Decoding hidden language: towards a reliable neural assessment of language comprehension in minimally-verbal autistic children

Selene Petit, MSc.

Department of Cognitive Science
ARC Centre of Excellence of Cognition and its Disorders

Faculty of Human Sciences
Macquarie University, Sydney, Australia

Presented for the degree of Doctor of Philosophy (PhD)

December 2019

Supervisors:

Dr.	Nicholas A. Badcock	Macquarie University
Dr.	Alexandra Woolgar	University of Cambridge
Professor	Elizabeth Pellicano	Macquarie University

Table of contents

Summary

Language is fundamental for cognition and social functioning, and something we often take for granted. In cases where people cannot use language to communicate, it can be difficult to evaluate cognitive abilities and the extent of spoken language understanding. In the case of minimally-verbal autistic people, accumulating evidence suggests a discrepancy between their receptive and productive language skills. In particular, it is becoming evident that at least some minimally-verbal children may understand more language than they can demonstrate. Because of the nature of their condition, including behavioural challenges and difficulties in social situations, autistic children often perform poorly on traditional measures of language such as standardised tests. Therefore, passive tests that do not require verbal answers may be more suitable. The aim of this thesis was to develop neural tests of language comprehension for minimally-verbal autistic children that do not require overt behavioural answers. Chapters 2 to 4 cover the development and validation of four paradigms in neurotypical children, and Chapter 5 presents the results of one of the paradigms applied to three minimally-verbal autistic children.

In Chapter 1, I present the scope and the challenges of this work, and I review the relevant literature. In Chapter 2, I developed two child-friendly auditory paradigms that use spoken words and sentences to detect lexico-semantic processing in neurotypical children (N = 31), using electroencephalography (EEG). In this work, I also evaluated the data recording quality of a low-cost portable EEG system, Emotiv EPOC+, which provides an easy-to-setup and affordable EEG option. I compared two data analyses approaches: univariate and multivariate pattern analyses (MVPA). Three main findings emerged: first, there was large inter-individual variability in neural signals, with only around 50% of individuals showing a statically

significant univariate effect. Second, the EPOC+ recorded similar signal to the research-grade system, when analysing the data with a univariate approach. Third, MVPA, which is robust to intra-individual differences in the time course and topology of effect, improved our reliability at the individual level to a maximum of 88% of children. This provided a promising avenue for a covert assessment of lexico-semantic processing in children. In Chapter 3, I extended this work and designed an updated paradigm in which I better controlled the stimuli, and added visual animations. I again found large inter-individual variability in the neural responses to semantic processing in children (N = 20), and the detection rate did not improve from Chapter 2. In Chapter 4, I turned to a novel paradigm that used functional transcranial Doppler ultrasound (fTCD) to assess mental task-following from the brain activity of neurotypical children (N = 20). I found that, for identical visual stimuli, children showed distinct lateralisation patterns when performing language and visuo-spatial memory tasks, with language being left-lateralised, and visuo-spatial memory being more bilateral. However, at the individual-subject level, I again found only half of participants showing statistically reliable task-following neural patterns. In Chapter 5, I selected the most promising paradigm (one of the EEG paradigms from Chapter 2) and tested three minimally-verbal autistic children (aged 5 to 15 years). I demonstrated that my techniques can be used with this population and I found evidence of language comprehension from one child's neural activity without requiring behavioural answers from him. Finally, in Chapter 6, I discuss the implications of these results for assessing language comprehension in minimally-verbal populations, and propose some future directions for this work.

Together, this research provides a rigorous exploration of the methodological issues in using neuroimaging to assess cognition at the individual subject level, and an initial proof-of-concept that we can measure intact lexico-semantic processing in populations that may otherwise struggle to communicate.

Author note

This thesis was prepared in the form of a thesis by publication, thus the experimental chapters were prepared in the form of stand-alone research articles. Chapter 2 was submitted to JSLHR and is currently under review, Chapter 3 was submitted to Scientific Reports and is currently under review, and Chapter 4 was submitted to Neuropsychologia and is currently under review. Chapter 5 is a small proof-of-concept study and as such has not been submitted. As each of the empirical chapters is an individual paper, there is some unavoidable repetition in the introduction and method sections. Bibliographies can be found at the end of each chapter.

All stimuli, presentation scripts, and analysis scripts, are available in Open Science Framework repositories. The repositories can be accessed as follows: Chapter 2 - https://osf.io/dhpgn/; Chapter 3 - https://osf.io/bv2dy/; Chapter 4 - https://osf.io/xygjv/; Chapter 5 - https://osf.io/q7myw/. Raw data will be made available upon publication of the manuscripts.

Author statement

I, Selene Petit, certify that the research presented in this thesis is my original work and is written by me. Any sources of information or assistance are appropriately cited and acknowledged. I am the main contributor to all projects presented for this degree. Some of the data presented in Chapter 2 (22 of the 31 participants) was collected by me during my Master's degree, which was written up as a research dissertation at Grenoble Alpes University, France. During my PhD I collected the 9 additional participants in that chapter, designed and carried out a substantially different set of analyses, refined the interpretation of results, and wrote the paper in its current form. All other data, analyses, and writing presented in this thesis are original and have not been previously submitted.

All research presented in this thesis was approved by the Macquarie University Human Research Ethics Committee (reference numbers: 5201200658 and 5201833584597, see Appendix).

Signed,

Selene Petit, December 2019

Acknowledgements

They said that doing a PhD may be difficult and, at times, lonely. Yet, throughout my PhD journey, I never once lacked support, and I can only begin to offer my sincerest gratitude to everyone who has contributed to this achievement.

First and foremost, I would like to thank my fantastic supervision team, starting with my mentor, Alexandra Woolgar. Alex, you have patiently watched me grow since I took my first researcher-wannabe steps, and I feel so fortunate to have learned from such an inspiring scientist. Your never-ending support, clever ideas, and infinite patience have been a textbook demonstration of what outstanding supervision looks like. You have always provided honest feedback and never made me feel foolish for asking too many questions. Your enthusiasm and brilliance are inspiring, and I am so grateful that our paths crossed. To my co-supervisor Nicholas Badcock, thank you for being so generous with your time and advice. I have never once left our meetings without feeling better about myself, and this has made all the difference. Also thank you to my associate supervisor Liz Pellicano for her support, positivity, and meticulous feedback on my work.

I am sincerely grateful for the support I have received at the Department of Cognitive Science at Macquarie University. I have met the most interesting people, collaborators, and friends. First, I would like to thank my collaborator and friend Tijl Grootswagers for patiently unravelling the mysteries of MVPA for me. Tijl, sitting next to you while you wrote a 100-line script in two minutes was surely the fastest way to learn, especially when complemented by witty but useful explanations. I extend my gratitude to the whole department, in particular to all the researchers who have provided feedback on my work, and the administrative team for helping me navigate the obscure world of bureaucracy and administrative paperwork.

I would also like to thank all the children and their families who generously gave their time to participate in my experiments. This work would not have been possible without them.

A huge thank you to my PhD companions Lina Teichmann and Denise Moerel. You have stood by me through the good, bad, and worse days and you have made this journey so much more fun. Thank you for sharing coffees and laughs, Aussie barbecues and Easter raclettes, champagne and European traditions, and the occasional red wine and nervous breakdown with me. Thank you for always making the time to support me, academically and personally.

I also want to thank my dear friends in Australia, in particular Zsanett Szilàgyi, Fix Decroix, Julien Millasseau, and James Cummins, who have made me feel a little more at home in this far-away country. Thank you for patiently listening to my complaining and bringing me happiness and joy all the way through my hardships. I know your door will always be open for me, through the good and the hard times, and I am forever grateful to you all for making my life better. A special thank you also goes to my dear friend and better half, Estelle Willemet. Although it was hard to watch me leave to the other side of the world, you always made sure to hide your tears and cheer for me instead. Thank you for keeping me in your life despite the distance, and for the constant reminder that worrying never makes things better, and laughing always does.

Finally, to my family, to whom I owe everything. I have missed you dearly, but the lessons you taught me followed me across the world. To my parents, Marie-Laure and Jean-Pierre, I am grateful to you both for more than could fit on this page. Thank you for teaching me everything that I did not learn in the books. Thank you for always believing in me so fervently, that today, I know how to do it for myself. Thank you for encouraging me to make brave decisions, even those that took me so far from home. To my brother Médéric and his growing family, thank you for always keeping your cool, and not letting the distance get in the way. My layovers in

Bangkok with you never fail to remind me how to be chill, which I often seem to forget! Lastly, to my grandma and personal heroine Odile, thank you for regularly fighting insubordinate technology to send me emails of encouragement. You have taught me that laughing is a cure, that age is only an excuse, and that optimism and curiosity matter more than perfect results. *Merci.*

Thank you.

Chapter 1

Introduction

The human ability to use speech is central for our social and cognitive functioning. In some cases, however, individuals who may otherwise have relatively intact mental functioning are not able to use spoken language to communicate. Such instances include individuals with cerebral palsy, minimally-verbal autism, or patients with disorders of consciousness. In these cases, it can be difficult to assess the true nature of their cognitive abilities, including how much language they actually understand. In the case of minimally-verbal autism, recent anecdotal and empirical evidence suggests that at least a subset of individuals may have preserved receptive language. For example, anecdotal reports from individuals who master an alternative mode of communication indicate that lack of speech does not always reflect an absence of language comprehension. Additionally, empirical findings from a subset of non-speaking[1] individuals suggest that neural processes associated with receptive language may be preserved, despite unreliable behavioural responses. This mounting evidence, and its potentially transformative implication, urges us to develop objective tests of language comprehension for these individuals.

[1] In this chapter, I will use the terms "non-speaking" and "minimally-verbal" (see below) as synonyms. "Non-speaking" is favoured by the autistic community, while "minimally-verbal" is the most commonly used term in the autism literature.

Minimally-verbal autistic[2] individuals are notoriously difficult to test due to their behavioural and social difficulties. For instance, during standardised tests of cognitive functioning, such as structured interviews or behavioural tests, they may struggle to adjust to the social context, feel anxious or frustrated, or lack motivation to complete the test. They may also lack behavioural skills such as pointing, which is often required in these tests (Helen Tager-Flusberg, 1999). Thus, most standardised assessments are inappropriate for this population, and yield unreliable results (Plesa Skwerer, Jordan, Brukilacchio, & Tager-Flusberg, 2016; Helen Tager-Flusberg & Kasari, 2013). Yet, accurate assessment of their cognitive and language abilities is essential for determining the most appropriate support. If indeed some individuals have more receptive language than is behaviourally apparent, this information could be transformative in terms of the way they are cared for, perceived, and interacted with. Accordingly, we urgently need to develop new tests of cognitive and language abilities that do not rely on behavioural responses.

In this thesis, I aimed to elicit and record neural signals that would indicate preserved language comprehension in children, and to provide a first application of my approach to a small group of minimally-verbal autistic children. I developed four paradigms, three using electroencephalography (EEG) and one using functional transcranial Doppler ultrasound, coupled with advanced analyses techniques, to observe neurotypical children's neural signals indicative of language comprehension. In the first and second experimental chapters, I looked for neural evidence of lexico-semantic processing in neurotypical children by presenting auditory words in different semantic contexts and recording their EEG. In the third experimental chapter, I looked for haemodynamic changes in response to task-following in children, indirectly reflecting their language comprehension, by asking them to follow spoken

[2] We use 'identify-first' language ('autistic person') rather than person-first language ('person with autism'), because it is the preferred term of autistic activists (e.g. Sinclair, 2013) and many autistic people and their families (Kenny et al., 2016) and is less associated with stigma (Gernsbacher, 2017).

instructions on two given tasks and by measuring brain metabolism using functional ultrasound. Finally, in the fourth experimental chapter, I applied the most successful paradigm in a case series of minimally-verbal autistic children.

In this introductory chapter, I begin by providing an overview of autism and the language profile of autistic individuals. I then review the research assessing cognition in non-communicative populations, including the development of neuroimaging and electrophysiological paradigms used in patients with disorders of consciousness, and those used specifically in autistic populations. Next, I present the methods that I used to collect and analyse neural data, and I conclude with an overview of this thesis.

1.1 Language and autism

1.1.1 Language profiles in autism

Autism is a neurodevelopmental condition, first described by the Austrian physician Kanner in 1943, in his report of 11 children who shared profound social and communication deficits, and unusual rigidity in their patterns of behaviour (Kanner, 1943). Following a better understanding of the wide diversity of symptoms and aetiology of autism, it is now referred to as autism spectrum disorder (ASD; hereafter, 'autism') in the Diagnostic and Statistical Manual of Mental Disorders, (5th ed.; DSM–5; American Psychiatric Association [APA], 2013). The DSM-5 describes autism according to difficulties in two core domains: (1) social communication and social interactions, typically seen as problems using nonverbal social cues, such as using eye contact, facial expressions, sustaining back-and-forth conversations, or understanding the pragmatics of a social situation, and (2) restrictive, repetitive and stereotyped patterns of behaviour, interests and activities, including hyper- or hyporeactivity to sensory aspects of the environment. Furthermore, the DSM-5 allows for the specification of whether these features co-occur with language and/or intellectual impairments.

Difficulties in verbal communication have been apparent since the first description of autism. All the children originally described by Kanner had atypical language that varied greatly in severity, from children only capable of meaningless speech or echolalia (immediate repetition of language heard by individuals), to children with no spoken language at all. Around the same time, the Austrian physician Asperger described a subgroup of children who likewise shared autistic traits in their social and ritualistic behaviours (Asperger, 1944). However, contrary to Kanner, Asperger's patients did not exhibit obvious language difficulties, with some even displaying above-average linguistic skills. Nevertheless, Asperger reported subtle atypicalities in his patients' verbal and non-verbal abilities, including with regard to syntax, verbosity, and prosody. Finally, Rutter (1978) proposed that language deficits were central to autism, based on earlier studies finding profound language impairments in most autistic children (Ornitz, 1973; Rutter, 1977; Rutter, 1970). From these early reports, the heterogeneity in the language abilities of autistic people was already evident.

