisition system.

Blink detection and removal. The most stringent solution to blink artifacts is to remove trials with blinks (Weiskrantz, Cowey, & Barbur, 1999), yet the relatively long trial (~15s) in TCG determines that the method would cause high data loss and even leave no trials left. Another commonly used method (Sirois & Jackson, 2012) assumes that the appearance of a blink is immediate, thus regarding the first missing value as the moment the blink starts and the recurrence of value as the end of the blink. However, as eyelid movement is not instantaneous, blinks are supposed to be characterized by a sharp decrement in pupil size, followed by an interval of missing data, then another rapid increase in pupil size (Caffier, Erdmann, & Ullsperger, 2003). In terms of gaze positions,

eyelid artifacts are also detrimental to capturing true position variations (Figure 3A, blink detection). The method that only considers missing samples as blink leaves the eyelid artifacts unremoved, leading to artificial differences in pupil size and gaze positions between conditions. Therefore, the preceding and following eyelid distortions to blinks are supposed to be detected and manipulated concurrently with missing values. Other algorithms that take eyelid signal into account either use a fixed wider window (Satterthwaite et al., 2007) or a fixed cutoff based on the velocity of change in pupil size (Mathôt, 2013) with the assumption that blinks are presented in similar patterns, which is however not true even within the same dataset (Caffier et al., 2003). Here, a more accurate pupillometry noise-based algorithm (Hershman et al., 2018) is adapted to include eyelid artifacts. The algorithm is based on the idea that the pupillometry signal recorded by the eye tracker is small-amplitude and high-frequency (above 20 Hz) fluctuations, resulting from measurement error inherent in the image-processing algorithms of eye tracking device instead of the physiological process since the latter, on the contrary, is characterized by relatively slow frequencies (Nowak, Hachoł, & Kasprzak, 2008). This kind of measurement error is eliminated when countering strong signals—the rapid movement of the eyelid during shutting and reopening, thus providing an opportunity to accurately detect the eyelid signal through the start and end of the measurement noise. In this project, the blink and eyelid signal detection is based on pupil size data and then applies the on/offset to position data, visualized by a data browser function contained in the FieldTrip toolbox (Figure 3A, left panel, pink areas). The manual selection function can be enabled in the data browser which would allow us to manually determine the neglected blinks or correct wrongly chosen ones. Yet, the manual selection wasn't applied to this dataset because of the already near-perfect accuracy.

After obtaining the relatively accurate on/offset of the blink and eyelid signal, the artifacts can either be removed—marked as *nan* (not a number) or interpolated using a linear or cubic-spline method (Mathôt, 2013). Comparing blink removal and blink interpolation method (Figure 3, heatmap of data after blink removal v.s. data after blink interpolation), it turns out that interpolation method would create another undesirable artifact "freeze"—highlighted curves or straight lines in the heatmap, especially when a large amount of data is interpolated. Besides, blink removal won't affect the results as long as missing data is not denoted as real 0 pupil size or gaze positions measurement. Therefore, blink artifacts or other missing data in the dataset received no special treatment but were marked as *nan*, which can be ignored in further analysis based on MATLAB. Yet the interpolation method (e.g., cubic-spline method) can still be useful for other studies. In this case, the choice between linear (Cohen et al., 2015) versus cubic interpolation (Mathôt, 2018) can have negligible effect considering the short duration of the blinks. Yet note that smoothing is usually performed before interpolation if a cubic-spline method is chosen (Geller et al., 2020).

Figure 3. Blink and related eyelid artifact detection and resolution. **A.** Visualization of original data with blink and eyelid signal (left), data with blink and eyelid artifact removal (middle), and data with blink and eyelid artifact interpolated (right) via an interactive databrowser. **B.** Heatmap of gaze positions contrast between original data and manipulated data—with blink and eyelid artifact removed or interpolated. Highlighted straight lines or curves can be observed in the heatmap if missing samples originating from blinks were interpolated.

Drift and linear trend removal. As explained above, our pursuit of continuous capture of unfolding communicative interaction throughout TCG restrained calibration during the task, which brought up apparent

artifacts in terms of gaze positions—drift and linear trend. These artifacts are progressive displacement of fixation registrations over time, possibly caused by participants' subtle movement of head or pupil dilation, which can not be avoided or controlled perfectly, even in a laboratory setting. As a result, position data is getting less accurate as the experiment proceeds. This effect can be exacerbated in the context of TCG, where the stimuli are not static and no calibration was performed between trials or blocks. As further analysis entails a closer examination of synchronization of gaze positions and stimulus footprints, a relatively accurate mapping of gaze positions to the absolute location of TCG game board is required—though not necessarily to be perfect because the following correlation analysis tolerates the systematic deviation (i.e., overall deviation to one direction) to a certain extent. Therefore, a band-pass filter between 0.01 Hz and 2 Hz is applied to the dataset initially to get rid of drifts that blur the edge of the game board in the heatmap (Figure 2B, original v.s. filtered data). Then detrending is adapted to further remove the linear trend over time and drag the center of gaze positions to the absolute coordinates of the center of the TCG game board in the heatmap (Figure 2B, filtered v.s. detrended data), thus achieving a relatively accurate mapping of systematically distorted gaze positions to the stimuli positions.

An additional optional module in the preprocessing pipeline is provided to perform offline recalibration of positions through best-fitting linear transformation (Vadillo et al., 2015). This approach requires the knowledge of ground truth in specific experiments that participants tend to look at one position at a certain time point since the offline recalibration method is based on a transformation matrix calculated by coordinates of the presented stimulus and at least 5-10 fixations that are likely to be directed to the stimulus. Compared with previous offline recalibration methods, e.g., calculating the disparity between fixation and the closest stimulus on the screen (Zhang & Hornof, 2011), this linear transformation approach has the flexibility to stretch or contract the fixation space and enables multiple trial-based recalibrations. Compared with above-mentioned filtering or detrending methods, the linear transformation approach is able to correct intra-trial drifts in a short term instead of slow drifts over time. Yet the additional recalibration method wasn't practically applied to the current TCG dataset as the filtering and detrending have already achieved good enough results (Figure 2B, detrended column), and additional recalibration based on participants' probable intended positions can bring undesired artificial difference or offset interesting difference of characteristics between individuals with ASD and TD. However, it can serve as a potential option for other studies troubled by the intra-trial drift problem.

Interactive artifacts definition and removal. Detection of squint and jump artifacts and position outliers is customized to each dataset through the medium of an interactive histogram—squint and jump artifacts, or an interactive heatmap—position outliers, that allows dragging on the plot to exclude spurious pupil size, positions with abnormal velocity and/or positions outside the TCG game board (Figure 4). The subject-dependent criterion is based on the idea that pupil sizes and gaze positions are inherently heterogeneous across experiments or individuals. Previous commonly used method includes converting data to *z* score based on mean and SD calculated for each trial or participant separately (Cohen et al., 2015; Wainstein et al., 2017). Then z score values above and/or below a user-defined threshold will be removed. The current subject-dependent method based on the distribution avoids the pre-defined and fixed criterion across different datasets or mandatory exclusion of partial data even it's good enough, thus reserving the flexibility to determine the cutoff contingent on each participant's idiosyncratic standard of normal range of pupil data (Mathôt et al., 2018). For example, based on the histogram of pupil area below (Figure 2A, left panel), it's reasonable to remove pupil area less than 800 and greater than 2700. Note that the typical range of pupil area is 800 to 2000 units, thus the current distribution is possibly caused

by a recording distortion. But since the pupillometry data requires further baseline correction or z-score conversion, the relative changes matter more than absolute values. All removed data is marked as *nan* consistent with the manipulation of blink and eyelid artifacts.