Seventy years later, we now have a clearer view of the language difficulties of autistic individuals. Although the current edition of the DSM specifies language problems as a co-occurring condition instead of part of a diagnosis criterion, this has not always been the case. As listed in the previous edition of the DSM (4th ed.; DSM-4-TR; APA, 2000), at least one of the following criteria regarding language difficulties needed to be met for a classification of autism: (1) a delay or absence of spoken language, not compensated by other modes of communication such as gestures, (2) for individuals with adequate speech, an inability to initiate or sustain normal conversations, (3) the use of repetitive and stereotyped, or idiosyncratic language, and (4) an inability to engage in social imitative or pretend play appropriate to developmental level. As can be seen from these criteria, the degree of severity of language difficulties can be diverse, and we can see a distinction between different aspects of language: while criterion (1) reflects problems with overall language production, criteria (2) and (3) relate to difficulties using functional language in context, purportedly because of difficulties with theory of mind (Paul, Chawarska, Cicchetti, & Volkmar, 2008), and the last criterion reflects an inadequate use of symbolism and social language (Tager-Flusberg, Edelson, & Luyster, 2011). As noted earlier, the most recent version of the DSM excludes language impairments as a diagnostic criterion, and instead distributes those four aspects between the two core social and non-social domains. This shift reflects changes in our understanding of language in autism, with recent research showing that language impairment is neither specific nor universal to autism (Grzadzinski, Huerta, & Lord, 2013; Levy et al., 2010).

For those autistic individuals with language difficulties, there are three main ways of classifying the type of difficulties. The first is the commonly-accepted view of a dissociation between the form (syntax and phonology) and the function (pragmatics) of language in autism (see Wilkinson, 1998). This dissociation is reflected by relatively preserved syntactic and phonological aspects of language in autistic children, whereas the social aspects of language are usually affected (Tager-Flusberg, 1981; Tager-Flusberg et al., 2011). The second type of classification suggests that language can be dissociated between comprehension and production. Discrepancies between the language that autistic individuals produce and what they understand appear common. It is unclear whether production and/or comprehension are affected, and how this varies between individuals. For example, some autistic children may have a reduced lag between their acquisition of language comprehension and their language production, compared to neurotypical children (Davidson & Weismer, 2017, but see Kwok, Brown, Smyth, & Oram Cardy, 2015 for a meta-analysis suggesting otherwise).

Finally, Allen and Rapin (1980) suggested a three-category clinical classification of language disorders in preschool children with communication difficulties (Allen & Rapin, 1980, applied in Rapin & Dunn, 2003). Their model distinguishes between three levels of difficulties: (1) mixed receptive/expressive disorders, which corresponds to impaired early phonological decoding resulting in problems with all the subsequent steps of language processing; (2) higher order processing disorders, in which phonology is not affected but instead the comprehension and formulation of discourse is impaired; and (3) expressive disorders, in which comprehension is unaffected but the expression is impaired. This model separates phonological, semantic, and expressive aspects of language. The authors applied the model to a large cohort of unselected children with language disorders (as opposed to most studies specifically selecting verbal or non-verbal autistic individuals), classifying autistic children based on historical data of language abilities, and language production during a short play session. Rapin and Dunn found that none of the autistic children had comprehension within the ‘normal’ range, and that a majority (63%) had difficulties at the phonological and syntactic levels of language. Despite the importance of these findings, which contradict the view that receptive phonology is usually preserved in autistic individuals (Tager-Flusberg, 1981; Tager-Flusberg et al., 2011), the classification method Rapin and Dunn used was not well described, making it difficult to replicate. Furthermore, they did not describe how they assessed whether children who did not speak could understand language (classified in (2)), or not (classified in (3)). Thus, from their model it is difficult to clearly separate non-speaking children with comprehension problems from those with preserved comprehension who struggle to retrieve or articulate words. Yet, this distinction is of paramount importance. Understanding which level of the language stream is affected would help us understand the causes and evolution of language impairment and improve prognostic accuracy (Brignell et al., 2018; Gillespie-Lynch et al., 2012; Hofvander et al., 2009; Howlin, 2003; Howlin & Moss, 2012; Szatmari, Bryson, Boyle, Streiner, & Duku, 2003). It would also help in designing targeted and more personalised clinical interventions, and influence general care of these individuals.

About 30% of the autistic population is estimated to remain incapable of using spoken language after reaching school age. These children are typically referred to as 'non-verbal' or 'minimally-verbal', although there is no universally-accepted definition of these terms (Tager-Flusberg & Kasari, 2013). It is unclear how many of these children will remain non-verbal throughout their lives, and how many are merely pre-verbal, i.e. they will later develop language. It is also unclear which factors influence a child towards one of these two alternatives, though early and targeted interventions seem important (Kasari, Paparella, Freeman, & Jahromi, 2008). Thus, even within the minimally-verbal population, there is heterogeneity in cognitive and linguistic skills, and in the trajectory of language development. Additionally, two difficulties are often encountered when testing these individuals: (1) their inherent difficulties using spoken language necessarily prevents them from responding to instructions or answering questions typical in cognitive tests; and (2) standard testing procedures often involve a strict testing environment, requiring interactions with an experimenter, and involving social and pragmatic skills that this population has not mastered. As a result of these challenges, minimally-verbal autistic individuals remain largely under-studied (Tager-Flusberg & Kasari, 2013).

Two sources of evidence suggest potentially hidden cognitive and language abilities in minimally-verbal autistic individuals, sparking interest in the scientific community. The first one arises from reports of minimally-verbal autistic individuals who use augmentative and alternative communication (AAC) methods. AAC is a generic term referring to communication methods that supplement or replace speech (American Speech-Language-Hearing Association, 1989, p. 107), such as sign language, picture exchange, pointing to letters written on paper, or touching pictures on a screen. AAC is used in many conditions such as cerebral palsy, intellectual disability, and autism (Hourcade, Everhart Pilotte, West, & Parette, 2004; Millar, Light, & Schlosser, 2006). Two example AAC methods are facilitated communication (FC) and rapid prompting method (RPM). FC consists of an individual pointing to letters or pictures with the physical aid of a facilitator (i.e. another person), usually touching the individual's hand or arm (Biklen, Morton, Gold, Berrigan, & Swaminathan, 1992). RPM consists of an individual pointing to words or letters from multiple-choice options, with verbal, visual, or tactile aid from a facilitator in order to maintain their attention to the task (Chen, Yoder, Ganzel, Goodwin, & Belmonte, 2012). These methods are controversial due to the possibility of facilitator influence on the messages typed. In other words, because these methods involve a facilitator physically interacting with the individual or the material, they may consciously or unconsciously move the individuals' hand or letterboard. A comprehensive review of the controversy of these methods is outside the scope of this thesis, but I acknowledge a crucial lack of scientific evidence supporting their validity. In particular, two recent reviews point to the severe lack of evidence for both FC (Hemsley et al., 2018) and RPM (Schlosser et al., 2019). Furthermore, several national organisations have publicly denounced the use of one or both of these methods as valid alternative means of communication, pointing to the risk of facilitators voluntarily or unconsciously modifying the messages of disabled individuals (American Association on Intellectual and Developmental Disabilities, 2009; American Psychological Association, 2003; American Speech-Language-Hearing Association, 2008). On the other side of this debate are numerous families and autistic people advocating for the use of these methods. Soma Mukhopadhyay, the developer of RPM, has rallied hundreds of families to use her method, which is now widely used around the world (Paynter et al., 2017; Paynter & Keen, 2015). Anecdotal reports in the form of blogs or books, allegedly written by disabled individuals using RPM, suggest intact language abilities in a subset of users (Higashida, 2016; Kedar, 2012). If these reports are true, condemning RPM dramatically impacts such individuals' only mode of communication. Resolving the deep-seated divide between scientific evidence and advocate families requires the development of rigorous and reliable methods to assess language abilities in minimally-verbal individuals.

The second source of evidence of language abilities in minimally-verbal individuals is empirical. DiStefano, Senturk, and Spurling Jeste (2019) used EEG with a group of minimally-verbal autistic children. The group exhibited different neural responses to words paired with a matching picture from words paired with a mis-matching picture, which the authors took as electrophysiological evidence of intact semantic processing. Cantiani et al. (2016) found a similar pattern in five of their 10 participants, indicating semantic processing in a subset of minimally-verbal autistic children. However, a third study using similar methods did not find such evidence (Coderre et al., 2019). I discuss these electrophysiological results and the implications of such variable findings below. Taken together, the known heterogeneity in the autistic population, anecdotal reports from autistic individuals, and mixed results from the emerging empirical reports on this population, provide a strong rationale for the development of an objective neural test of language processing in minimally-verbal autistic individuals.

1.1.2 Testing minimally-verbal autistic children

Despite the challenges associated with testing minimally-verbal autistic individuals, recent efforts to include this population in research has led to the development of assessment guidelines. Kasari et al. (2013) summarized the available methods for assessing minimally-verbal individuals and emphasized the need to integrate standardized testing materials with child observation and parent report. They conclude that no single method is ideal to assess this population, and encourage the development of appropriate assessment methods. Another report by Plesa Skwerer et al. (2016) also emphasizes the heterogeneity of receptive language between individuals and across methods, and urge for developing individualised approaches to test these children. More recently, Tager-Flusberg et al. (2017) provided specific guidelines to acquire behavioural and neuroimaging data in these children, including breaking down tasks, using positive reinforcement to guide participants, and dealing with problematic behaviours.

These reports converge on the awareness that standard testing methods for cognitive and language abilities are often inadequate for the minimally-verbal population (Tager-Flusberg, 1999, 2000). Instead, in this thesis, I aimed to design a test using brain imaging that addresses the challenges associated with testing autistic children. I focused my research on two cognitive functions, language comprehension (chapters 2 and 3) and task-following (chapter 4). I addressed the following issues: First, contrary to most of the testing materials that evaluate cognitive skills, my methods minimised motor responses from children. Second, my methods contained few demands on individuals' social skills, such as direct interactions with, and requirements from, the experimenter. This reduced stress associated with testing and ensured that impaired social abilities did not confound the estimate of language. Third, I tailored the materials to be well tolerated by autistic children, namely by offering an alternative EEG system that is easier to setup on participants' heads. This more suitable system accounted for potential sensory atypicalities, especially hyper-sensitivity to touch, which is common in autistic people (Posar & Visconti, 2018; Tomchek & Dunn, 2007). Finally, I aimed to develop methods that were sensitive at the individual level, allowing us to screen individuals for preserved language comprehension, as opposed to the widely-used group studies where the data from several participants are averaged together and do not allow conclusions to be drawn for individuals.

1.2 Investigating Cognition in Non-Communicative Populations

1.2.1 Functional neuroimaging assessment of cognition

Recent years have seen the development of innovative brain imaging and electrophysiological approaches to assess covert cognitive abilities in non-communicative individuals. Most of these methods have been developed to assess patients with disorders of consciousness, such as those in a vegetative state. These patients are usually behaviourally under-responsive and show limited awareness of themselves or their environment. However, some may retain preserved cognitive abilities that are not revealed by standard clinical tests (Owen, 2008). This might lead to misdiagnosis, with patients having a more preserved mental life than their behaviour suggests. The use of functional neuroimaging in this difficult-to-assess population has provided evidence for intact cognitive functioning in some patients. Various levels of cognition have been studied using functional magnetic resonance imaging (fMRI), ranging from basic sensory perception to higher cognitive functions, including language comprehension, command following, and communication. Of interest to my PhD research are two functions I sought to assess in minimally-verbal autistic children, language comprehension and command-following. I now review the strategies developed to assess these functions in non-communicative individuals.

In a ground-breaking study, Owen et al. (2006) introduced an innovative method to detect residual cognitive function in patients in vegetative state. Using fMRI, Owen et al. measured neural activity of a patient while she performed different passive and covert tasks. They first confirmed that the patient had intact speech-specific and semantic processing neural activity. To do this, they presented spoken sentences in contrast to noise sequences, and semantically correct sentences in contrast to semantically anomalous sentences, respectively. They observed that her brain activation to speech and to semantic anomalies was similar to that of healthy adults, indicating preserved semantic processing. Because her intact responses to these two experiments could reflect autonomous speech and semantic processing, and not necessarily conscious awareness, they subsequently observed her neural responses to instructions to either imagine playing tennis (supported by supplementary motor areas) or imagine navigating inside her own house (supported by spatial navigation areas). Despite the patient's apparent unresponsiveness, Owen and his team observed voluntary modulation of her neural activity in response to these instructions, indicating wilful, conscious task-following, and ultimately pointing to preserved covert awareness.

These results have since been replicated and expanded to offer communication opportunities for these patients. Using Owen et al.'s paradigm, Monti et al. (2010) found that five patients with a disorder of consciousness could wilfully modulate their neural activity in response to instructions. Additionally, they introduced an avenue for communication by asking one patient who had reliable task-related responses to undergo a further experiment. They asked this patient, and healthy adults, to answer yes/no autobiographical questions by mentally performing one of the tasks (motor imagery or spatial imagery) for "yes" and the other task for "no". This paradigm allowed a blinded experimenter to determine the answer to all the questions in 16 healthy participants, as well as five out of six questions in the patient, simply by observing patterns of neural activity. In this particular case, this approach successfully allowed for communication in the absence of motor behaviours.

Recent work suggests that the aforementioned paradigms may be too complex, limiting the applications to patients who retain high levels of cognitive functioning but not patients with more impaired cognition. Bekinschtein et al. (2011) sought to use a simpler motor-preparation paradigm to assess voluntary movement planning. They first screened patients for intact basic auditory and speech processing by observing their electrophysiological responses to sound and speech with EEG. Then, they asked patients to wilfully move their right or left hand while they underwent an fMRI recording. They found that, despite the absence of observable movement, two out of five patients showed increased activation in the left premotor cortex when instructed to move their right hand, indicating movement preparation. As the paradigm used simple instructions, it potentially allows for the expansion of this line of research to patients with lower cognitive functioning who may retain voluntary motor planning and intact command following.

With a similar goal in mind, Naci et al. (2014, 2017) aimed to create a paradigm that avoided patients voluntarily following instructions while recording their fMRI response. Instead, Naci et al. designed a naturalistic paradigm that was intrinsically engaging: presenting participants with a suspenseful movie. They found a correlation between the timing of activation of brain regions that support executive function, and behavioural ratings of how suspenseful the movie was across time. From this, they concluded the existence of neural correlates of executive function that correspond to conscious experience of real-life events unfolding over time. This "neural code for consciousness" was shared across healthy adults, and was also found in a brain-injured patient, reflecting preserved consciousness despite an absence of behaviour.