Squints are defined as spurious pupil sizes that are not realistic, caused by not fully open or closed eyelid or temporarily distorted recording (Figure 4A, left panel). The size of the human pupil ranges from around 2 to 8 mm while the intra range variances are mostly caused by the luminance of the environment and distance to fixation—pupil dilates in response to decreased brightness and increased distance (for a review, see Mathôt, 2018). The third account that is of interest exists, namely the psychological process that "activates the mind" (Goldwater, 1972) but with a far smaller effect on pupil size compared to the former two accounts. Therefore, to capture the subtle effect of the desired psychological process out of uninformative noise, controlling over two former accounts was performed during data acquisition. Additionally, removal of artificial pupil size is also required to avoid contaminating the potential capturing of the subtle psychological effect. The squint problem is less detectable in the dataset than blinks which are at least marked as missing data by eye trackers, yet through the current medium of an interactive histogram, those pupil sizes that appear much less frequently can be regarded as artificial and removed from further analysis (Figure 4B, left panel). Besides, in pupillometry studies, the unrealistic change of pupil size is supposed to be considered as well. The size of the pupil changes at most by a factor of about 4 or around 16 in terms of pupil surface (McDougal & Gamlin, 2008), indicating any change above the range could result from a distorted recording. As we did not include the pupillometry aspect in the current analysis, the function of removing abnormal velocity of change of pupil size is to be developed in the future versions of the preprocessing pipeline.

Jumps are defined as gaze positions with abnormally high velocity (Figure 4A, middle panel), possibly caused by sudden stimulus-irrelevant saccades during participants' distracted moments, and calculated by the square root of the sum of squares of x and y positions minus those at one previous time point. The exclusion of jumps prepares a relatively clean dataset containing informative gaze position data that is more probable to be relevant to the stimulus presentation. The dragging method is the same as squint manipulation and the distribution of an example jump-excluded datasets were presented in the middle panel in Figure 4B.

Outlier gaze positions with coordinates that fall outside the actual size of the screen—1920*1080 pixels in this study, can be removed as outliers by enabling the function of a bounding box (Figure 4 right panel, red frame). Further definition and exclusion of user-defined area, e.g., TCG game board consisting of gaze positions in the heatmap, can be achieved by dragging the edge of the area of interest, i.e., dotted orange straight line, to the edge of the screen with the direction pointed by the dotted orange arrows.

Figure 4. Interactive definition and removal of artifacts and outliers. **A.** Artifacts and outliers definition via a graphic user interface. Original distribution of pupil area (squint definition), position velocity (jump definition) and heatmap of original position (outlier definition) is displayed sequentially, allowing users to drag the plot, i.e., start from the dotted orange line towards the direction of the arrow to exclude spurious pupil size, position with abnormal velocity and positions outside screen or TCG game board. The unit of pupil area is pixels and the typical range is 800-2000 units, though this distribution may differ per subject. The red frame in the heatmap represents the actual size of the screen frame (1920*1080 pixels). **B.** Distribution and heatmap after removal of artifacts and outliers. The white area within the read frame indicates no data point exists.

Optional modules. Besides the above-discussed *position recalibration* function provided additionally by the preprocessing pipeline, *time-course visualization* (e.g., Figure 3A) and *heatmap visualization* (e.g., Figure 3B) can also be utilized to achieve a better overview of the dataset or examination of the data quality. The rest *fixation extraction* module directly imports eye movement definition from the eye tracker's online parser, which analyzes eye position data into meaningful events and states, e.g., saccades, fixations, and blinks. Here, saccades are defined by the combination of three indicators—motion threshold of 0.15 (degrees), velocity threshold of 30 (degrees/sec), and acceleration threshold of 8000 (degrees/sec^2). Except for missing samples that are marked as blinks, saccades and fixations were defined and presented alternatively in the recorded files. The function of fixation and saccade extraction enables clear visualization of the scanpath that is informative in terms of the number of fixations and saccades, position and duration of each fixation, as well as saccadic path and sequence during the stimulus presentation (e.g., Figure 1B, scanpath was plotted against footprints of geometric shapes). Through the observation of scanpath, the gaze movement spatiotemporal patterns can be qualitatively identified and quantitatively accessed between individuals with TD and ASD. Future versions of the pipeline can include offline eye movement definition methods, e.g., a deep-learning approach introduced by Startsev, Agtzidis, & Dorr (2019) that avoids user-defined thresholds (i.e., hand-tuned parameters in eye tracking settings) and allows more fine-grained differentiation between fixation and more saccade-like smooth pursuit events or compensating the possible misdefinition of events in suboptimal datasets.

Trial definition and status report. After removal of a series of drifts and outliers from continuous data by following the steps above, the *ft_definetrial* function and a paradigm-specific *trialfun* script in the FieldTrip toolbox can be used to define trials. Finally, the status report will display the percentage of samples marked as *nan* in the current dataset. Subjects who have more than 20% missing data (nan) can be considered for exclusion (Winn et al., 2018). This exclusion criterion left 17 TD datasets and 14 ASD datasets for further analysis.

Preliminary observations

With artifacts-free datasets at hand, further investigation of the research problem—neurocognitive sources of autistic communicative challenges, appears to be more approachable. As suggested in the introduction, the comparison of the *addressee's* spatiotemporal patterns of gaze positions in *pre-target movement* intervals can examine the *altered perceptual function* account since contextual effects of top-down priors are expected to remain constant. Similarly, the *conceptual alignment* account can be tested via a comparison of the *addressee's* gaze patterns in the interval of *post-target movement*, during which top-down priors are likely to alter the gaze pattern through the context-mediated realization of the communicatively relevant portions in movement, i.e., the addressee learns from the previously established shared context that the communicator has finished instructing and starts to complete his/her own configuration, though the clutch cannot be explicitly told. To be specific, I hypothesize that (1) *pre-target movement* intervals witness no significant difference of *addressee's* gaze patterns between *Novel* and *Known* or *ASD* and *TD*, while the latter is tenable if *altered perceptual function* count cannot fully explain the autistic communicative challenges; (2) regarding *post-target movement* intervals, if the *conceptual alignment* consideration can be a potential explanation, different *addressee's* gaze patterns can be found between conditions, which can serve as an indicator that the extent of exploitation of shared context differs

between the novel and known trials, or individuals with ASD or TD. To test the hypotheses, one barrier that is still in the way is our lack of a previously well-established approach to recognize or further compare spatiotemporal patterns of gaze positions that can mark the emergence of understanding both qualitatively and quantitatively. Here, I report attempts to establish a theoretically-reasonable and practically-viable approach to capture the emergence of understanding in the TCG platform and a few preliminary observations when applying the method to current ASD and TD datasets.

Qualifying the emergence of understanding in eye tracking. The first step to investigate the eye tracking marker of emergence of understanding during communicative interaction without any previous knowledge or reference is to look for the potentially shared patterns across participants in actual datasets. Paired participants' scanpath against respective stimulus footprint per trial were plotted to obtain a general and intuitive view of communicative relevant dynamic gaze behavior (Figure 5 presents an example pair). Stimulus footprints were classified into four categories—*pre-target movement*, *on-target adjustment*, *post-target movement* (both during the phase of *communicator's moving*), and *addressee's moving*. The *pre-target movement* was defined as the communicator's initial landing of the addressee's target location; and the *on-target adjustment* refers to the period after the communicator has arrived at the addressee's target location but before the communicator leaves for his/her own target location, during which the communicator usually attempts to develop idiosyncratic strategies to indicate addressee's target orientation (e.g., Figure 1A, the step-out movement). Note the *on-target adjustment* can be absent in a few simple trials that do not require rotation indication, e.g., the trial displayed in Figure 5. The above two categories constitute a *communicative movement*, indicating that the communicator attempts to deliver relevant information via the shape movement. In contrast, *post-target movement*, also called *instrumental movement*, indicates that the communicator aims to leave for his/her own target location with the desired orientation after completing the aforementioned *communicative movement*, thus the movement doesn't entail communicatively informative meaning. The above three categories are all in the experimental phase of *communicator's moving*, thus the division is not predefined or fixed, but highly dependent on trial types, individual communicators, and the ongoing interaction. Besides, to test the *altered perceptual account* versus *conceptual alignment* account, the intervals of the most interest are communicator's communicative and instrumental movement. Therefore, a coarse manual annotation of categories was performed based on the criteria described above to extract the *pre-target movement* and *post-target movement*.