This pioneering work is promising and particularly relevant in the context of detecting preserved language comprehension in individuals who cannot behaviourally communicate. However, it requires an MRI scanner, which is expensive with limited availability, and a high degree of compliance from participants, who need to lie still for long periods with loud noises from the scanner. For these reasons, this method is unsuitable to test all populations, especially autistic children. Instead, it is necessary to turn to portable and affordable methods to assess this population's cognitive abilities. EEG is one such solution. It records the brain's electrical activity over time, allowing the precise assessment of neural responses to specific stimuli. It is far less expensive and more available than MRI, and is portable, making it a viable option both for clinical assessment of cognition, and for testing outside the laboratory, such as at home or in schools.

1.2.2 Electrophysiological assessment of cognition

Two approaches to cognitive assessment using EEG have emerged in recent years. The first one is based on analysing changes in the EEG signal in response to different stimuli or task instructions. The second one is based on the real-time decoding of these signals to implement brain-computer interfaces (BCI). Considering the first, researchers have adapted the aforementioned fMRI command-following paradigms for EEG, and observed changes in electrical oscillations in response to task-following. Cruse et al. (2011, 2012) used a motor-imagery paradigm in sixteen vegetative patients. They instructed patients to imagine moving different body parts and analysed the evoked EEG oscillations in response to the different instructions. They observed successful modulation of oscillatory activity in four out of 16 patients, indicating preserved task-following. Other researchers have used event-related potential (ERP) paradigms. ERPs are changes in the brain's electrical activity in response to events such as the presentation of stimuli. They are extracted by averaging the EEG response time-locked to the stimulus presentation, and usually compared across different conditions (Luck, 2014). ERPs typically present different peaks over time, which are commonly named based on their polarity (negative or positive), and their timing, and relate to specific perceptual or cognitive events. For instance, the P300 is a positive deflection of the ERP waveform occurring around 300ms after the presentation of a stimulus, and is sensitive to participants' attention to the stimulus. To give on example, Schnakers et al. (2008, 2009) presented patients with a series of names including their own first name, and asked patients to listen passively or count the occurrence of their name. Researchers observed a larger P300 ERP response to the patient's own name compared to other names, and this response was largest when participants actively counted their name, relative to passive listening. This pattern indicates command-following and thus preserved voluntary brain functioning.

The second approach to cognitive assessment using EEG is BCI. The use of BCI has emerged as an approach to facilitate communication in the absence of spoken language. BCIs translate the intentions of an individual into computer control signals without requiring motor control. More specifically, they translate the spontaneous or evoked neural signals into control commands that can be externally observed, such as letters displayed on an on-screen keyboard (Moghimi, Kushki, Marie Guerguerian, & Chau, 2013). Various neural signals have been used for BCI, including spontaneous activity, and two types of evoked activities: the P300 ERP and the steady-state, visually-evoked potential (SSVEP). The P300 approach has been used to decode which letter a participant is thinking of, by presenting an array of letters, with flashes highlighting different letters sequentially. A larger P300 is observed when the letter chosen and attended by the participant is highlighted (Farwell & Donchin, 1988). This method has allowed paralysed patients to initiate communication, albeit at a slow rate of less than six letters per minute (Donchin, Spencer, & Wijesinghe, 2000; Nijboer et al., 2008; Serby, Yom-Tov, & Inbar, 2005). The SSVEP approach relies on the neural response to presentations of visual stimuli flickering at high frequencies. It is possible to record oscillations in the occipital lobe corresponding to the presentation frequency. By presenting an array of letters, each flickering at a different frequency, it is possible to determine which letter the participant fixates on (Lee et al., 2010; Wang, Gao, Hong, Jia, & Gao, 2008; Wang, Wang, Gao, Hong, & Gao, 2006).

The approaches mentioned above (ERP analyses and BCIs) come with challenges when we consider their utility for minimally-verbal autistic individuals. Motor-imagery paradigms require imagined physical movements, making them well suited to paralysed patients. However, in the case of minimally-verbal autistic children, this type of paradigm is unsuitable for several reasons. First, it requires participants to remain perfectly still, as actual motor movements would contaminate the EEG signal. Second, in addition to understanding the instructions, participants must imagine their body moving, which might in itself be difficult, for example because this population can experience altered proprioceptive responses (Masterton & Biederman, 1983; Paton, Hohwy, & Enticott, 2012). Finally, these children may have difficulties with motor control (Hughes, 1996; Ming, Brimacombe, & Wagner, 2007), and it is not clear to what extent this affects movement imagery. BCIs, in their current form, might also be unsuitable. Some promising results using the P300 in BCI settings in healthy populations suggest possible communication through a visually-presented keyboard, without overt responses (Cheng, Gao, Gao, & Xu, 2002; Serby et al., 2005). However, its reliability in disabled individuals is yet to be demonstrated, and it has not been used in autistic populations. This is likely due to the long training sometimes required for BCI pattern recognition. Additionally, commonly used evoked responses, such as the P300 or SSVEP, are typically elicited in paradigms with flickering visual stimuli, which might be inappropriate for autistic populations considering their altered visual processing (Behrmann, Thomas, & Humphreys, 2006) and higher risk of seizures (Bolton et al., 2011). Thus, as an additional challenge to assess autistic individuals, we must consider the associated sensory and behavioural atypicalities, including difficulties complying with task demands and remaining still. Due to these challenges, it is necessary to depart from paradigms involving motor imagery, motor demands, or flickering visual stimuli, and turn to implicit methods that minimise cognitive and perceptual demands.

Two implicit methods that have been used thus far in autistic children, eye-tracking and EEG, introduce promising avenues for passive investigation of cognition. Eye-tracking can measure the movements of participants' eyes and the changes in pupil diameter (Duchowski, 2007). It has been used in pre-verbal infants and autistic children using preferential looking paradigms (Brady, Anderson, Hahn, Obermeier, & Kapa, 2014; Gredebäck, Johnson, & von Hofsten, 2009; Piotroski & Naigles, 2012). In these paradigms, an auditory word is presented, followed by two visual images, one of them matching the auditory word. If children know the word, they typically look longer at the corresponding picture, yielding an estimate of their vocabulary knowledge. In order to avoid reliance on visual processing skills I did not explore the eye tracking avenue in this thesis, but I acknowledge that it may be relevant for future work. Of most relevance to this thesis are the extant studies using EEG in the general autistic population and the few published studies in minimally-verbal autistic individuals, which I review below. In particular, I am interested to measure whether words presented in semantically congruent or incongruent contexts would elicit distinguishable neural signals, which would indicate lexico-semantic processing.

1.2.3 The N400 as a marker of language comprehension

1.2.3.1 Introduction to the N400

The N400 is a well-studied ERP component, first discovered in 1980 by recording the EEG activity of adults while they read sentences with either congruent or incongruent endings (e.g. "he took his coffee with cream and *sugar*" versus "he took his coffee with cream and *dog*"). Kutas and Hillyard (1980) coined the term N400 when they found a negative potential peaking around 400ms after the onset of the final word. This potential was more negative in response to incongruent than congruent words, a differenced coined the N400 *effect*. In the following years, many studies have characterised the stimulus properties that influence this component. Studies indicate that the N400 is unaffected by anomalies in physical properties of stimuli (e.g. by the contrast of the two written sentences "he took his coffee with cream and SUGAR" and "he took his coffee with cream and sugar"). It is also unaffected by grammatical anomalies, such as a missing plural "s" in "all dogs have four leg" (Hillyard & Kutas, 1983; Kutas, Van Petten, Ackles, Jennings, & Coles, 1988). Instead, the congruency of the semantic and/or predictive context is key to eliciting the effect. Moreover, the context does not need to be a full sentence, but a single probe word presented before a related or unrelated word is enough to elicit a larger N400 in the unrelated condition (e.g. a larger N400 is elicited by "nest" when it follows "arm" compared to when it follows "bird" - Kutas et al., 1988). The N400 is amodal, manifesting in written, auditory, or American Sign Language presentations (Holcomb, Coffey, & Neville, 1992, reviewed in Kutas et al., 1994). Finally, the N400 can be elicited by non-linguistic stimuli, such as in response to line drawings that complete written sentences in matched versus mismatched conditions (Ganis, Kutas, & Sereno, 1996), or in response to similar versus different faces (Caldara, Jermann, Arango, & Van der Linden, 2004).

As evidenced by the variety of stimuli that elicit N400 effects, the N400 does not merely reflect violations of linguistic semantic context. Instead, there is ongoing debate on which cognitive processes the N400 truly reflects (Kutas & Federmeier, 2011). For instance, it is unclear to what extent the N400 is modulated by attention (Deacon & Shelley-Tremblay, 2000), or to which step of language comprehension the N400 relates. Specifically, a common assumption is that neural processing of words starts with perceptual recognition and ends with lexical processing, which allows the access to word recognition and semantics. While some authors argue that the N400 occurs early in this stream of processes (Deacon, Dynowska, Ritter & Grose-Fifer, 2004), others suggest it to be later, at the level where the word's semantic information is linked to the semantic context held in working memory (Brown & Hagoort, 1993; Hagoort, Baggio, & Willems, 2009). Notwithstanding this ongoing debate, I hereby present its utility as a tool for studying vocabulary and language comprehension through manipulating the semantic context of linguistic stimuli. By comparing neural responses to identical words presented in different semantic contexts, we can test for evidence of lexico-semantic processing of linguistic stimuli.

Using this logic, the N400 has been used to test vocabulary and semantic processing in populations that are unable to communicate behaviourally (reviewed in Münte, Urbach, Duzel, & Kutas, 2000). In a seminal study, Byrne, Dywan, and Connolly (1995) utilised the N400 to test vocabulary in a 17-year-old boy with cerebral palsy who was unable to communicate. They recorded the N400 to auditory presentations of words that were congruent or incongruent with a previously presented picture. The stimuli were also presented to three age-matched, neurotypical participants. The words were selected to be known (within the participants' age range) or unknown (beyond the participants' age range). For all four participants the known words elicited N400 effects, indicating semantic processing. For the unknown words, this pattern was attenuated for one control, and reversed for the other three participants. The contrast thus demonstrates the N400 to be sensitive to vocabulary. Moreover, the results came in the absence of overt responses, making it a promising avenue for testing minimally-verbal individuals. Although encouraging, this study was based on only four participants, and there was no account for multiple comparisons across electrodes in the statistical analyses. Nonetheless, this approach has since been used successfully for individuals from various populations, such as brain-injury-induced aphasia (Connolly, Mate-Kole, & Joyce, 1999), stroke (D'Arcy et al., 2003; Wilson et al., 2012), and in more cases of cerebral palsy (reviewed in Geytenbeek et al., 2010).

Two outstanding questions remain to be addressed in pursuing this approach as a neural test of language comprehension in minimally-verbal autistic children. The first is the sensitivity of the N400 effect at the individual subject, as opposed to group, level. The second is the sensitivity of N400 effects in autistic populations.

1.2.3.2 Individual reliability of the N400

With the exception of a few experiments, most of the studies mentioned above examined the N400 effect in groups; that is, by averaging the results from multiple individuals to obtain a grand average across participants. This approach increases the signal of interest by averaging out the noise across individuals' data. Yet, to develop an individualised test, it is crucial to examine the reliability of these neural signals in individuals. To my knowledge, only a few studies have systematically assessed the presence of N400 effects in individuals (Beukema et al., 2016; Cruse et al., 2011; Kiang, Patriciu, Roy, Christensen, & Zipursky, 2013; Rohaut et al., 2015). Cruse et al. reported N400 effects in neurotypical adults in response to different auditory paradigms and task instructions. Despite finding a robust N400 at the group level, only a subset of the participants showed a reliable N400 effect at the individual level. The number of participants showing an effect – hereafter referred to as the "detection rate" – was affected by stimulus type and task instructions. Specifically, their detection rate was highest when contrasting normatively associated word pairs (i.e. pairs in which the second word was strongly predicted by the first word) and unassociated word pairs (6/12 participants), and lower when contrasting congruent and incongruent sentences (2/12 participants). Furthermore, for their normatively-associated versus unassociated pairs, they also found an effect of task demands, with their overt task (i.e. indicate whether each word-pair is related or unrelated with a button press) yielding a higher detection rate (9/12 participants) than their covert (i.e. decide in your head whether each word pair is related or unrelated - 7/12 participants) or passive task (i.e. listen to the sentences - 0/12 participants). These results suggest high inter-individual variability in the presence of N400 effects. Cruse et al. did not report whether these neural events also varied in timing and topography between individuals. Using similar approaches, Beukema et al. (2016) and Rohaut et al. (2015) report comparable results, with statistically significant N400 effects recorded in 88% and 58% of neurotypical adults, respectively. These scarce results point to heterogeneity in neural response in the healthy adult population, and our lack of understanding of inter-individual variability in neural responses.

In addition to limited reporting of the presence of N400 in individuals, a growing number of studies report large variability in N400 latency and strength across individuals (reviewed in Tanner, Goldshtein, & Weissman, 2018). Kim, Oines, and Miyake (2018) presented adults with written sentences that were either correct or grammatically and semantically incorrect. The incorrect sentences elicited both an N400 and a P600. The P600 is an ERP component linked with syntactic and structural processing. The authors found a negative correlation between the size of these two components, that is, the smaller the N400, the bigger the P600. This suggests that different participants may have relied on different cognitive processes to analyse the semantic content of sentences. Moreover, verbal working memory correlated negatively with the amplitude of the N400, potentially reflecting individual differences in processing semantic information. Other studies found a correlation between working memory and the size and topography (localisation of the effect on the scalp) of the N400 effect in adults (Boudewyn, Long, & Swaab, 2013). Moreover, variations have been found between populations, such as a smaller and/or delayed N400 in subgroups of schizophrenia patients (Adams et al., 1993; Grillon, Ameli, & Glazer, 1991; Kostova, Passerieux, Laurent, & Hardy-Baylé, 2005). Overall, these studies suggest that despite strong N400 effects in groups of participants, differences exist between individuals. We expect these variations to be, if anything, even more pronounced in children whose brain development may vary across age. Thus, one aim of this thesis is to rigorously investigate the idiosyncratic neural responses to semantic manipulations in individual neurotypical children.

1.2.3.3 The N400 in autism

The second necessary step before using the N400 as a marker of lexico-semantic processing in autism is to review the current knowledge on the N400 in autistic individuals. Most studies have been conducted on autistic individuals with language and IQ within the typical range. Contrasting results emerge, with studies finding either similar, smaller, or absent N400 in autistic individuals compared to neurotypical controls. McCleery et al. (2010) used a picture-word paradigm in which participants saw a picture and then heard a related or unrelated spoken word or sound (e.g. a picture of a ball followed by the word “ball” or “car” or by the sound of a ball bouncing or of a car engine). They found an N400 effect in autistic children in response to sounds but not to words, pointing to preserved semantic processing in the non-verbal relative to the verbal domain. Similarly, contrasting EEG responses to pictures related or unrelated to heard sentences, Ribeiro et al. (2013) found N400 effects in neurotypical children but not in autistic children, indicating a lack of sensitivity for lexico-semantic processing in this group. Similarly, Pijnacker et al. (2010) found no N400 effects in a group of cognitively-able autistic adults in response to written congruent and incongruent sentences.

However, two of these studies (Pijnacker et al., 2010; Ribeiro et al., 2013) found a late positive component in the autistic group, starting around 700 ms post-stimulus onset, and which differentiated the conditions. This component may reflect delayed memory-retrieval of semantic information (Coulson & Van Petten, 2002). Finally, one study found smaller but not absent N400 effects in autistic adults in response to congruent versus incongruent spoken sentences (Fishman, Yam, Bellugi, Lincoln, & Mills, 2011). In contrast, several studies have found significant N400 effects in autistic adults (Braeutigam, Swithenby, & Bailey, 2008; Coderre, Chernenok, Gordon, & Ledoux, 2017; Russo, Mottron, Burack, & Jemel, 2012), albeit sometimes occurring at unexpected timings and locations, such as delayed onset (Méndez, Sans, Abril, & Valdizan, 2009), or more right-lateralised (Coderre et al., 2017). These mixed findings seem likely to reflect large heterogeneity in autism and call for the need to develop individualised neural tests.

To date, three studies have reported the electrophysiological markers of lexico-semantic in minimally-verbal autism. First, Cantiani et al. (2016) presented pictures followed by their corresponding spoken-word label or a mismatched label (i.e. a picture-word paradigm) to ten minimally-verbal autistic children and ten neurotypical children. There were two key findings. First, the autistic group showed basic auditory and visual responses that were similar in amplitude, but delayed, compared to the neurotypical group, suggesting delayed basic sensory processing in autism. Second, contrary to neurotypical children, the autistic group did not show a significant N400 effect. Yet, by further examining individual participants' data, Cantiani et al. report that eight out of ten neurotypical and five out of ten autistic children showed N400 effects. From this, the authors conclude that at least a subset of the autistic participants had intact lexico-semantic processing. The authors suggest that averaging all participants in the grand-average obscured participants that did show an N400 effect, and recommend examining individual ERPs. One limitation of this study is the choice of analysing the N400 effect only in one scalp region (around the vertex), potentially missing effects occurring at other locations.

More recently, DiStefano et al. (2019) used a similar picture-word paradigm to compare the EEG responses of verbal autistic, minimally-verbal autistic, and neurotypical children. Contrary to Cantiani et al., DiStefano et al. found significant N400 effects in all groups, albeit at a later latency in the minimally-verbal autistic group. Furthermore, a cluster analysis revealed subgroups of autistic children, with one cluster of children showing reversed but marked N400-like effects (i.e. larger N400 in response to related compared to unrelated words). Unfortunately, the authors did not provide individual participants' waveforms or individual statistical analyses, making it difficult to assess the individual-subject level results. Finally, Coderre et al. (2019) used three implicit measures of vocabulary knowledge, eye movements, pupillometry, and ERPs in a picture-word paradigm, to assess five severely autistic adults (according to their DSM-5 classification), including two minimally-verbal participants. They found high variability between methods and participants, with different measures being sensitive in different individuals. The minimally-verbal participants did not show evidence of vocabulary knowledge with any method. However, this study did not include a group of neurotypical adults, making it difficult to determine whether their methods were reliable in individual neurotypical adults. Their small sample size (N = 5) also makes it difficult to draw conclusions on the minimally-verbal population.

A limitation of the three EEG studies on minimally-verbal individuals mentioned above is their reliance on intact visual attention and processing, for the processing of picture stimuli. Yet, abnormalities in visual attention have been found in autistic individuals (Behrmann et al., 2006; Townsend, Harris, & Courchesne, 1996), which make visual paradigms sub-optimal for this population. For this reason, the first paradigm that I developed relied on auditory processing alone.

Overall, previous results obtained using EEG in the minimally-verbal population are scant and highly variable, potentially reflecting the large heterogeneity in autism. Moreover, no study has provided detailed individual results of atypical or neurotypical children. Thus, one of the objectives of my PhD was to determine the inter-individual reliability of neural markers of lexico-semantic processing in neurotypical children. Once this is established, we can determine suitability for application to minimally-verbal populations.

1.3 Overcoming methodological and conceptual limitations

In addition to the lack of rigorous analyses of individual data, the development of N400 paradigms as a neural marker of language processing has been hampered by three limitations that I set out to address in this thesis. Two are methodological concerns, while the third is conceptual.

1.3.1 An alternative solution to research-grade EEG systems setup: Emotiv EPOC+

The first methodological concern arising from the empirical work on EEG relates to the type of system used. Given that hypersensitivity is common in autism (Posar & Visconti, 2018; Tomchek & Dunn, 2007), research-grade EEG systems may be hard to tolerate. Research-grade EEG setups involve skin preparation and applying gel on the scalp, as well as placing a cap on the head and securing it under the chin. Even though habituating and desensitizing participants to the cap has proven effective to improve testing compliance, easier-to-setup EEG systems must be considered. In my first study (Chapter 2), I introduce Emotiv's EPOC+ system as such an alternative. EPOC+ consists of a lightweight headset with 14 flexible plastic arms holding electrodes that connect to the scalp with saline-soaked cotton pieces. In addition to being fast to setup, this system is portable, lending itself to use beyond the laboratory, such as homes or schools. EPOC (the previous version of EPOC+) has been validated for recording ERPs by simultaneous recordings with a research-grade system. High correlations between EPOC and traditional EEG systems have been found for auditory ERPs (Badcock et al., 2013, 2015;

Barham et al., 2017), the face-sensitive N170 (de Lissa, Sörensen, Badcock, Thie, & McArthur, 2015), and the P300 in some cases (Duvinage et al., 2013, but see Vos, Kroesen, Emkes, & Debener, 2014). For my purpose of developing neural tests of language comprehension in autistic children, I set out to validate the EPOC+ for recording N400s as a portable and easy-to-tolerate EEG system.

1.3.2 Improved analysis of individual EEG data: multivariate decoding

Second, all the N400 work presented above reports univariate analyses of ERP waveforms. Univariate analyses consist of comparing the brain's electrical activation over time between two conditions of interest. To do this, the EEG data from all trials is averaged and compared between conditions. For example, the N400 effect reflects the comparison between the brain's electrical activity in response to congruent versus incongruent words, and this comparison is made at each time point or by averaging over time points, in a pre-determined time-window or epoch. These comparisons can be made for each electrode individually or in regions comprising several electrodes averaged together. Univariate analysis is the most common way to examine EEG data but it presents some disadvantages for our purposes. First, the conditions are compared over multiple time points and across multiple electrodes, usually with t-tests. Such a large number of statistical tests necessitates correction for type I error (i.e. incorrectly concluding that a significant difference exists between conditions, which is set at 5%, by scientific convention). This means that either unfeasibly large numbers of participants (or trials) are needed or only large effects can be detected. This issue is usually circumvented by *a priori* selecting electrodes (or clusters of electrodes) and time-windows of interest, and only testing for differences in these regions, to reduce the number of comparisons. However, this approach overlooks differences at other electrodes and/or time points, potentially missing important discriminative signals. Due to the intrinsic noise in individuals' data, and the potential variability between individuals, it is essential to retain as much statistical power as possible, while still accounting for unexpected locations and timing of effects. For these reasons, univariate analyses may be sub-optimal for analysing individual data. Instead, in the empirical work of this thesis, I also use multivariate pattern analyses (MVPA) on my EEG data (Chapters 2, 3, and 5).

MVPA, also called *classification* or *decoding* analysis, is a machine-learning technique that has been adopted in the last two decades for analysis of neuroimaging data. Although it has mainly been used for analysis of fine spatial patterns in fMRI data (e.g. Hanson & Bunzl, 2010; Haxby, 2012; Haxby et al., 2001), it has more recently been applied to timeseries data such as EEG or magnetoencephalography (MEG; Carlson, Grootswagers, & Robinson, 2019; Carlson, Hogendoorn, Kanai, Mesik, & Turret, 2011; Schaefer, Farquhar, Blokland, Sadakata, & Desain, 2011). MVPA consists of training a computer algorithm to extract reliable differences in the pattern of neural activity between conditions. In the context of electrophysiological data, the EEG data are split so that some proportion of trials constitute a training set and the remaining are “held out” as a testing set. Then, a machine-learning algorithm, such as a *linear support vector machine*, is presented with the training data recorded at one time point, from all channels. These data have dimensionality equal to the number of EEG channels, and the classifier is trained to find a decision boundary in the multidimensional data-space that best distinguishes between the conditions. After this, we present the classifier with the “held out” data, i.e. the testing set, and record its classification of these trials. We then compute the proportion of test data correctly classified, i.e. that fell on the correct side of the decision boundary. For example, in my case I present spoken words in two different conditions: congruent and incongruent (Figure 1A). I then measure the EEG response in all channels (Figure 1B) and train the classifier on the EEG data from all trials but two, holding out one trial from each condition. The classifier finds the boundary that best separates the two conditions in the training data (Figure 1C). I then present the classifier with the two untrained trials and observe whether the trials are correctly classified as congruent or incongruent according to which side of the boundary they fall, that is, the classifier accuracy (Figure 1D). I can then repeat this process, by withholding a different trial for each repetition. Finally, I average the accuracy values from the different iterations to derive a single classification accuracy for this time point. If the average accuracy is significantly above chance (50% in a two-condition experiment), I conclude that there is information in the pattern of neural activity that differs between conditions. I can then repeat this process across time, over all the time points, which gives a profile of when the brain makes a distinction between the conditions (Figure 1E).

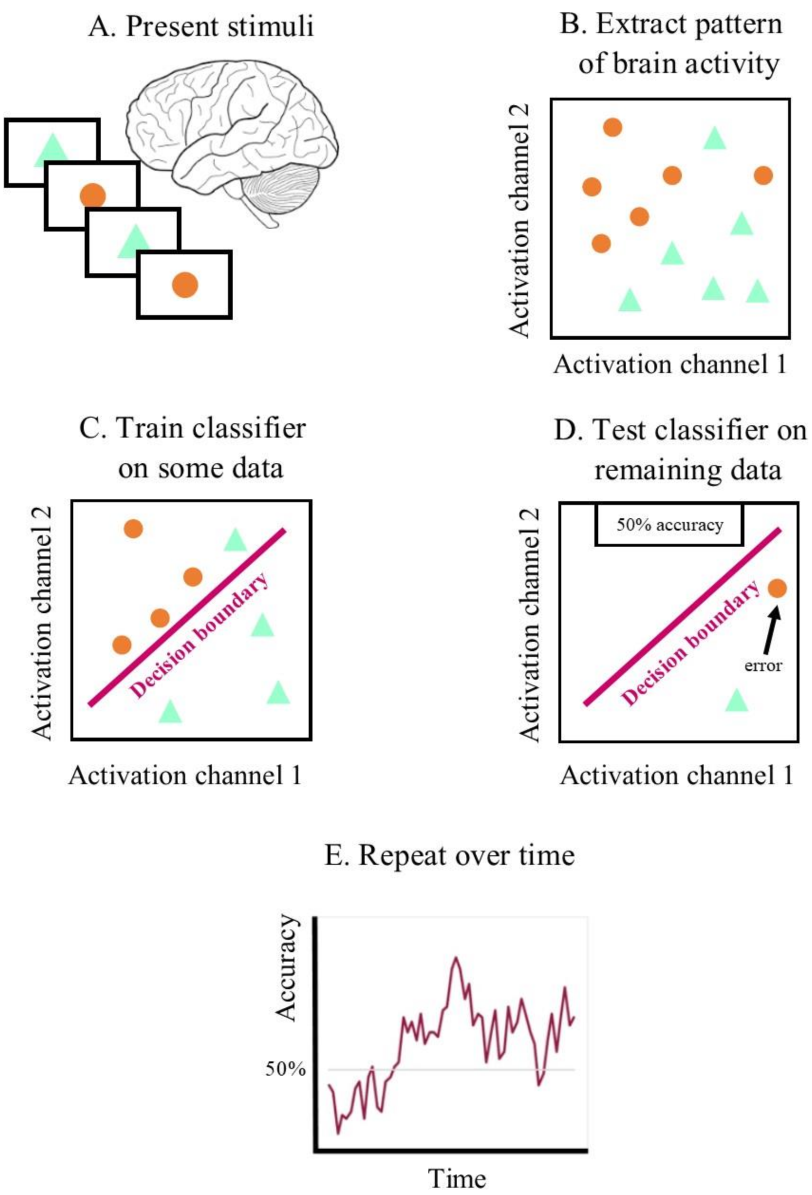


Figure 1. MVPA decoding for EEG data. A. First, I present stimuli to participants and record the brain data for each condition (here orange circles represent congruent words, and green triangles represent incongruent words). B. I record and extract the pattern of neural activity across EEG channels. Note that experimental data are typically recorded from many channels, but for illustration purposes, only two channels are presented here. C. A machine-learning classifier is then trained to find a decision boundary that best distinguishes between the conditions of interest, based on a subset of the data. D. The classifier is presented with unknown data and the accuracy is computed (i.e. the proportion of trials that fell on the correct side of the decision boundary). E. This decoding process is repeated at every time point to create a profile of accuracy over time. For a two condition experiment, chance is 50%.

To validate that decoding above 50% does not occur by chance, I implement a *permutation test* of statistical significance. To do this, I first randomly re-label the EEG data with the two labels ("condition 1" and "condition 2"). This altered dataset should not contain any information that discriminates between conditions. I then proceed through the classification described above for this permuted dataset and record the resulting classification accuracy. I repeat this procedure multiple times (usually 10,000), with different random permutations each time, to obtain a null distribution of classifier accuracies. Actual decoding is considered significant if it is higher than 95% of the random permutations (i.e. less than 5% chance of finding the actual decoding accuracy by chance, $\alpha = .05$). We can perform this analysis at every time point, to estimate when significant differences between conditions occur.