Three spatiotemporal synchronization patterns of scanpath and stimulus footprints were then identified. (1) *Tracking mode*. The tracking mode describes a scenario in which an individual's scanpath closely tracks the stimulus footprint. It commonly presents when no contextual information can be referred to and no prediction can be made contingent on the current situation, e.g., addressee's scanpath in the *pre-target movement* (Figure 5). (2) *Contenting mode*. The tracking mode was found especially in the addressee's scanpath in the *post-target movement* interval, which marks a situation where the addressee understands the intended location the communicator tries to indicate though not explicitly be told. The manifestation and time spent before realization can vary among different individuals or even trials depending on the extent the addressee understand and exploit the shared context, e.g., the addressee in Figure 5 (addressee's scanpath, *post-target movement*) shows a pattern that after watching the communicator's pre-target movement, the gaze position stays close to the resolved target location at the first two steps instead of continuing tracking closely of stimulus footprints as in the pre-target interval, then the eye trace keeps shifting between the target location and communicator's current location. The

mode is named "contenting" because the addressee's eye patterns—regardless of the specific manifestation of patterns—indicate that they usually stop closely tracking the footprints once they feel they have understood the intended target location and orientation. Besides, it's theoretically unnecessary for the addressee to concentrate on communicative-irrelevant instrumental moving as well. Another manifestation of the contenting mode can be that the addressee stares at the target location and does not move gaze anymore after the emergence of understanding. (3) *Predicting mode.* The predicting mode is characterized by the situation that scanpath precedes the stimulus footprints. This mode usually appears in the condition with strong prediction, e.g., one is observing his/her own shape movement with the knowledge of intended location in mind—the communicator in Figure 5 during the *post-target moving* stares at the target position long before s/he actually moves the shape there. Another example is the communicator's scanpath during the phase of *addressee moving* when the communicator has a strong prediction over the position that the address is supposed to land and perhaps is actively checking if the addressee has understood or not.

Through the classification of movement categories and spatiotemporal synchronization patterns, a more organized way of description can be developed. From my observation, the *pre-target movement* mainly consists of the single pattern of *tracking mode* in the addressee's scanpath due to a lack of contextual information. Without access to the target configuration or any other clues, the addressee has to rely on the communicator's steps to resolve the target location, which, reflected by the gaze behavior, is close tracking of the presented stimulus footprints. Yet, as the addressee is aware of the geometric shapes assigned to the communicator once the *pre-target movement* starts, the addressee can have some predictions on the difficulty of the current trial. However, in the *post-target movement*, the synchronization patterns become more complicated with the additional effect from contextual information which may vary between types of trials, individuals, and even occasional random factors. Yet the different manifestations of the synchronization patterns obtain shared traits, one of which is the less correlated scanpath with stimulus footprints, probably resulting from either the unnecessity to attend to the current presented stimulus or a higher cognitive process such as actively zooming out to consider a broad picture instead of simply driven by the presented stimulus. Thus, a closer investigation of the correlation between the addressee's scanpath and stimulus footprints can potentially provide deeper insight into the eye-tracking marker of the emergence of understanding in the TCG platform and further, the potentially different pattern between individuals with ASD and TD. Though a more systematic manual annotation of different movement categories and spatiotemporal synchronization patterns is required in the future to draw a more tenable conclusion as well as to compare and validate the qualitative and quantitative analysis, preliminary attempts can already be made to examine the validity of a potential quantitative approach—cross-correlation to measure the similarity between spatiotemporal synchronization patterns.

Figure 5. Overlays of scanpath and stimulus footprint during a successful trial of a neurotypical participant pair. The trial's target configuration is shown on the right. Fixation locations are indicated with circles containing a plus sign, with circle size corresponding to the duration of the fixation. Saccadic movements are indicated with colored traces, with trace color corresponding to the temporal order of the saccade. Pre-target, post-target components in the communicator moving period were annotated manually. The most interesting eye tracking component is the *addressee's* scanpath (bottom panel) during the communicator moving interval, i.e., the addressee is watching the communicator moving the shape. At this moment, the communicator is also watching his/her own movement (top panel). (1) In the *pre-target moving* interval, the communicator moves the shape to the addressee's target location, thus forming a *communicative moving* to the addressee. Without any other clues, the addressee has to closely track the communicator's shape movement to infer her target location. Reflected in the scanpath (bottom panel, pre-target row), the addressee follows the communicator's movement downwards initially and then rightwards. (2) In the *post-target moving* interval, the communicator starts to move the shape to his/her own target location with desired orientation, which is not communicative

but constitutes an *instrumental moving*. Thus it's theoretically not necessary for the addressee to follow the stimulus footprints anymore (bottom figure, post-target row), e.g., the addressee seems to be staring at the resolved target location initially (first two steps) and then switches to a distinctive pattern that involves constant shift between addressee's target location and communicator's current location (the third to sixth step) and finally gives up following the shape and lands on his/her target location (the seventh step). Note that this is only one of many possible patterns that are observed in the post-target phase compared with the relatively undiversified pre-target phase occupied by the closely-tracking pattern.

Cross-correlation as a quantitative method. Cross-correlation, also known as the sliding dot product, is an approach to measure the similarity of two time series as a function of the displacement of one relative to the other. The two outputs—correlation coefficient (*ccf*) and time lag can provide a closer insight into the examination of the synchronization of scanpath and stimulus footprints. To be specific, the *ccf* is an indicator of the extent of point-wise similarity of two series, e.g., the level of similarity between the coordinates of scanpath and stimulus footprints; and the time lag is an indicator of the preceding or lagging pattern between two time series, e.g., the scanpath slightly precedes the stimulus footprints indicating a tracking mode and the reverse can reflect a predicting mode instead. A combination of low *ccf* and long time lag can indicate the scanpath's lost track of stimulus footprints. Furthermore, based on the previous qualitative observation that pre-target movements mostly consist of a single tracking mode while post-target movements are more diverse in terms of the variety of spatiotemporal synchronization modes (e.g., different manifestation of *contenting mode*), the cross-correlation method seems to be promising to capture the similarity of patterns quantitatively.

Cross-correlation was then performed on both intervals (*pre-target* and *post-target movement*) respectively between the addressee's scanpath and stimulus footprints per participant on a trial-by-trial basis. Then inter-trial and inter-subject values that above or below 3 SD from mean in terms of correlation coefficient or time lag were excluded as outliers. Figure 6A demonstrates an example of the variance of *ccf* after outliers exclusion, either in terms of inter-subject variance per trial or inter-trial variance per subject. Note that the exclusion of outliers in analyses of time lag or other intervals was performed similarly but not plotted here. Then mixed ANOVAs were performed on both *ccf* and time lag with the between-group factor (levels ASD and TD) and the within-group factor *trial type* (levels Novel and Known).