When performing this analysis at each time point, it is still necessary to account for multiple comparisons made across time. However, MVPA takes into account the pattern of neural activation combined across all electrodes, dramatically reducing the number of multiple comparisons relative to performing univariate analyses at all sensors. Additionally, by examining patterns of neural activity, MVPA is not restricted to detecting any particular unidirectional change in voltage between conditions (e.g. voltage in condition 1 > voltage in condition 2), and does not require that changes are in the same direction in different channels. Instead, more subtle combinations of voltage changes across space, which may remain undetected with univariate analyses, may give insight into the discriminative neural signals between conditions. Moreover, in Chapter 3, I use an even more statistically powerful version of this analysis, which uses all the data from all time points in a single MVPA analysis. This approach reduces the total number of statistical comparisons for each person to one and thus removes the need for multiple-comparisons correction (as done in Geuze, Gerven, Farquhar, & Desain, 2013; Tanaka, Watanabe, Maki, Sakriani, & Nakamura, 2019). As the empirical data of Chapters 2 and 3 show, MVPA allows us to increase our sensitivity to detect individual differences in neural processes between lexico-semantic conditions.

1.3.3 Assessing language comprehension through task-following

Finally, there is a conceptual limitation of using the N400 to index language comprehension. The N400 provides strong evidence for lexico-semantic processing, and therefore can give indications on language comprehension. However, it is possible that some lexico-semantic processing will occur in the absence of language comprehension. For example, in the case of the picture-word paradigms described above, it is possible that the N400 merely reflects vocabulary knowledge, and may not inform sentence-level comprehension. For this reason, paradigms contrasting congruent and incongruent sentences, like the ones presented in Chapters 2 and 3, are stronger evidence for holistic comprehension of sentences. If we observe N400 effects in response to congruent and incongruent sentences in minimally-verbal autistic children, this would be evidence that they understood the meaning of the sentences sufficiently well enough to process the final word differently depending on its fit, making it a promising marker of semantic comprehension. However, it remains possible that N400 effects reflect the association of key words in the sentence (e.g. “birthday” and “cake”) rather than true sentence-level comprehension. In turn, it is possible that neurophysiological indicators could reflect higher co-occurrences of the keywords in natural situations (“birthday” and “cake” are heard together more often than “birthday” and “door”) rather than deep semantic processing of those items.

Stronger evidence may be found if participants follow a verbal command – they would have to understand the command in order to follow it. Therefore, in Chapter 4, I explored the possibility of using neuroimaging to assess command-following in children. For this work, I used another accessible neuroimaging device, transcranial functional Doppler ultrasound (fTCD). I measured the hemispheric haemodynamic activity of neurotypical children while they performed one of two mental tasks that typically activate the left or the right hemisphere respectively. The rationale for this study is that the detection of changes in hemispheric activity that reliably co-occur with children performing one or the other mental task, infers preserved task-following abilities, thus preserved comprehension of task instructions.

FTCD allows for the continuous recording of blood flow velocity in the two middle cerebral arteries, through ultrasound emission and reception by two probes placed on each temple (Willie et al., 2011). It thus allows for the real-time, non-invasive measurement of global changes in the brain's vascular response. The middle cerebral arteries are the main source of irrigation for most of the frontal, lateral, and temporal areas of the brain (Navarro-Orozco & Sánchez-Manso, 2019). By designing tasks that rely more on one hemisphere over the other, it is possible to observe increased haemodynamic response in that hemisphere, seen with fTCD as an increase in blood flow velocity. This logic has been used in several studies seeking to assess the reliability of fTCD for estimating lateralisation of brain functions without requiring invasive or expensive tests used previously (Loring & Meador, 2000). In these fTCD studies, participants are typically asked to perform either a language task, such as silently generating words that start with a given letter, or a visuospatial memory task, such as remembering the location of visual stimuli. Results from these studies are consistent with reports using fMRI, with predominant left lateralisation of language function in groups of adults (Bishop, Watt, & Papadatou-Pastou, 2009; Groen, Whitehouse, Badcock, & Bishop, 2012; Rosch, Bishop, & Badcock, 2012; Whitehouse & Bishop, 2009) and children (Groen et al., 2012; Haag et al., 2010; Lohmann, Dräger, Müller-Ehrenberg, Deppe, & Knecht, 2005). Moreover, studies have found predominantly right lateralisation of visuospatial memory in groups of adults and children (Groen, Whitehouse, Badcock, & Bishop, 2011; Groen et al., 2012; Rosch et al., 2012). Despite these group results confirming lateralisation of these neural functions, individual results are more heterogenous, with some individuals having opposite or mixed patterns of lateralisation (Badzakova-Trajkov, Corballis, & Häberling, 2016; Whitehouse & Bishop, 2009). Previous research suggests that lateralisation in autistic individuals may be atypical (Lindell, Notice, & Withers, 2009; Preslar, Kushner, Marino, & Pearce, 2014; Yoshimura et al., 2013). However, I argue that lateralisation *per se* is not crucial to answer the question of language comprehension. Indeed, an individual may have both language and visuospatial memory lateralised to the left side, but still have different patterns of hemispheric activation in the two tasks (e.g. differences in the extent of lateralisation or time course of activation). Thus, in my study, I examined the difference in the left and right patterns of activation between tasks across time, instead of reporting only lateralisation *per se*. Furthermore, the visual stimuli presented in both tasks were strictly identical so that the conditions differed only in the auditory instruction and the task to perform. As such, any difference in the patterns would indicate different neural processes engaged in the tasks, evidence for preserved task-instruction comprehension.

1.4 Thesis overview

In this thesis, I present the results from four paradigms developed to objectively assess language comprehension in individual children and a small case series using one of these paradigms in minimally-verbal children. In Chapters 2 and 3, I used EEG to measure lexico-semantic processing in neurotypical children by measuring the electrophysiological response to semantic manipulations of spoken language. In Chapter 4, I used fTCD to indirectly assess language comprehension in neurotypical children by observing the brain’s haemodynamic response during task-following. Because the paradigm used for the second experiment in Chapter 2 yielded the highest detection rate in individuals, in Chapter 5, I applied the Chapter 2 EEG to assess language comprehension in three minimally-verbal autistic children. The overarching goal of these studies was to design an implicit and objective neural test of language comprehension for populations who are unable to reliably communicate and for whom comprehension abilities are unknown. In the following section, I present a brief overview of the empirical chapters of this thesis.

1.4.1 Chapter 2 - Towards an individualised neural assessment of receptive language in children

In Chapter 2, I used two EEG systems to measure the brain's electrical activity of neurotypical children in response to manipulations of lexico-semantic context. I presented identical spoken words in two semantic conditions, congruent versus incongruent, and measure the ERP. For this I developed two auditory paradigms, one contrasting related and unrelated word pairs (e.g. "cat – dog" versus "arm – dog"), and one contrasting congruent and incongruent sentences (e.g. "she wore a necklace around her neck" – "she wore a necklace around her milk"). Typical ERP research reports group results, not concerned with inter-individual variability. Yet, for potential clinical applications of a neural test of language comprehension, it is important to establish the reliability of neural signals in individuals. Thus, the first goal of this study was to assess the heterogeneity of neural signals in individual neurotypical children. Second, due to the behavioural and sensorial difficulties of autistic children, easy-to-tolerate recording systems are preferred. Thus, the second goal was to assess the validity of a portable EEG system (Emotiv EPOC+) as an alternative to research-grade systems. Finally, the third goal was to address the issue of inter-individual variability in neural signals by analysing the data with two different analyses: univariate N400 analyses and timeseries MVPA. Timeseries MVPA analyses patterns of activity from all the electrodes at each time point and accounts for variation in the location of neural responses, which is critical when dealing with individual's data. This study established the heterogeneity of neural signals in the neurotypical population, with variations in topology, time course, and strength of neural response to lexico-semantic manipulations. Despite these variations, the sentences paradigm yielded the highest reliability (88% of individuals), when analysing research-grade EEG system's data with MVPA.

1.4.2 Chapter 3 - Even neurotypical children are heterogeneous: better individual subject analysis of EEG data with temporally and spatially unconstrained multivariate decoding

In Chapter 3, I further evaluated the heterogeneity in neurotypical children's neural signals in response to lexico-semantic manipulations. I also assessed whether a more engaging paradigm may yield a better detection rate. To this end, I designed a new auditory EEG paradigm contrasting congruent and incongruent sentences, together with short visual animations that matched the auditory sentences. I assessed whether this visual component increased the strength of neural responses to semantic manipulations by boosting the semantic context of sentences, and increasing participant engagement. Furthermore, I added to the analyses conducted in Chapter 2 by using temporally and spatially unconstrained MVPA. Unconstrained MVPA analyses patterns of neural activity across electrodes and time, allowing for variations in the location and timing of neural responses and reducing multiple comparisons, possibly increasing reliability. These methodological and analysis improvements allowed me to replicate and extend the results found in Chapter 2. However, these did not improve the detection rate compared to the Chapter 2 sentences paradigm.

1.4.3 Chapter 4 - Finding hidden treasures: a child-friendly neural test of task-following in children using functional transcranial Doppler ultrasound

In Chapter 4, I set out to measure neural responses to task following in neurotypical children using fTCD. This study explored another avenue to measure language comprehension from neural signals by assessing task-following abilities. I based this work on previous fMRI research finding preserved task-following abilities in non-communicative patients (Bekinschtein et al., 2011; Naci et al., 2017; Owen et al., 2006), and tested whether an inexpensive and portable method, fTCD, could be used to this end. By contrasting language (i.e. generating words starting with a presented letter) and visuospatial tasks (i.e. remembering the location of letters on the screen), this paradigm offered an objective assessment of task-following, and therefore language comprehension. There was, again, large inter-individual variability in the neural responses recorded, and this paradigm was less reliable at the individual level than the Chapter 2 sentences paradigm.

1.4.4 Chapter 5 – Neural assessment of lexico-semantic processing in minimally-verbal autism: a case-series approach

In Chapter 5, I used the Chapter 2 sentences paradigm with EEG, contrasting congruent and incongruent spoken sentences, to test three minimally-verbal autistic children. This direct application of a carefully-designed paradigm allowed me to observe the neural response of these children to lexico-semantic anomalies. I found strong neural evidence for lexico-semantic processing in one child with extremely limited spoken language, which was apparent in both the MVPA decoding and a statistical N400 effect. This case series is a proof-of-concept for the application of my EEG paradigms to study this under-tested population.

1.4.5 Chapter 6 – Discussion

Finally, in Chapter 6, I draw the results together and discuss the implications of this work. By developing four neuroimaging paradigms and systematically assessing inter-individual reliability of neural signals in response to linguistic and task-following conditions, I provide strong empirical evidence of the large inter-individual variability in neural responses of children. Nevertheless, by applying the most reliable paradigm, together with powerful data analyses that allow for individual subject variation in time course and topology of effect, I was able to detect neural responses to lexico-semantic manipulations in one minimally-verbal autistic child.

1.5 Conclusion

Despite recent efforts to include minimally-verbal autistic children in research, still little is known about their cognitive abilities. To what extent does their behaviour correlate with their mental life? Is their language comprehension intact? If yes, at what level is language processing affected? To what extent does their inability to communicate relate to other factors such as motor control, or verbal and non-verbal intelligence? Which interventions are more suited for these individuals? In this thesis, I start addressing this large knowledge gap by developing implicit and objective neural tests of language comprehension that do not require behavioural responses. By developing rigorous and objective neural measures and combining them with accessible neuroimaging technologies, I hope to make it easier to include this understudied population in research. I hope my work will form a basis through which future research can add to theoretical knowledge of the neural and the cognitive profile in minimally-verbal autism. This work has the potential to radically influence the assessment, treatment and care of minimally-verbal individuals.

1.6 References

Adams, J., Faux, S. F., Nestor, P. G., Shenton, M., Marcy, B., Smith, S., & McCarley, R. W. (1993). ERP abnormalities during semantic processing in schizophrenia. *Schizophrenia Research*, *10*(3), 247–257. https://doi.org/10.1016/0920-9964(93)90059-R

Allen, D. A., & Rapin, I. (1980). Language disorders in preschool children: Predictors of outcome -- a preliminary report --. *Brain & Development*, *2*(1), 73–80.

Asperger, H. (1944). Die „Autistischen Psychopathen” im Kindesalter. *Archiv für Psychiatrie und Nervenkrankheiten*, *117*(1), 76–136. https://doi.org/10.1007/BF01837709

Badcock, N. A., Mousikou, P., Mahajan, Y., de Lissa, P., Thie, J., & McArthur, G. (2013). Validation of the Emotiv EPOC® EEG gaming system for measuring research quality auditory ERPs. *PeerJ*, *1*, e38. https://doi.org/10.7717/peerj.38

Badcock, N. A., Preece, K. A., Wit, B., Glenn, K., Fieder, N., Thie, J., & McArthur, G. (2015). Validation of the Emotiv EPOC EEG system for research quality auditory event-related potentials in children. *PeerJ*, *3*, e907. https://doi.org/10.7717/peerj.907

Badzakova-Trajkov, G., Corballis, M. C., & Häberling, I. S. (2016). Complementarity or independence of hemispheric specializations? A brief review. *Neuropsychologia*, *93*, 386–393. https://doi.org/10.1016/j.neuropsychologia.2015.12.018

Barham, M. P., Clark, G. M., Hayden, M. J., Enticott, P. G., Conduit, R., & Lum, J. A. G. (2017). Acquiring research-grade ERPs on a shoestring budget: A comparison of a modified Emotiv and commercial SynAmps EEG system. *Psychophysiology*, *54*(9), 1393–1404. https://doi.org/10.1111/psyp.12888

Behrmann, M., Thomas, C., & Humphreys, K. (2006). Seeing it differently: Visual processing in autism. *Trends in Cognitive Sciences*, *10*(6), 258–264. https://doi.org/10.1016/j.tics.2006.05.001

Bekinschtein, T. A., Manes, F. F., Villarreal, M., Owen, A. M., & Della Maggiore, V. (2011). Functional Imaging Reveals Movement Preparatory Activity in the Vegetative State. *Frontiers in Human Neuroscience*, *5*. https://doi.org/10.3389/fnhum.2011.00005

Beukema, S., Gonzalez-Lara, L. E., Finoia, P., Kamau, E., Allanson, J., Chennu, S., … Cruse, D. (2016). A hierarchy of event-related potential markers of auditory processing in disorders of consciousness. *NeuroImage : Clinical*, *12*, 359–371. https://doi.org/10.1016/j.nicl.2016.08.003

Biklen, D., Morton, M. W., Gold, D., Berrigan, C., & Swaminathan, S. (1992). Facilitated communication: Implications for individuals with autism. *Topics in Language Disorders*, *12*(4), 1–28.