Pre-target synchronization. Novel trials obtain significantly higher *ccf* (Figure 6B) than known trials ($F(1,30) = 23.057$, $p < 0.001$**) but no difference was found between individuals with ASD and TD ($F(1,30) = 1.977$, $p = 0.170$) and the interaction between the two variables is insignificant either ($F(1,30) = 1.584$, $p = 0.218$). In terms of the indicator of *time lag* (Figure 6C), neither between-group ($F(1,30) = 0.385$, $p = 0.540$) or within-group($F(1,30) = 2.002$, $p = 0.168$) difference was found. Yet, when comparing the trendlines (left panel) that represent trend of *ccf* with the progress of trials between conditions, a significant main effect of *trial type* ($F(1,30) = 1076.524$, $p < 0.001$**) and a significant main effect of *group* ($F(1,30) = 96.653$, $p < 0.001$**) were found while the interaction is also significant ($F(1,30) = 48.193$, $p < 0.001$**). Besides, the main effect of trendline of *time lag* is also significant (*trial type*, $F(1,30) = 84.786$, $p < 0.001$**; *group*, $F(1,30) = 47.445$, $p < 0.001$**), but the interaction is not yet significant ($F(1,30) = 2.513$, $p = 0.116$).

The findings are consistent with our hypothesis that during the pre-target interval when the communicator is demonstrating the addressee's target location, the addressee has minimal prediction and no contextual clues. In this case, the addressee's spatiotemporal synchronization is presented in the form of the single *tracking mode*, leading to the observed similar time lag between the novel and known conditions—the differences between trial

types only lie in the manipulation of communicative context. Yet a significantly higher *ccf* in novel trials was found, which can reflect the prediction of difficulty of the trials since the addressee has already obtained the knowledge of both player's assigned geometric shapes in the pre-target interval. For example, if the addressee is assigned a triangle but s/he sees the communicator is moving a circle during the pre-target moving interval, the addressee can know that it will be a tough trial as the communicator will have to use idiosyncratic strategy to explain the target orientation of the addressee's triangle instead of simply rotating the shape. Thus the prediction of difficulty of trials can lead to an allocation of more effort into tracking the communicator's shape and trying to make sense of the movement. As the novel trials are designed to force players to create new strategies without shared communicative context, the difficulty of novel trials is significantly higher than that of known trials. Reflected in gaze behavior, the addressee may concentrate on a smaller area around the moving shape in novel trials, leading to the current higher *ccf* in novel trials compared with known trials. The significant difference of *ccf* in the contrast between the novel and known trials can also suggest that *ccf* can be an indicator of trial difficulty while difficult trials can observe an increased *ccf* between the addressee's scanpath and stimulus footprints. Besides, the decreasing trend of *ccf* and increasing trend of *time lag* as the trial proceeds can reflect a decreased subjective trial difficulty over time. The decreased *ccf* and/or prolonged *time lag* implies that the addressee is probably adopting a looser *tracking mode*—though still higher or longer than those in post-target interval (Figure 7). It can also suggest that the addressees with previous procedural experience can be more comfortable exploiting available information (e.g., from the assigned shapes) to predict and complete the communicative interaction.

Individuals with ASD and TD present a similar significant contrast pattern between *Novel* and *Known* in terms of either *ccf* (Figure 6B) or *time lag* (Figure 6C) indicator, potentially suggesting that the perceptual function involved in TCG is not altered as neurotypical individuals. Yet in terms of the trend over time, the significant main effect of *group* in both *ccf* and *time lag* suggests that they obtain different patterns. Compared with neurotypical individuals, autistic individuals show a lower velocity of change during known trials—either decreasing in terms of *ccf* or increasing in terms of *time lag* overtime but a higher velocity of change during novel trials. For neurotypical addressees, the maintenance of short *time lag* and high *ccf* (closely tracking mode) in novel trials throughout but a rapid change of velocity of time lag or ccf (switch to contenting mode earlier and earlier over time) in known trials reflect a well-balanced allocation of cognitive resources according to different context. The trait requires (1) a quick judgement of the difficulty of the current trial based on the only clue—assigned shapes to both players; (2) the ability to flexibly adjust the gaze pattern accordingly. Therefore, neurotypical individuals are more flexible between the switch of conditions—they are able to quickly judge the difficulty of the current trial and adjust the gaze patterns to the option that they feel reasonable to use (effortful tracking mode v.s. relaxing contenting mode) even if the novel and known trials keep alternating throughout. On the contrary, autistic individuals don't seem to either actively judge the difficulty of the current trial or determine which gaze pattern should be made soon, or both. However, the different patterns between groups regarding trends do not necessarily reflect an altered perceptual function, but can be a higher cognitive ability—decision making. Therefore, the between-group significance of trends shown in the pre-target interval does not indicate that autistic perceptual function is altered. On the contrary, the insignificance of overall *time lag* values between groups despite the significant trends over time renders the evidence more possible to support an unaltered perceptual function in ASD.

Figure 6. Cross-correlation of *addressee's* scanpath and stimulus footprints during *pre-target movement* intervals. **A.** Demonstration of two boxplots for inter-subject variance per trial and inter-trial variance per subject, where variance represents the variance of correlation coefficient (*ccf)* after exclusion of outliers (red plus sign, above 3 SD from mean). The bottom blue curve represents SD over time. **B.** Comparison of the correlation coefficient between conditions, i.e., *ASD* v.s. *TD* and *Novel* v.s. *Known*. In the left panel, the red and blue circles represent *ccf* per trial for novel and known trials, respectively. The red and blue straight lines on the top panel indicate the mean value of *ccf* across all trials after initial subject-wise averaging in novel and known conditions. The dotted black lines are trend lines, indicating the change of *ccf* over trials. In the right panel, a bar plot with error bar and mean value for each participant between conditions is presented. Novel trials obtain higher *ccf* than known trials in both groups ($F(1,30) = 23.057$, $p < 0.001$**). As for the trenda significant main effect of trial type ($F(1,30) = 1076.524$, $p < 0.001$**) and a significant main effect of group ($F(1,30) = 96.653$, $p < 0.001$**) were found while the interaction is also significant ($F(1,30) = 48.193$, $p < 0.001$**) **C.** Comparison of the time lag (ms) between conditions. The legend is the same as B but reflecting the time lag between the *addressee's* scanpath and stimulus footprints. Positive time lag implicates that stimulus footprints appear before the corresponding scanpath while the negative indicates the reverse. Error bars represent s.e.m. (TD num =17, ASD num = 14). * $p < 0.05$; ** $p < 0.001$.

Post-target synchronization. Significant difference of *ccf* was observed in the contrast between *Novel* and *Known* conditions ($F(1,30) = 8.042$, $p = 0.008$*), yet neither between-group difference ($F(1,30) = 0.065$, $p = 0.8$) nor interaction was found significant ($F(1,30) = 0.615$, $p = 0.439$). Regarding the indicator of time lag, a significant main effect of *trial type* was found ($F(1,30) = 5.542$, $p = 0.026$*) while the main effect of *group* ($F(1,30) = 0.018$, $p = 0.894$) or interaction ($F(1,30) = 0.009$, $p = 0.924$) was not significant. As for the trendline, main effects of *trial type* and *group* were significant in terms of the indicator of *ccf* ($F(1,30) = 71.976$, $p < 0.01$**; $F(1,30) = 4.795$, $p = 0.031$*) or time lag ($F(1,30) = 165.987$, $p < 0.001$**; $F(1,30) = 7.772$, $p = 0.006$*), while neither interaction is significant ($F(1,30) = 0.432$, $p = 0.513$; $F(1,30) = 0.283$, $p = 0.596$).

The findings are consistent with our hypothesis in the sense that the presence of diverse spatiotemporal synchronization patterns during the post-target phase will produce overall decreased *ccf* compared with those in pre-target phase (Figure 6). This is probably due to the addressee's switch to a contenting mode after the emergence of understanding and then stop closely tracking the communicator's less informative instrumental movement. Yet the significant high *ccf* in known trials compared with novel trials can be mixed results out of a combination of effects of prediction on trial difficulty and the potential shared contextual knowledge, especially considering that the *Novel* v.s. *Known* contrast of *ccf* is also significant in the pre-target interval (Figure 6B). These entangled effects indicate that *ccf* may not be an option for representing the emergence of mutual understanding but an indicator of prediction on the difficulty of trials based on assigned shapes, or a general indicator of top-down processes. In contrast, the significant *Novel* v.s. *Known* contrast of *time lag* during and only during the post-target period suggests that the *time lag*, as a measurement of temporal synchronization, can capture the synchronization pattern that distinguishes between pre-target and post-target period, thus marking the emergence of understanding. Specifically, the *time lag* during post-target moving in novel trials dramatically fell to near 0 especially in neurotypical participants (Figure 7B) compared to around 50 in pre-target moving, creating the significant difference between novel and known trials. This can reflect a difficulty to exploit enough contextual information in novel trials to alleviate the burden of stimulus-driven tracking of the communicator's movements, because new shared strategies (i.e. method to indicate the addressee's target orientation) are supposed to be established in novel trials by negotiating and the mutual understanding can take quite a few rounds to achieve.

In terms of the trends over time, similar trends of decreased *ccf* and prolonged *time lag* (except for the opposite trend of shortened time lag over time for TD in novel trials, Figure 7B, left middle panel) is observed in the post-target interval. The main effect of *trial type* was significant regarding the trends, indicating that individuals tend to obtain a steeper decrease of *ccf* in novel trials than known trials, which can be caused by (1) addressee' tendency to stare at the target place in case s/he forgot the newly resolved location and/or orientation in difficult novel trials; (2) addressee's zooming out behavior during the gap to look at a broader picture to understand the communicator's intent. Based on the argument above, *ccf* entails a mixture of multiple processes, thus the potential reason underlying the trending pattern can not be elucidated with current information at hand. Besides, regarding the trend of *time lag*, the known trials here observe an apparent difference from that in the pre-target movement that the *time lag* increases with a higher velocity over time, suggesting that more diverse spatiotemporal patterns are utilized in the post-target interval and the addressee's ability to detect the clutch as early as possible is better over time. Due to the atypical decreasing trending pattern for TD in novel trials (Figure 7B, left middle panel), no conclusion can be drawn from the trends regarding the marker of understanding or the distinguished pattern autistic individuals may exhibit. This will be discussed in the discussion section.

Figure 7. Cross-correlation of *addressee's* scanpath and stimulus footprints during *post-target movement* intervals. Legends are the same as Figure 6. **A.** Comparison of the correlation coefficient between conditions, i.e., *ASD* v.s. *TD* and *Novel* v.s. *Known*. Significant difference of *ccf* was observed in the contrast between *Novel* and *Known* conditions ($F(1,30) = 8.042$, $p = 0.008$*). **B.** Comparison of the time lag (ms) between conditions. *Known* conditions obtain prolonged time lag compared with *Novel* conditions ($F(1,30) = 5.542$, $p = 0.026$*). Error bars represent s.e.m. (TD num =17, ASD num = 14). * $p < 0.05$; ** $p < 0.001$.

Discussion

The current preliminary findings capitalize that a method that can quantify the spatiotemporal pattern information is crucial to capture the desired emergence of understanding in TCG platform (e.g., *time lag* in cross-correlation method) instead of simple correlation coefficient values, which can be contaminated by other neurocognitive effects. Yet the current quantitative method of *time lag* is still too coarse to capture some subtle but not trivial information, e.g., a more fine-grained classification of movement behaviors. For example, neurotypical individual's *time lag* in novel trials (Figure 7B) shows a decreasing trend over time and even reaches a negative value at the last few trials, thus with the addressee's scanpath leading the communicator's movement. This observation challenges our previous assumption that only a larger positive time lag can reflect the higher cognitive processes, e.g., zooming out for a broader picture. Instead, the data shows a possibility that neurotypical addressees perform so excellently in the communicative task that they can even predict the communicator's behavior during the post-target stage. This raises a technical question that how can we confirm that the counter-intuitive decreasing trend observed in neurotypical participants instead of the increasing trend found in autistic individuals in the post-target stage is a reflection of successful communication. It's certain that additional behavioral data of trial-by-trial feedback indicates that neurotypical participants perform better in the communicative task than autistic participants. But a more sensitive measurement is still required for traits that can only be measured by eye tracker, e..g, specifically at which step the addressee suddenly understands the

communicator's intention; how does addressee process communicator's on-target adjustment. Besides, most trends except for the above-discussed one show a decreased trend in terms of *ccf* and increased trend in terms of *time lag* in both pre-target and post-target intervals while the between-subject or within-subject contrasts are all significant in both intervals. Therefore, though we have confirmed the validity of *time lag* in characterizing the overall difference of contrast between novel and known trials as well as pre-target and post-target intervals, the *time lag* indicator did not perform well regarding the trend over time. Based on the considerations above, a more fine-grained, dynamic and sensitive measurement of spatiotemporal pattern similarity is supposed to be established for allowing the capture of the communicative dynamics over time.

A more fine-grained quantitative method

As our optimal goal is to compare the similarity of scanpath and stimulus footprints to look for the pattern that potentially marks the emergence of understanding in the TCG platform, thus the essence of the question can be categorized as pattern comparison, where the pattern can be regarded as three-dimensional—x and y coordinates and time. With the term "pattern", we implicitly deliver a request that the optimal approach is supposed to be invariant to translation, scaling or even rotation as these pattern-irrelevant traits can vary between trials and individuals, thus exerting an undesired effect on the results. The currently applied indicators such as correlation coefficient, are sensitive to noise or outliers. Though an outlier exclusion was applied in the study, it remains unknown to what extent the current results are driven by noise. The previous eye tracking studies that include scanpath comparison into analysis focus on the comparison between two scanpath where eye tracking-related features can be extracted and compared (Anderson et al., 2015), yet I am mostly interested in the scanpath-footprints relationship that has potential to mark the emergence of understanding. Therefore, an approach from the pattern recognition domain namely canonical time warping is considered to enable the capture of dynamic scanpath patterns relatively to stimulus footprints.

Canonical time warping method. The canonical time warping method (CTW, Zhou & De la Torre, 2009) is a combination of dynamic time warping (DTW) and canonical correlation analysis (CCA). DTW is an algorithm to measure the similarity between two time series which may vary in speed and can be of different lengths. The former trait allows us to quantify the dynamic change of spatiotemporal patterns even in one segment (continuous warping), e.g., switching from *tracking mode* to *contenting mode*. In contrast, the indicator *time lag* in cross-correlation can only produce one value given two user-defined segments. The latter trait enables the possibility to use the original number of stimulus footprints and gaze data instead of stretching the two time series beforehand. DTW offers a more flexible and dynamic measurement in the temporal domain, while the CCA extracts the real patterns from irrelevant noise in the spatial domain. Canonical warping distance is invariant to translation, scaling, and rotation, which is quite preferable for my eye tracking datasets, not only in the sense that it allows the capture of real patterns without detrimental noise but also in terms of freeing the datasets from potential contamination of drifts artifacts. As I explained before, the eye tracking datasets suffer from an unignorable influence of drifts and linear trends as no inter-trial calibration was performed during data collection out of the pursuit of continuous interactive communication. Yet the application of CCA is similar to performing a position recalibration provided as an optional module in the preprocessing pipeline, which is based on linear transformation. At the same time,

there is no need to worry about the side effects brought by the position recalibration, e.g., create artificial differences between conditions or cancel out desired traits. An updated version of CTW (Zhou & De la Torre, 2016) further enables the comparison between multi-modal sequences. The new characteristic allows us to regard stimus footprints as an area instead of purely a point in the center position. In sum, CTW, as the combination of the temporal flexibility in DTW and spatially pattern extraction in CCA, provides a promising approach to quantitatively examine the spatiotemporal synchronization patterns of scanpath and stimulus footprints.