Bishop, D. V. M., Watt, H., & Papadatou-Pastou, M. (2009). An efficient and reliable method for measuring cerebral lateralization during speech with functional transcranial Doppler ultrasound. *Neuropsychologia*, *47*(2), 587–590. https://doi.org/10.1016/j.neuropsychologia.2008.09.013

Bolton, P. F., Carcani-Rathwell, I., Hutton, J., Goode, S., Howlin, P., & Rutter, M. (2011). Epilepsy in autism: Features and correlates. *The British Journal of Psychiatry*, *198*(4), 289–294. https://doi.org/10.1192/bjp.bp.109.076877

Boudewyn, M. A., Long, D. L., & Swaab, T. Y. (2013). Effects of Working Memory Span on Processing of Lexical Associations and Congruence in Spoken Discourse. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00060

Brady, N. C., Anderson, C. J., Hahn, L. J., Obermeier, S. M., & Kapa, L. L. (2014). Eye tracking as a measure of receptive vocabulary in children with autism spectrum disorders. *Augmentative and Alternative Communication*, *30*(2), 147–159.

Braeutigam, S., Swithenby, S. J., & Bailey, A. J. (2008). Contextual integration the unusual way: A magnetoencephalographic study of responses to semantic violation in individuals with autism spectrum disorders. *European Journal of Neuroscience*, *27*(4), 1026–1036. https://doi.org/10.1111/j.1460-9568.2008.06064.x

Brignell, A., Morgan, A. T., Woolfenden, S., Klopper, F., May, T., Sarkozy, V., & Williams, K. (2018). A systematic review and meta-analysis of the prognosis of language outcomes for individuals with autism spectrum disorder. *Autism & Developmental Language Impairments*, *3*, 2396941518767610. https://doi.org/10.1177/2396941518767610

Brown, C., & Hagoort, P. (1993). The Processing Nature of the N400: Evidence from Masked Priming. *Journal of Cognitive Neuroscience*, *5*(1), 34–44. https://doi.org/10.1162/jocn.1993.5.1.34

Byrne, J. M., Dywan, C. A., & Connolly, J. F. (1995). An innovative method to assess the receptive vocabulary of children with cerebral palsy using event-related brain potentials. *Journal of Clinical and Experimental Neuropsychology*, *17*(1), 9–19. https://doi.org/10.1080/13803399508406576

Caldara, R., Jermann, F., Arango, G. L., & Van der Linden, M. (2004). Is the N400 category-specific? A face and language processing study. *Neuroreport*, *15*(17), 2589–2593. https://doi.org/10.1097/00001756-200412030-00006

Cantiani, C., Choudhury, N. A., Yu, Y. H., Shafer, V. L., Schwartz, R. G., & Benasich, A. A. (2016). From Sensory Perception to Lexical-Semantic Processing: An ERP Study in Non-Verbal Children with Autism. *PLOS ONE*, *11*(8), e0161637. https://doi.org/10.1371/journal.pone.0161637

Carlson, T. A., Grootswagers, T., & Robinson, A. K. (2019). An introduction to time-resolved decoding analysis for M/EEG. *ArXiv:1905.04820 [q-Bio]*. Retrieved from http://arxiv.org/abs/1905.04820

Carlson, T. A., Hogendoorn, H., Kanai, R., Mesik, J., & Turret, J. (2011). High temporal resolution decoding of object position and category. *Journal of Vision*, *11*(10). https://doi.org/10.1167/11.10.9

Chen, G. M., Yoder, K. J., Ganzel, B. L., Goodwin, M. S., & Belmonte, M. K. (2012). Harnessing repetitive behaviours to engage attention and learning in a novel therapy for autism: An exploratory analysis. *Frontiers in Psychology*, *3*, 12. https://doi.org/10.3389/fpsyg.2012.00012

Coderre, E. L., Chernenok, M., Gordon, B., & Ledoux, K. (2017). Linguistic and Non-Linguistic Semantic Processing in Individuals with Autism Spectrum Disorders: An ERP Study. *Journal of Autism and Developmental Disorders*, *47*(3), 795–812. https://doi.org/10.1007/s10803-016-2985-0

Coderre, E. L., Chernenok, M., O'Grady, J., Bosley, L., Gordon, B., & Ledoux, K. (2019). Implicit Measures of Receptive Vocabulary Knowledge in Individuals With Level 3 Autism. *Cognitive and Behavioral Neurology*, *32*(2), 95. https://doi.org/10.1097/WNN.0000000000000194

Connolly, J. F., Mate-Kole, C. C., & Joyce, B. M. (1999). Global aphasia: An innovative assessment approach. *Archives of Physical Medicine and Rehabilitation*, *80*(10), 1309–1315.

Coulson, S., & Van Petten, C. (2002). Conceptual integration and metaphor: An event-related potential study. *Memory & Cognition*, *30*(6), 958–968. https://doi.org/10.3758/BF03195780

Cruse, D., Chennu, S., Chatelle, C., Bekinschtein, T. A., Fernández-Espejo, D., Pickard, J. D., … Owen, A. M. (2011). Bedside detection of awareness in the vegetative state: A cohort study. *The Lancet*, *378*(9809), 2088–2094. https://doi.org/10.1016/S0140-6736(11)61224-5

Cruse, D., Chennu, S., Fernández-Espejo, D., Payne, W. L., Young, G. B., & Owen, A. M. (2012). Detecting Awareness in the Vegetative State: Electroencephalographic Evidence for Attempted Movements to Command. *PLOS ONE*, *7*(11), e49933. https://doi.org/10.1371/journal.pone.0049933

D'Arcy, R. C. N., Marchand, Y., Eskes, G. A., Harrison, E. R., Phillips, S. J., Major, A., & Connolly, J. F. (2003). Electrophysiological assessment of language function following stroke. *Clinical Neurophysiology*, *114*(4), 662–672. https://doi.org/10.1016/S1388-2457(03)00007-5

Davidson, M. M., & Weismer, S. E. (2017). A Discrepancy in Comprehension and Production in Early Language Development in ASD: Is it Clinically Relevant? *Journal of Autism and Developmental Disorders*, *47*(7), 2163–2175. https://doi.org/10.1007/s10803-017-3135-z

de Lissa, P., Sörensen, S., Badcock, N., Thie, J., & McArthur, G. (2015). Measuring the face-sensitive N170 with a gaming EEG system: A validation study. *Journal of Neuroscience Methods*, *253*, 47–54. https://doi.org/10.1016/j.jneumeth.2015.05.025

Deacon, D., & Shelley-Tremblay, J. (2000). How automatically is meaning accessed: A review of the effects of attention on semantic processing. *Frontiers in Bioscience: A Journal and Virtual Library*, *5*, E82-94.

Deacon, Diana, Dynowska, A., Ritter, W., & Grose-Fifer, J. (2004). Repetition and semantic priming of nonwords: Implications for theories of N400 and word recognition. *Psychophysiology*, *41*(1), 60–74. https://doi.org/10.1111/1469-8986.00120

DiStefano, C., Senturk, D., & Spurling Jeste, S. (2019). ERP Evidence of Semantic Processing in Children with ASD. *Developmental Cognitive Neuroscience*, 100640. https://doi.org/10.1016/j.dcn.2019.100640

Donchin, E., Spencer, K. M., & Wijesinghe, R. (2000). The mental prosthesis: Assessing the speed of a P300-based brain-computer interface. *IEEE Transactions on Rehabilitation Engineering*, *8*(2), 174–179. https://doi.org/10.1109/86.847808

Duchowski, A. T. (2007). Eye tracking methodology. *Theory and Practice*, *328*(614), 2–3.

Duvinage, M., Castermans, T., Petieau, M., Hoellinger, T., Cheron, G., & Dutoit, T. (2013). Performance of the Emotiv Epoc headset for P300-based applications. *BioMedical Engineering OnLine*, *12*, 56. https://doi.org/10.1186/1475-925X-12-56

Farwell, L. A., & Donchin, E. (1988). Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*, *70*(6), 510–523. https://doi.org/10.1016/0013-4694(88)90149-6

Fishman, I., Yam, A., Bellugi, U., Lincoln, A., & Mills, D. (2011). Contrasting patterns of language-associated brain activity in autism and Williams syndrome. *Social Cognitive and Affective Neuroscience*, *6*(5), 630–638. https://doi.org/10.1093/scan/nsq075

Ganis, G., Kutas, M., & Sereno, M. I. (1996). The search for "common sense": An electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience*, *8*(2), 89–106. https://doi.org/10.1162/jocn.1996.8.2.89

Gernsbacher, M. A. (2017). Editorial Perspective: The use of person-first language in scholarly writing may accentuate stigma. *Journal of Child Psychology and Psychiatry*, *58*(7), 859–861. https://doi.org/10.1111/jcpp.12706

Geuze, J., Gerven, M. A. J. van, Farquhar, J., & Desain, P. (2013). Detecting Semantic Priming at the Single-Trial Level. *PLOS ONE*, *8*(4), e60377. https://doi.org/10.1371/journal.pone.0060377

Geytenbeek, J., Harlaar, L., Stam, M., Ket, H., Becher, J. G., Oostrom, K., & Vermeulen, J. (2010). Utility of language comprehension tests for unintelligible or non-speaking children with cerebral palsy: A systematic review. *Developmental Medicine and Child Neurology*, *52*(12), e267-277. https://doi.org/10.1111/j.1469-8749.2010.03807.x

Gillespie-Lynch, K., Sepeta, L., Wang, Y., Marshall, S., Gomez, L., Sigman, M., & Hutman, T. (2012). Early Childhood Predictors of the Social Competence of Adults with Autism. *Journal of Autism and Developmental Disorders*, *42*(2), 161–174. https://doi.org/10.1007/s10803-011-1222-0

Gredebäck, G., Johnson, S., & von Hofsten, C. (2009). Eye tracking in infancy research. *Developmental Neuropsychology*, *35*(1), 1–19.

Grillon, C., Ameli, R., & Glazer, W. M. (1991). N400 and semantic categorization in schizophrenia. *Biological Psychiatry*, *29*(5), 467–480. https://doi.org/10.1016/0006-3223(91)90269-R

Groen, M. A., Whitehouse, A. J. O., Badcock, N. A., & Bishop, D. V. M. (2011). Where were those rabbits? A new paradigm to determine cerebral lateralisation of visuospatial memory function in children. *Neuropsychologia*, *49*(12), 3265–3271. https://doi.org/10.1016/j.neuropsychologia.2011.07.031

Groen, M. A., Whitehouse, A. J. O., Badcock, N. A., & Bishop, D. V. M. (2012). Does cerebral lateralization develop? A study using functional transcranial Doppler ultrasound assessing lateralization for language production and visuospatial memory. *Brain and Behavior*, *2*(3), 256–269. https://doi.org/10.1002/brb3.56

Grzadzinski, R., Huerta, M., & Lord, C. (2013). DSM-5 and autism spectrum disorders (ASDs): An opportunity for identifying ASD subtypes. *Molecular Autism*, *4*(1), 12. https://doi.org/10.1186/2040-2392-4-12

Haag, A., Moeller, N., Knake, S., Hermsen, A., Oertel, W. H., Rosenow, F., & Hamer, H. M. (2010). Language lateralization in children using functional transcranial Doppler sonography. *Developmental Medicine & Child Neurology*, *52*(4), 331–336. https://doi.org/10.1111/j.1469-8749.2009.03362.x

Hagoort, P., Baggio, G., & Wlllems, R. M. (2009). Semantic unification. In *The cognitive neurosciences*, *4th ed* (pp. 819–835). Cambridge, MA, US: Massachusetts Institute of Technology.

Hanson, S. J., & Bunzl, M. (2010). *Foundational Issues in Human Brain Mapping*. MIT Press.

Haxby, J. V. (2012). Multivariate pattern analysis of fMRI: The early beginnings. *NeuroImage*, *62*(2), 852–855. https://doi.org/10.1016/j.neuroimage.2012.03.016

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science*, *293*(5539), 2425–2430. https://doi.org/10.1126/science.1063736

Hemsley, B., Bryant, L., Schlosser, R. W., Shane, H. C., Lang, R., Paul, D., … Ireland, M. (2018). Systematic review of facilitated communication 2014–2018 finds no new evidence that messages delivered using facilitated communication are authored by the person with disability. *Autism & Developmental Language Impairments*, *3*, 2396941518821570. https://doi.org/10.1177/2396941518821570

Higashida, N. (2016). *The Reason I Jump: The Inner Voice of a Thirteen-Year-Old Boy with Autism* (Reprint edition; K. A. Yoshida & D. Mitchell, Trans.). New York: Random House Trade Paperbacks.

Hillyard, S. A., & Kutas, M. (1983). Electrophysiology of cognitive processing. *Annual Review of Psychology*, *34*, 33–61. https://doi.org/10.1146/annurev.ps.34.020183.000341

Hofvander, B., Delorme, R., Chaste, P., Nydén, A., Wentz, E., Ståhlberg, O. Leboyer, M. (2009). Psychiatric and psychosocial problems in adults with normal-intelligence autism spectrum disorders. *BMC Psychiatry*, *9*(1), 35. https://doi.org/10.1186/1471-244X-9-35

Holcomb, P. J., Coffey, S. A., & Neville, H. J. (1992). Visual and auditory sentence processing: A developmental analysis using event-related brain potentials. *Developmental Neuropsychology*, *8*(2–3), 203–241. https://doi.org/10.1080/87565649209540525

Hourcade, J., Everhart Pilotte, T., West, E., & Parette, P. (2004). A History of Augmentative and Alternative Communication for Individuals with Severe and Profound Disabilities. *Focus on Autism and Other Developmental Disabilities*, *19*(4), 235–244. https://doi.org/10.1177/10883576040190040501

Howlin, P. (2003). Outcome in high-functioning adults with autism with and without early language delays: Implications for the differentiation between autism and Asperger syndrome. *Journal of Autism and Developmental Disorders*, *33*(1), 3–13.