Dual-EEG methods

Methodological opportunities to further examine the neurocognitive sources of autistic communicative challenges can also be manifested from the electrophysiological perspective. While eye tracking studies reveal the emergence of understanding via identifying and comparing different spatiotemporal synchronization patterns of scanpath and stimulus footprints, a dual-EEG setup can enable the delineation of temporal and spectral properties of conceptual mis/alignment in individuals with ASD and TD, thus illuminating the neural mechanisms underlying autistic communicative challenges. To be specific, if the conceptual alignment account is true, autistic pairs may have difficulty using the newly-established shared communicative context, which will lead to a lower neural alignment in novel trials compared with neurotypical individuals.

Based on the above assumption, a dual-EEG setup (Figure 8A) was established via the connection of two EEG labs at Donders Center for Cognitive Neuroimaging (DCCN). Paired participants can thus sit individually in each lab and communicate via the same TCG platform on the screen. Pilot studies were performed before formal data collection for a sanity check (Figure 8B). The purpose of the sanity check is to examine the validity of the dual-EEG setup via investigation of underlying brain oscillations. With the same frequency band between 8 and 12 Hz, posterior *alpha* power is considered to indicate visual activities while the central *mu* power reflects sensorimotor activities. Then a cross-correlation was performed and 8-12 Hz power coefficient was found to be most evident over central and posterior regions. Similar results were also observed in terms of the indicator of time lag—channels with the most synchronized trials are in central and posterior regions. Therefore, the dual-EEG setup is successful in terms of simultaneous display of visual stimulus and recording of motor response. The project was presented at the Project Proposal Meeting at DCCN in April. To date, five pairs of neurotypical datasets have been collected. Further analysis of the electrophysiological datasets can provide deeper insight into the question of neurocognitive sources of autistic communicative challenges.

Figure 8. Dual-EEG setup and sanity check. **A.** Demonstration of dual-EEG setup in practice (left) and example trials showing successful simultaneous recording (right). The sender precedes the receiver in terms of sensorimotor activities as the sender presses the buttons before the receiver; but synchronizes with the receiver in terms of visual activities, indicating the successful simultaneous display of visual stimulus. **B.** Synchronization of posterior *alpha* power and central *mu* power measured by cross-correlation. The most evident 8-12 Hz power coefficient appears over central and posterior regions (left). Channels with the most synchronized trials (with black dotted frame) are also in central and posterior regions (right), indicating visual and sensorimotor synchronization. Heatmap shows the distribution of time lag of trials per channel, and the right topography plots the central time-lag bins. The TCG setup in a dual-EEG platform is successful.

Conclusion

To trace the cognitive sources of communicative challenges in ASD, i.e., *altered perceptual account* v.s. *conceptual alignment* account, a live communication platform (TCG) in combination with eye tracking methods were established to capture both the emergence of understanding and the potential difference in terms of spatiotemporal synchronization patterns between individuals with ASD and TD. Specifically, to achieve clean eye tracking datasets, a preprocessing pipeline was built initially to exclude all kinds of artifacts and outliers observed in the actual datasets (e.g, blinks and eyelid signal, drifts, etc). Without previous knowledge of the eye tracking marker of understanding during TCG, a qualitative method was then applied to classify the moving behaviors into four categories, identify three kinds of spatiotemporal synchronization patterns of scanpath and stimulus footprints, and form hypotheses to enable the examination of both emergence of understanding and autistic individuals' potential different gaze patterns. Then a quantitative approach of cross-correlation was applied to the moving intervals of interest while one of the outputs—*time lag* that can differentiate between different synchronization patterns was shown to be a more valid indicator compared with the correlation coefficient. This validation is based on a within-group contrast and then applies to the ASD v.s. TD contrast. Preliminary evidence inclines towards an unaltered perceptual function in ASD but lacks further evidence to confirm the shared conceptual account for autistic communicative challenges, possibly due to the current coarse quantitative method that neglects subtle but not trivial characteristics during communicative interaction. Therefore, a more advanced and fine-grained quantitative method was then discussed theoretically to suggest the potential fitness and benefits. Finally a combined dual-EEG and eye tracking study set up and validation were reported to further open the possibility to investigate the neurocognitive sources of communicative challenges in ASD from an electrophysiological perspective.

References

Akmajian, A., Farmer, A. K., Bickmore, L., Demers, R. A., & Harnish, R. M. (2017). *Linguistics: An introduction to language and communication*. MIT press.

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (DSM-5®). American Psychiatric Pub.

Amso, D., Haas, S., Tenenbaum, E., Markant, J., & Sheinkopf, S. J. (2014). Bottom-up attention orienting in young children with autism. *Journal of autism and developmental disorders*, *44*(3), 664-673.

Anderson, N. C., Anderson, F., Kingstone, A., & Bischof, W. F. (2015). A comparison of scanpath comparison methods. *Behavior research methods*, *47*(4), 1377-1392.

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

Bhat, A. N., Galloway, J. C., & Landa, R. J. (2012). Relation between early motor delay and later communication delay in infants at risk for autism. *Infant Behavior and Development*, *35*(4), 838-846.

Bolis, D., & Schilbach, L. (2018). Observing and participating in social interactions: action perception and action control across the autistic spectrum. *Developmental cognitive neuroscience*, *29*, 168-175.

Booth, R. D., & Happé, F. G. (2018). Evidence of reduced global processing in autism spectrum disorder. *Journal of autism and developmental disorders*, *48*(4), 1397-1408.

Brennan, J. R., Lajiness-O'Neill, R., Bowyer, S., Kovelman, I., & Hale, J. T. (2019). Predictive sentence comprehension during story-listening in autism spectrum disorder. *Language, Cognition and Neuroscience*, *34*(4), 428-439.

Caffier, P. P., Erdmann, U., & Ullsperger, P. (2003). Experimental evaluation of eye-blink parameters as a drowsiness measure. *European journal of applied physiology*, *89*(3), 319-325.

Casartelli, L., Federici, A., Fumagalli, L., Cesareo, A., Nicoli, M., Ronconi, L., ... & Sinigaglia, C. (2020). Neurotypical individuals fail to understand action vitality form in children with autism spectrum disorder. *Proceedings of the National Academy of Sciences*, *117*(44), 27712-27718.

Carter, B. T., & Luke, S. G. (2018). Individuals' eye movements in reading are highly consistent across time and trial. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(3), 482.

Chambon, V., Farrer, C., Pacherie, E., Jacquet, P. O., Leboyer, M., & Zalla, T. (2017). Reduced sensitivity to social priors during action prediction in adults with autism spectrum disorders. *Cognition*, *160*, 17-26.

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in cognitive sciences*, *16*(4), 231-239.

Chita-Tegmark, M. (2016). Attention allocation in ASD: a review and meta-analysis of eye-tracking studies. *Review Journal of Autism and Developmental Disorders*, *3*(3), 209-223.

Chita-Tegmark, M. (2016). Social attention in ASD: A review and meta-analysis of eye-tracking studies. *Research in developmental disabilities*, *48*, 79-93.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences*, *36*(3), 181-204.

Clements, C. C., Zoltowski, A. R., Yankowitz, L. D., Yerys, B. E., Schultz, R. T., & Herrington, J. D. (2018). Evaluation of the social motivation hypothesis of autism: a systematic review and meta-analysis. *JAMA psychiatry*, *75*(8), 797-808.

Constantino, J. N., Kennon-McGill, S., Weichselbaum, C., Marrus, N., Haider, A., Glowinski, A. L., ... & Jones, W. (2017). Infant viewing of social scenes is under genetic control and is atypical in autism. *Nature*, *547*(7663), 340-344.

Cusack, J. P., Williams, J. H., & Neri, P. (2015). Action perception is intact in autism spectrum disorder. *Journal of Neuroscience*, *35*(5), 1849-1857.

Davis, R., & Crompton, C. J. (2021). What Do New Findings About Social Interaction in Autistic Adults Mean for Neurodevelopmental Research?. *Perspectives on Psychological Science*, *16*(3), 649-653.

Demetriou, E. A., Lampit, A., Quintana, D. S., Naismith, S. L., Song, Y. J. C., Pye, J. E., ... & Guastella, A. J. (2018). Autism spectrum disorders: a meta-analysis of executive function. *Molecular psychiatry*, *23*(5), 1198-1204.

de Vries, L., Fouquaet, I., Boets, B., Naulaers, G., & Steyaert, J. (2020). Autism spectrum disorder and pupillometry: A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews*.

Friston, K. J., Lawson, R., & Frith, C. D. (2013). On hyperpriors and hypopriors: comment on Pellicano and Burr. *Trends in cognitive sciences*, *17*(1), 1.

Feigin, H., Shalom-Sperber, S., Zachor, D. A., & Zaidel, A. (2021). Increased influence of prior choices on perceptual decisions in autism. *Elife*, *10*, e61595.

Friston, K. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society B: Biological*

sciences, *360*(1456), 815-836.

Geller, J., Winn, M. B., Mahr, T., & Mirman, D. (2020). GazeR: A package for processing gaze position and pupil size data. *Behavior research methods*, *52*(5), 2232-2255.

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.

Goldwater, B. C. (1972). Psychological significance of pupillary movements. *Psychological bulletin*, *77*(5), 340.

Hanke, M., Mathôt, S., Ort, E., Peitek, N., Stadler, J., & Wagner, A. (2019). A practical guide to functional magnetic resonance imaging with simultaneous eye tracking for cognitive neuroimaging research. In *Spatial Learning and Attention Guidance* (pp. 291-305). Humana, New York, NY.

Happé, F. G. (1993). Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition*, *48*(2), 101-119.

Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S., & Keysers, C. (2012). Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends in cognitive sciences*, *16*(2), 114-121.

Hellmer, K., & Nyström, P. (2017). Infant acetylcholine, dopamine, and melatonin dysregulation: Neonatal biomarkers and causal factors for ASD and ADHD phenotypes. *Medical hypotheses*, *100*, 64-66.

Hershman, R., Henik, A., & Cohen, N. (2018). A novel blink detection method based on pupillometry noise. *Behavior research methods*, *50*(1), 107-114.

Hershman, R., Henik, A., & Cohen, N. (2019). CHAP: Open-source software for processing and analyzing pupillometry data. *Behavior Research Methods*, *51*(3), 1059-1074.

Hill, E. L. (2004). Executive dysfunction in autism. *Trends in cognitive sciences*, *8*(1), 26-32.

Jones, K. E., Craver-Lemley, C., & Barrett, A. M. (2008). Asymmetrical visual-spatial attention in college students diagnosed with ADD/ADHD. *Cognitive and behavioral neurology: official journal of the Society for Behavioral and Cognitive Neurology*, *21*(3), 176.

Jung, M., Tu, Y., Lang, C. A., Ortiz, A., Park, J., Jorgenson, K., ... & Kong, J. (2019). Decreased structural connectivity and resting-state brain activity in the lateral occipital cortex is associated with social communication deficits in boys with autism spectrum disorder. *Neuroimage*, *190*, 205-212.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological review*, *87*(4), 329.

Kalandadze, T., Norbury, C., Nærland, T., & Næss, K. A. B. (2018). Figurative language comprehension in individuals with autism spectrum disorder: A meta-analytic review. *Autism*, *22*(2), 99-117.

Knapen, T., Swisher, J. D., Tong, F., & Cavanagh, P. (2016). Oculomotor remapping of visual information to foveal retinotopic cortex. *Frontiers in systems neuroscience*, *10*, 54.

Koldewyn, K., Jiang, Y. V., Weigelt, S., & Kanwisher, N. (2013). Global/local processing in autism: Not a disability, but a disinclination. *Journal of autism and developmental disorders*, *43*(10), 2329-2340.

Kret, M. E., & Sjak-Shie, E. E. (2019). Preprocessing pupil size data: Guidelines and code. *Behavior research methods*, *51*(3), 1336-1342.

Lawson, R. P., Mathys, C., & Rees, G. (2017). Adults with autism overestimate the volatility of the sensory environment. *Nature neuroscience*, *20*(9), 1293-1299.

LeBarton, E. S., & Landa, R. J. (2019). Infant motor skill predicts later expressive language and autism spectrum disorder diagnosis. *Infant Behavior and Development*, *54*, 37-47.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1-36.

Malkin, L., & Abbot-Smith, K. (2021). How set switching affects the use of context-appropriate language by autistic and neuro-typical children. *Autism*, 13623613211012860.

Marandi, R. Z., Madeleine, P., Omland, Ø., Vuillerme, N., & Samani, A. (2018). Eye movement characteristics reflected fatigue development in both young and elderly individuals. *Scientific reports*, *8*(1), 1-10.

Mathôt, S. (2018). Pupillometry: psychology, physiology, and function. *Journal of Cognition*, *1*(1).

Mathôt, S., Fabius, J., Van Heusden, E., & Van der Stigchel, S. (2018). Safe and sensible preprocessing and baseline correction of pupil-size data. *Behavior research methods*, *50*(1), 94-106.

Mathôt, S., Van der Linden, L., Grainger, J., & Vitu, F. (2013). The pupillary light response reveals the focus of covert visual attention. *PloS one*, *8*(10), e78168.

Mathys, C. D., Lomakina, E. I., Daunizeau, J., Iglesias, S., Brodersen, K. H., Friston, K. J., & Stephan, K. E. (2014). Uncertainty in perception and the Hierarchical Gaussian Filter. *Frontiers in human neuroscience*, *8*, 825.

McDougal, D. H., & Gamlin, P. D. (2010). The influence of intrinsically-photosensitive retinal ganglion cells on the spectral sensitivity and response dynamics of the human pupillary light reflex. *Vision research*, *50*(1), 72-87.

Mody, M., Shui, A. M., Nowinski, L. A., Golas, S. B., Ferrone, C., O'Rourke, J. A., & McDougle, C. J. (2017). Communication deficits and the motor system: exploring patterns of associations in autism spectrum disorder (ASD). *Journal of autism and developmental disorders*, *47*(1), 155-162.

Mottron, L., Dawson, M., Soulieres, I., Hubert, B., & Burack, J. (2006). Enhanced perceptual functioning in autism: an update, and eight principles of autistic perception. *Journal of autism and developmental disorders*, *36*(1), 27-43.

Nowak, W., Hachoł, A., & Kasprzak, H. (2008). Time-frequency analysis of spontaneous fluctuation of the pupil size of the human eye. *Optica Applicata*, *38*(2).

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational intelligence and neuroscience*, *2011*.

Palmer, C. J., Lawson, R. P., & Hohwy, J. (2017). Bayesian approaches to autism: Towards volatility, action, and behavior. *Psychological bulletin*, *143*(5), 521.

Palmer, C. J., Seth, A. K., & Hohwy, J. (2015). The felt presence of other minds: Predictive processing, counterfactual predictions, and mentalizing in autism. *Consciousness and Cognition*, *36*, 376-389.

Pellicano, E., & Burr, D. (2012). When the world becomes 'too real': a Bayesian explanation of autistic perception. *Trends in cognitive sciences*, *16*(10), 504-510.