Howlin, P., & Moss, P. (2012). Adults with autism spectrum disorders. *Canadian Journal of Psychiatry. Revue Canadienne De Psychiatrie*, *57*(5), 275–283. https://doi.org/10.1177/070674371205700502

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*, *2*, 217–250.

Kasari, C., Brady, N., Lord, C., & Tager-Flusberg, H. (2013). Assessing the minimally verbal school-aged child with autism spectrum disorder. *Autism Research: Official Journal of the International Society for Autism Research*, *6*(6), 479–493. https://doi.org/10.1002/aur.1334

Kasari, C., Paparella, T., Freeman, S., & Jahromi, L. B. (2008). Language outcome in autism: Randomized comparison of joint attention and play interventions. *Journal of Consulting and Clinical Psychology*, *76*(1), 125–137. https://doi.org/10.1037/0022-006X.76.1.125

Kedar, I. (2012). *Ido in Autismland: Climbing Out of Autism's Silent Prison*. Sharon Kedar.

Kenny, L., Hattersley, C., Molins, B., Buckley, C., Povey, C., & Pellicano, E. (2016). Which terms should be used to describe autism? Perspectives from the UK autism community. *Autism: The International Journal of Research and Practice*, *20*(4), 442–462. https://doi.org/10.1177/1362361315588200

Kiang, M., Patriciu, I., Roy, C., Christensen, B. K., & Zipursky, R. B. (2013). Test-retest reliability and stability of N400 effects in a word-pair semantic priming paradigm. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *124*(4), 667–674. https://doi.org/10.1016/j.clinph.2012.09.029

Kim, A. E., Oines, L., & Miyake, A. (2018). Individual differences in verbal working memory underlie a tradeoff between semantic and structural processing difficulty during language comprehension: An ERP investigation. *Journal of Experimental Psychology.*

Learning, Memory, and Cognition, *44*(3), 406–420. https://doi.org/10.1037/xlm0000457

Kostova, M., Passerieux, C., Laurent, J.-P., & Hardy-Baylé, M.-C. (2005). N400 anomalies in schizophrenia are correlated with the severity of formal thought disorder. *Schizophrenia Research*, *78*(2), 285–291. https://doi.org/10.1016/j.schres.2005.05.015

Kutas, M, Van Petten, C., Ackles, P., Jennings, J., & Coles, M. (1988). *Advances in psychophysiology*.

Kutas, Marta, & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, Marta, & Hillyard, S. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*, 203–205. https://doi.org/10.1126/science.7350657

Kwok, E. Y. L., Brown, H. M., Smyth, R. E., & Oram Cardy, J. (2015). Meta-analysis of receptive and expressive language skills in autism spectrum disorder. *Research in Autism Spectrum Disorders*, *9*, 202–222. https://doi.org/10.1016/j.rasd.2014.10.008

Lee, P.-L., Sie, J.-J., Liu, Y.-J., Wu, C.-H., Lee, M.-H., Shu, C.-H., Shyu, K.-K. (2010). An SSVEP-Actuated Brain Computer Interface Using Phase-Tagged Flickering Sequences: A Cursor System. *Annals of Biomedical Engineering*, *38*(7), 2383–2397. https://doi.org/10.1007/s10439-010-9964-y

Levy, S. E., Giarelli, E., Lee, L.-C., Schieve, L. A., Kirby, R. S., Cunniff, C., … Rice, C. E. (2010). Autism spectrum disorder and co-occurring developmental, psychiatric, and medical conditions among children in multiple populations of the United States. *Journal of Developmental and Behavioral Pediatrics: JDBP*, *31*(4), 267–275. https://doi.org/10.1097/DBP.0b013e3181d5d03b

Lindell, A. K., Notice, K., & Withers, K. (2009). Reduced language processing asymmetry in non-autistic individuals with high levels of autism traits. *Laterality*, *14*(5), 457–472.

Lohmann, H., Dräger, B., Müller-Ehrenberg, S., Deppe, M., & Knecht, S. (2005). Language lateralization in young children assessed by functional transcranial Doppler sonography. *NeuroImage*, *24*(3), 780–790. https://doi.org/10.1016/j.neuroimage.2004.08.053

Loring, D. W., & Meador, K. J. (2000). Pre-surgical evaluation for epilepsy surgery. *Neurosciences*, *5*(3), 143–150.

Luck, S. J. (2014). *An introduction to the event-related potential technique*.

Masterton, B. A., & Biederman, G. B. (1983). Proprioceptive versus visual control in autistic children. *Journal of Autism and Developmental Disorders*, *13*(2), 141–152. https://doi.org/10.1007/BF01531815

McCleery, J. P., Ceponiene, R., Burner, K. M., Townsend, J., Kinnear, M., & Schreibman, L. (2010). Neural correlates of verbal and nonverbal semantic integration in children with autism spectrum disorders. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *51*(3), 277–286. https://doi.org/10.1111/j.1469-7610.2009.02157.x

Méndez, M., Sans, O., Abril, B., & Valdizan, J. R. (2009). 13. Event-related potentials (N 400) in autistic children. *Clinical Neurophysiology*, *120*(4), e136. https://doi.org/10.1016/j.clinph.2008.09.041

Millar, Light, & Schlosser. (2006). The Impact of Augmentative and Alternative Communication Intervention on the Speech Production of Individuals With Developmental Disabilities: A Research Review. *Journal of Speech, Language, and Hearing Research*, *49*(2), 248–264. https://doi.org/10.1044/1092-4388(2006/021)

Ming Cheng, Xiaorong Gao, Shangkai Gao, & Dingfeng Xu. (2002). Design and implementation of a brain-computer interface with high transfer rates. *IEEE Transactions on Biomedical Engineering*, *49*(10), 1181–1186. https://doi.org/10.1109/TBME.2002.803536

Moghimi, S., Kushki, A., Marie Guerguerian, A., & Chau, T. (2013). A Review of EEG-Based Brain-Computer Interfaces as Access Pathways for Individuals with Severe Disabilities. *Assistive Technology*, *25*(2), 99–110. https://doi.org/10.1080/10400435.2012.723298

Monti, M. M., Vanhaudenhuyse, A., Coleman, M. R., Boly, M., Pickard, J. D., Tshibanda, L., … Laureys, S. (2010). Willful Modulation of Brain Activity in Disorders of Consciousness. *New England Journal of Medicine*, *362*(7), 579–589. https://doi.org/10.1056/NEJMoa0905370

Münte, T., Urbach, T., Duzel, E., & Kutas, M. (2000). Event-related brain potentials in the study of human cognition and neuropsychology. *Handbook of Neuropsychology*, *1*.

Naci, L., Cusack, R., Anello, M., & Owen, A. M. (2014). A common neural code for similar conscious experiences in different individuals. *Proceedings of the National Academy of Sciences*, *111*(39), 14277–14282. https://doi.org/10.1073/pnas.1407007111

Naci, L., Sinai, L., & Owen, A. M. (2017). Detecting and interpreting conscious experiences in behaviorally non-responsive patients. *NeuroImage*, *145*, 304–313. https://doi.org/10.1016/j.neuroimage.2015.11.059

Navarro-Orozco, D., & Sánchez-Manso, J. C. (2019). Neuroanatomy, Middle Cerebral Artery. In *StatPearls*. Retrieved from http://www.ncbi.nlm.nih.gov/books/NBK526002/

Nijboer, F., Sellers, E. W., Mellinger, J., Jordan, M. A., Matuz, T., Furdea, A., Kübler, A. (2008). A P300-based brain–computer interface for people with amyotrophic lateral sclerosis. *Clinical Neurophysiology*, *119*(8), 1909–1916. https://doi.org/10.1016/j.clinph.2008.03.034

Ornitz, E. M. (1973). Childhood autism—A review of the clinical and experimental literature. *California Medicine*, *118*(4), 21.

Owen, A. M. (2008). Disorders of consciousness. *Annals of the New York Academy of Sciences*, *1124*, 225–238. https://doi.org/10.1196/annals.1440.013

Owen, A. M., Coleman, M. R., Boly, M., Davis, M. H., Laureys, S., & Pickard, J. D. (2006). Detecting Awareness in the Vegetative State. *Science*, *313*(5792), 1402. https://doi.org/10.1126/science.1130197

Paton, B., Hohwy, J., & Enticott, P. G. (2012). The Rubber Hand Illusion Reveals Proprioceptive and Sensorimotor Differences in Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *42*(9), 1870–1883. https://doi.org/10.1007/s10803-011-1430-7

Paul, R., Chawarska, K., Cicchetti, D., & Volkmar, F. (2008). Language Outcomes of Toddlers With Autism Spectrum Disorders: A Two Year Follow-Up. *Autism Research : Official Journal of the International Society for Autism Research*, *1*(2), 97–107. https://doi.org/10.1002/aur.12

Paynter, J. M., Ferguson, S., Fordyce, K., Joosten, A., Paku, S., Stephens, M., … Keen, D. (2017). Utilisation of evidence-based practices by ASD early intervention service providers. *Autism*, *21*(2), 167–180. https://doi.org/10.1177/1362361316633032

Paynter, J. M., & Keen, D. (2015). Knowledge and use of intervention practices by community-based early intervention service providers. *Journal of Autism and Developmental Disorders*, *45*(6), 1614–1623. https://doi.org/10.1007/s10803-014-2316-2

Pijnacker, J., Geurts, B., van Lambalgen, M., Buitelaar, J., & Hagoort, P. (2010). Exceptions and anomalies: An ERP study on context sensitivity in autism. *Neuropsychologia*, *48*(10), 2940–2951. https://doi.org/10.1016/j.neuropsychologia.2010.06.003

Piotroski, J., & Naigles, L. (2012). Intermodal preferential looking. *Research Methods in Child Language*, 17–28.

Plesa Skwerer, D., Jordan, S. E., Brukilacchio, B. H., & Tager-Flusberg, H. (2016). Comparing methods for assessing receptive language skills in minimally verbal children and adolescents with autism spectrum disorders. *Autism*, *20*(5), 591–604. https://doi.org/10.1177/1362361315600146

Posar, A., & Visconti, P. (2018). Sensory abnormalities in children with autism spectrum disorder. *Jornal De Pediatria*, *94*(4), 342–350. https://doi.org/10.1016/j.jped.2017.08.008

Preslar, J., Kushner, H. I., Marino, L., & Pearce, B. (2014). Autism, lateralisation, and handedness: A review of the literature and meta-analysis. *Laterality: Asymmetries of Body, Brain and Cognition*, *19*(1), 64–95.

Rapin, I., & Dunn, M. (2003). Update on the language disorders of individuals on the autistic spectrum. *Brain & Development*, *25*(3), 166–172.

Ribeiro, T. C., Valasek, C. A., Minati, L., & Boggio, P. S. (2013). Altered semantic integration in autism beyond language: A cross-modal event-related potentials study. [Miscellaneous Article]. *Neuroreport*, *24*(8), 414–418. https://doi.org/10.1097/WNR.0b013e328361315e

Rohaut, B., Faugeras, F., Chausson, N., King, J.-R., Karoui, I. E., Cohen, L., & Naccache, L. (2015). Probing ERP correlates of verbal semantic processing in patients with impaired consciousness. *Neuropsychologia*, *66*, 279–292. https://doi.org/10.1016/j.neuropsychologia.2014.10.014

Rosch, R. E., Bishop, D. V. M., & Badcock, N. A. (2012). Lateralised visual attention is unrelated to language lateralisation, and not influenced by task difficulty – A functional transcranial Doppler study. *Neuropsychologia*, *50*(5), 810–815. https://doi.org/10.1016/j.neuropsychologia.2012.01.015

Russo, N., Mottron, L., Burack, J. A., & Jemel, B. (2012). Parameters of semantic multisensory integration depend on timing and modality order among people on the autism spectrum: Evidence from event-related potentials. *Neuropsychologia*, *50*(9), 2131–2141. https://doi.org/10.1016/j.neuropsychologia.2012.05.003

Rutter, M. (1977). Infantile autism and other psychoses. *Child Psychiatry: Modern Approaches*, 717–747.

Rutter, Michael. (1970). Autistic children: Infancy to adulthood. *Seminars in Psychiatry*, *2*, 435.

Rutter, Michael. (1978). Language Disorder and Infantile Autism. In Michael Rutter & E. Schopler (Eds.), *Autism: A Reappraisal of Concepts and Treatment* (pp. 85–104). https://doi.org/10.1007/978-1-4684-0787-7_6

Schaefer, R. S., Farquhar, J., Blokland, Y., Sadakata, M., & Desain, P. (2011). Name that tune: Decoding music from the listening brain. *NeuroImage*, *56*(2), 843–849. https://doi.org/10.1016/j.neuroimage.2010.05.084

Schlosser, R. W., Hemsley, B., Shane, H., Todd, J., Lang, R., Lilienfeld, S. O., Odom, S. (2019). Rapid Prompting Method and Autism Spectrum Disorder: Systematic Review Exposes Lack of Evidence. *Review Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s40489-019-00175-w

Schnakers, C., Perrin, F., Schabus, M., Majerus, S., Ledoux, D., Damas, P., Laureys, S. (2008). Voluntary brain processing in disorders of consciousness. *Neurology*, *71*(20), 1614. https://doi.org/10.1212/01.wnl.0000334754.15330.69

Schnakers, Caroline, Perrin, F., Schabus, M., Hustinx, R., Majerus, S., Moonen, G., Laureys, S. (2009). Detecting consciousness in a total locked-in syndrome: An active event-related paradigm. *Neurocase*, *15*(4), 271–277. https://doi.org/10.1080/13554790902724904

Serby, H., Yom-Tov, E., & Inbar, G. F. (2005). An improved P300-based brain-computer interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *13*(1), 89–98. https://doi.org/10.1109/TNSRE.2004.841878

Sinclair, J. (2013). Why I dislike “person first” language. *Autonomy, the Critical Journal of Interdisciplinary Autism Studies*, *1*(2).

Szatmari, P., Bryson, S. E., Boyle, M. H., Streiner, D. L., & Duku, E. (2003). Predictors of outcome among high functioning children with autism and Asperger syndrome. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *44*(4), 520–528.