Pennington, B. F., & Ozonoff, S. (1996). Executive functions and developmental psychopathology. *Journal of child psychology and psychiatry*, *37*(1), 51-87.

Porges, S. W. (2005). The role of social engagement in attachment and bonding. *Attachment and bonding*, *3*, 33-54.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and brain sciences*, *36*(4), 329-347.

Pulvermüller, F., Garagnani, M., & Wennekers, T. (2014). Thinking in circuits: toward neurobiological explanation in cognitive neuroscience. *Biological cybernetics*, *108*(5), 573-593.

Quiñones-Camacho, L. E., Fishburn, F. A., Belardi, K., Williams, D. L., Huppert, T. J., & Perlman, S. B. (2021). Dysfunction in interpersonal neural synchronization as a mechanism for social impairment in autism spectrum disorder. *Autism Research*.

Rayner, K., & Reingold, E. M. (2015). Evidence for direct cognitive control of fixation durations during reading. *Current opinion in behavioral sciences*, *1*, 107-112.

Redcay, E., & Schilbach, L. (2019). Using second-person neuroscience to elucidate the mechanisms of social interaction. *Nature Reviews Neuroscience*, *20*(8), 495-505.

Satterthwaite, T. D., Wolf, D. H., Ruparel, K., Erus, G., Elliott, M. A., Eickhoff, S. B., ... & Gur, R. C. (2013). Heterogeneous impact of motion on fundamental patterns of developmental changes in functional connectivity during youth. *Neuroimage*, *83*, 45-57.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience 1. *Behavioral and brain sciences*, *36*(4), 393-414.

Selten, R., & Warglien, M. (2007). The emergence of simple languages in an experimental coordination game. *Proceedings of the National Academy of Sciences*, *104*(18), 7361-7366.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, *27*(3), 379-423.

Sirois, S., & Jackson, I. R. (2012). Pupil dilation and object permanence in infants. *Infancy*, *17*(1), 61-78.

Sogo, H. (2013). GazeParser: an open-source and multiplatform library for low-cost eye tracking and analysis. *Behavior research methods*, *45*(3), 684-695.

Stark, E., Stacey, J., Mandy, W., Kringelbach, M. L., & Happé, F. (2021). Autistic Cognition: Charting Routes to Anxiety. *Trends in Cognitive Sciences*.

Stern, J. A., Boyer, D., & Schroeder, D. (1994). Blink rate: a possible measure of fatigue. *Human factors*, *36*(2), 285-297.

Startsev, M., Agtzidis, I., & Dorr, M. (2019). 1D CNN with BLSTM for automated classification of fixations, saccades, and smooth pursuits. *Behavior Research Methods*, *51*(2), 556-572.

Stolk, A., Bašnáková, J., & Toni, I. (2020). Joint epistemic engineering: The neglected process of context construction in human communication.

Stolk, A., Noordzij, M. L., Volman, I., Verhagen, L., Overeem, S., van Elswijk, G., ... & Toni, I. (2014). Understanding communicative actions: A repetitive TMS study. *Cortex*, *51*, 25-34.

Stolk, A., Todorovic, A., Schoffelen, J. M., & Oostenveld, R. (2013). Online and offline tools for head movement compensation in MEG. *Neuroimage*, *68*, 39-48.

Stolk, A., Verhagen, L., Schoffelen, J. M., Oostenveld, R., Blokpoel, M., Hagoort, P., ... & Toni, I. (2013). Neural mechanisms of communicative innovation. *Proceedings of the National Academy of Sciences*, *110*(36), 14574-14579.

Stolk, A., Verhagen, L., & Toni, I. (2016). Conceptual alignment: How brains achieve mutual understanding. *Trends in cognitive sciences*, *20*(3), 180-191.

Tamietto, M., Castelli, L., Vighetti, S., Perozzo, P., Geminiani, G., Weiskrantz, L., & de Gelder, B. (2009). Unseen facial and bodily expressions trigger fast emotional reactions. *Proceedings of the National Academy of*

Sciences, *106*(42), 17661-17666.

Tewolde, F. G., Bishop, D. V., & Manning, C. (2018). Visual motion prediction and verbal false memory performance in autistic children. *Autism Research*, *11*(3), 509-518.

Thomas, M. S., Davis, R., Karmiloff-Smith, A., Knowland, V. C., & Charman, T. (2016). The over-pruning hypothesis of autism. *Developmental Science*, *19*(2), 284-305.

Toni, I., & Stolk, A. (2019). Conceptual alignment as a neurocognitive mechanism for human communicative interactions. *Human Language: From Genes and Brains to Behavior*, 249.

Vadillo, M. A., Street, C. N., Beesley, T., & Shanks, D. R. (2015). A simple algorithm for the offline recalibration of eye-tracking data through best-fitting linear transformation. *Behavior research methods*, *47*(4), 1365-1376.

Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: predictive coding in autism. *Psychological review*, *121*(4), 649.

Wadge, H., Brewer, R., Bird, G., Toni, I., & Stolk, A. (2019). Communicative misalignment in autism spectrum disorder. *Cortex*, *115*, 15-26.

Wainstein, G., Rojas-Líbano, D., Crossley, N. A., Carrasco, X., Aboitiz, F., & Ossandón, T. (2017). Pupil size tracks attentional performance in attention-deficit/hyperactivity disorder. *Scientific reports*, *7*(1), 1-9.

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.

Watanabe, S., Miki, K., & Kakigi, R. (2002). Gaze direction affects face perception in humans. *Neuroscience letters*, *325*(3), 163-166.

Weiskrantz, L., Cowey, A., & Barbur, J. L. (1999). Differential pupillary constriction and awareness in the absence of striate cortex. *Brain*, *122*(8), 1533-1538.

Winn, M. B., Wendt, D., Koelewijn, T., & Kuchinsky, S. E. (2018). Best practices and advice for using pupillometry to measure listening effort: An introduction for those who want to get started. *Trends in hearing*, *22*, 2331216518800869.

Wheatley, T., Boncz, A., Toni, I., & Stolk, A. (2019). Beyond the isolated brain: the promise and challenge of interacting minds. *Neuron*, *103*(2), 186-188.

Wittgenstein, L. (2010). *Philosophical investigations*. John Wiley & Sons.

Zalla, T., Amsellem, F., Chaste, P., Ervas, F., Leboyer, M., & Champagne-Lavau, M. (2014). Individuals with autism spectrum disorders do not use social stereotypes in irony comprehension. *PloS one*, *9*(4), e95568.

Zhang, Y., & Hornof, A. J. (2011). Mode-of-disparities error correction of eye-tracking data. *Behavior research methods*, *43*(3), 834-842.

Zhegallo, A. V., & Marmalyuk, P. A. (2015). ETRAN—R extension package for eye tracking results analysis. *Perception*, *44*(8-9), 1129-1135.

Zhou, J., Park, C. Y., Theesfeld, C. L., Wong, A. K., Yuan, Y., Scheckel, C., ... & Troyanskaya, O. G. (2019). Whole-genome deep-learning analysis identifies the contribution of noncoding mutations to autism

risk. *Nature genetics*, *51*(6), 973-980.

Zhou, F., & De la Torre, F. (2016). Spatio-temporal matching for human pose estimation in video. *IEEE transactions on pattern analysis and machine intelligence*, *38*(8), 1492-1504.

Zhou, F., & Torre, F. (2009). Canonical time warping for alignment of human behavior. *Advances in neural information processing systems*, *22*, 2286-2294.

Zwaigenbaum, L., Bryson, S., Rogers, T., Roberts, W., Brian, J., & Szatmari, P. (2005). Behavioral manifestations of autism in the first year of life. *International journal of developmental neuroscience*, *23*(2-3), 143-152.