Tager-Flusberg, H. (1981). On the nature of linguistic functioning in early infantile autism. *Journal of Autism and Developmental Disorders*, *11*(1), 45–56.

Tager-Flusberg, Helen. (1999). A Psychological Approach to Understanding the Social and Language Impairments in Autism. *International Review of Psychiatry (Abingdon, England)*, *11*(4), 325–334. https://doi.org/10.1080/09540269974203

Tager-Flusberg, Helen. (2000). The challenge of studying language development in children with autism. *Methods for Studying Language Production*, 313–332.

Tager-Flusberg, Helen, Edelson, L., & Luyster, R. (2011). Language and Communication in Autism Spectrum Disorders. In *Autism Spectrum Disorders*.

Tager-Flusberg, Helen, & Kasari, C. (2013). Minimally Verbal School-Aged Children with Autism Spectrum Disorder: The Neglected End of the Spectrum. *Autism Research*, *6*(6), 468–478. https://doi.org/10.1002/aur.1329

Tager-Flusberg, Helen, Plesa Skwerer, D., Joseph, R. M., Brukilacchio, B., Decker, J., Eggleston, B., Yoder, A. (2017). Conducting research with minimally verbal participants with autism spectrum disorder. *Autism: The International Journal of Research and Practice*, *21*(7), 852–861. https://doi.org/10.1177/1362361316654605

Tanaka, H., Watanabe, H., Maki, H., Sakriani, S., & Nakamura, S. (2019). Electroencephalogram-Based Single-Trial Detection of Language Expectation Violations in Listening to Speech. *Frontiers in Computational Neuroscience*, *13*. https://doi.org/10.3389/fncom.2019.00015

Tanner, D., Goldshtein, M., & Weissman, B. (2018). Individual differences in the real-time neural dynamics of language comprehension. *Psychology of Learning and Motivation*, *68*, 299–335.

Tomchek, S. D., & Dunn, W. (2007). Sensory processing in children with and without autism: A comparative study using the short sensory profile. *The American Journal of Occupational Therapy: Official Publication of the American Occupational Therapy Association*, *61*(2), 190–200. https://doi.org/10.5014/ajot.61.2.190

Townsend, J., Harris, N. S., & Courchesne, E. (1996). Visual attention abnormalities in autism: Delayed orienting to location. *Journal of the International Neuropsychological Society*, *2*(6), 541–550. https://doi.org/10.1017/S1355617700001715

Vos, M. D., Kroesen, M., Emkes, R., & Debener, S. (2014). P300 speller BCI with a mobile EEG system: Comparison to a traditional amplifier. *Journal of Neural Engineering*, *11*(3), 036008. https://doi.org/10.1088/1741-2560/11/3/036008

Wang, Y., Gao, X., Hong, B., Jia, C., & Gao, S. (2008). Brain-Computer Interfaces Based on Visual Evoked Potentials. *IEEE Engineering in Medicine and Biology Magazine*, *27*(5), 64–71. https://doi.org/10.1109/MEMB.2008.923958

Whitehouse, A. J. O., & Bishop, D. V. M. (2009). Hemispheric division of function is the result of independent probabilistic biases. *Neuropsychologia*, *47*(8), 1938–1943. https://doi.org/10.1016/j.neuropsychologia.2009.03.005

Wilkinson, K. M. (1998). Profiles of language and communication skills in autism. *Mental Retardation and Developmental Disabilities Research Reviews*, *4*(2), 73–79. https://doi.org/10.1002/(SICI)1098-2779(1998)4:2<73::AID-MRDD3>3.0.CO;2-Y

Willie, C., Colino, F., Bailey, D., Tzeng, Y., Binsted, G., Jones, L., others. (2011). Utility of transcranial Doppler ultrasound for the integrative assessment of cerebrovascular function. *Journal of Neuroscience Methods*, *196*(2), 221–237.

Wilson, K. R., O'Rourke, H., Wozniak, L. A., Kostopoulos, E., Marchand, Y., & Newman, A. J. (2012). Changes in N400 topography following intensive speech language therapy for individuals with aphasia. *Brain and Language*, *123*(2), 94–103. https://doi.org/10.1016/j.bandl.2012.06.005

Yijun Wang, Ruiping Wang, Xiaorong Gao, Bo Hong, & Shangkai Gao. (2006). A practical VEP-based brain-computer interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *14*(2), 234–240. https://doi.org/10.1109/TNSRE.2006.875576

Yoshimura, Y., Kikuchi, M., Shitamichi, K., Ueno, S., Munesue, T., Ono, Y., others. (2013). Atypical brain lateralisation in the auditory cortex and language performance in 3-to 7-year-old children with high-functioning autism spectrum disorder: A child-customised magnetoencephalography (MEG) study. *Molecular Autism*, *4*(1), 38.

Chapter 2

Towards an individualised neural assessment of receptive language in children

Selene Petit[*,1], Nicholas A. Badcock[1], Tijl Grootswagers[1,2], Anina N. Rich[1], Jon Brock[1], Lyndsey Nickels[1], Denise Moerel[1], Nadene Dermody[1,3], Shu Yau[1,4], Elaine Schmidt[3,5], Alexandra Woolgar[1,3]

[1] Department of Cognitive Science, Macquarie University, Australia

[2] School of Psychology, University of Sydney, Australia

[3] Medical Research Council (UK), Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, UK

[4] School of Psychology & Exercise Science, Murdoch University, Western Australia.

[5] Department of Linguistics, Macquarie University, Australia

* Corresponding author at: Australian Hearing Hub, Level 3, 16 University Avenue, Macquarie University, NSW 2109, Australia.

This chapter is currently under review at *JSLHR*:

Petit S., Badcock N. A., Grootswagers T., Rich A. N., Brock J., Nickels L., Moerel D., Dermody N., Yau S., Schmidt E., Woolgar A. (under review). Towards an individualised neural assessment of receptive language in children.

Abstract

Purpose: We aimed to develop a non-invasive neural test of language comprehension to use with non-speaking children for whom standard behavioural testing is unreliable (e.g. minimally verbal autism). In particular, our aims were three-fold. First, we sought to establish the reliability of neural responses to spoken language in individual neurotypical children. Second, we aimed to validate the use of a portable and accessible electroencephalography (EEG) system, by comparing its recordings to those of a research-grade system. Third, in light of substantial inter-individual variability in individuals' neural responses, we assessed whether multivariate decoding methods could improve reliability.

Methods: We tested the reliability of two child-friendly N400 paradigms. Thirty-one typically developing children listened to identical spoken words that were either strongly predicted by the preceding context or violated lexical-semantic expectations. We measured EEG concurrently from a research-grade system, Neuroscan's SynAmps2, and an adapted gaming system, Emotiv's EPOC+.

Results: We found substantial inter-individual variability in the timing and topology of N400-like effects. For both paradigms and EEG systems, traditional N400 effects at the expected sensors and time points were statistically significant in around 50% of individuals. Using multivariate analyses, detection rate increased to 88% of individuals, illustrating the robustness of this method in the face of inter-individual variations in topography.

Conclusions: There was large inter-individual variability in neural responses to semantic expectations, suggesting inter-individual variation in either the cognitive response to lexical-semantic violations, and/or the neural substrate of that response. Around half of our neurotypical participants showed the expected N400 effect at the expected location and time point. A low-cost, accessible EEG system provides comparable data for univariate analysis but

was not well suited to multivariate decoding. However, multivariate analyses with research-grade EEG increased our detection rate to 88% of individuals, providing a promising avenue to index language comprehension in children with limited communication.

2.1 Introduction

Language is a crucial part of everyday life and something we often take for granted. In cases where people cannot speak or reliably communicate, it can be difficult to assess whether an individual is able to understand spoken language. Examples include disorders of consciousness, minimally-verbal autism, or cerebral palsy (Giacino and Smart, 2007; Harrison and Connolly, 2013; Tager-Flusberg and Kasari, 2013). In these cases, there is an urgent need for passive and reliable individualised measures of receptive language processing. Such a test could transform both individual assessment and treatment, and our scientific understanding of cognitive profile in these populations.

Electroencephalography (EEG) is a passive way to record neural responses, and offers the opportunity to measure language understanding in the absence of reliable behavioural responses. In particular, the N400 event-related potential (ERP) is elicited by hearing or reading words, with the N400 being larger when the words violate the semantic or predictive context in which they are presented (e.g. Kutas, 1993; Kutas and A. Hillyard, 1980; Kutas and Federmeier, 2011). For instance, the word "door" elicits a larger N400 ERP when it is presented in the sentence "The clouds are high up in the *door*", compared to the sentence "I had no key to open the *door*". This difference in the N400 amplitude is called the N400 effect. It is well-documented and has been recorded in groups of adults, children, and special populations (for a comprehensive review see Kutas and Federmeier, 2011). The N400 has previously been used as a neural index of linguistic processing in special populations, such as cerebral palsy (Byrne et al., 1995), traumatic brain injury (Connolly et al., 1999), stroke (D'Arcy et al., 2003), autism

(Cantiani et al., 2016; Coderre et al., 2019; DiStefano et al., 2019), and disorders of consciousness (Schoenle and Witzke 2004, Steppacher et al 2013), making it a promising marker of lexico-semantic processing. However, most of these studies either relied on visual inspection of the waveforms and did not conduct statistical analyses of the N400 or lacked a sufficient control group to make strong inference in their clinical sample.

In addition, only a few studies have systematically assessed the sensitivity of statistical N400 effects in individuals (Beukema et al., 2016; Cruse et al., 2014; Rohaut et al., 2015), and to our knowledge, only one has reported them for individual children (Cantiani et al., 2016). The sparse results on neurotypical individuals suggest a medium sensitivity of auditory N400 paradigms, with around half of the individuals showing an N400 effect to semantic manipulations. This has important implications both for neuro-cognitive understanding of lexico-semantic processing, and for the possibility of using N400 as a future neural marker of language processing in non-verbal individuals. Here we sort to redress this knowledge gap by establishing the sensitivity of two auditory paradigms for detecting lexico-semantic processing in individual neurotypical children.

In the first experiment, neurotypical children listened to pairs of spoken words which were either normatively associated (e.g. “arm – *leg*”) or unrelated (e.g. “boat – *leg*”). In the second, children listened to spoken sentences with either a congruent completion (i.e. “she wore a necklace around her *neck*”) or an incongruent/anomalous completion (i.e. “the princess may someday become a *neck*”). These paradigms elicit strong and reliable N400 effects in adults and children (Borovsky et al., 2012; Friedrich and Friederici, 2005; Kiang et al., 2013; Rämä et al., 2013; Torkildsen et al., 2007), so we predicted that the ERP evoked by the identical spoken word token (in our example, *neck*) would vary according to semantic context. To make the paradigms suitable for children we created game-like tasks where children encountered friendly or evil aliens.

With a view to future clinical applications, we also sought to validate an accessible EEG system that avoids many of the typical setup inconveniences. A traditional 32-channel, gel-based EEG system takes around 35 minutes to setup and is somewhat uncomfortable, involving rubbing the participant's scalp with gel in order to bridge the EEG electrodes to the skin. In addition, most typical EEG setups are not portable, with extensive wiring compelling participants to remain seated, and signals are best recorded in an electrically shielded room, which can be intimidating for some subjects. Although previous research have successfully used laboratory EEG systems to test special populations such as autistic participants (Coderre et al., 2019; McCleery et al., 2010) or Rett syndrome (Laan and Vein, 2002), alternative portable solutions for children or adults with cognitive disorders may allow the inclusion of more individuals in research. Recently, more accessible and more portable EEG systems have become available, one of which has been adapted and validated for the measurement of auditory ERPs in adults (Badcock et al., 2013; de Lissa et al., 2015), children (Badcock et al., 2015), and autistic children (Yau et al., 2015). The Emotiv EPOC+ system, hereafter referred to as "EPOC+", was originally designed for gaming purposes, and consists of a wireless headset with 14 electrodes that connect to the scalp via saline solution-soaked cotton-rolls. The setup is fast (approx. 5-10 minutes) and it is not necessary to rub the scalp. This system is also low in cost compared with research-grade systems and is wireless and portable, allowing its use outside of the laboratory (e.g. in homes or schools).

Although the EPOC system (the predecessor of EPOC+) has been validated against research-grade systems for recording early ERPs, such as auditory ERPs (Badcock et al., 2015, 2013; Barham et al., 2017), and face-sensitive N170 (de Lissa et al., 2015), studies on later components such as the P300 have yielded less consistent results. Vos et al. (2014) report similar performance of an Emotiv amplifier compared to a research-grade amplifier, and Elsawy et al. (2014) found acceptable results when using a classifier on P300 EPOC data, while

Duvinage et al. (2013) report that the EPOC recorded a significantly noisier signal compared to the research-grade ANT system (Advanced Neuro Technology, ANT, Enschede, The Netherlands). To our knowledge, no studies have tested it on N400 ERPs. Using a portable and fast-to-setup system to record N400s would be a significant step towards assessing language comprehension in children who do not speak and may not tolerate a research-grade EEG system. For this reason, we tested the fidelity of the adapted Emotiv EPOC+ EEG system against data recorded concurrently from a research-grade Neuroscan system during the experimental tasks.

We assessed the ERP differences between semantic conditions using both typical univariate N400 analyses and multivariate pattern analyses (MVPA). Traditional univariate analyses of ERPs usually require an *a priori* choice of electrodes and time points of interest. When testing individual participants, especially children and special populations, this *a priori* knowledge may not be available. Thus, in addition to univariate analyses of the N400 effect, we tested the sensitivity of MVPA on our EEG data. We trained a linear classifier to discriminate between the two semantic conditions (congruent and incongruent) on individual-subject EEG data. This targets the information contained in the pattern of activation across sensors, making it robust to individual differences in signal direction and topology (Grootswagers et al., 2017; Haynes, 2015; Hebart and Baker, 2018).

Despite robust group-level univariate N400 effects, we found substantial inter-subject variability in topology and time course of response, with an individual subject detection rate of around 50% in both paradigms, in both research-grade and consumer-grade systems, using classical analysis. In the sentences paradigm, MVPA increased the detection rate to 88% but only for the research-grade EEG system. These data suggest heterogeneity in the neural responses to lexico-semantic processing in neurotypical children and may help direct future development of paradigms for assessing language comprehension in non-speaking children.

2.2 Experiment one: normatively associated word pairs

2.2.1 Methods

2.2.1.1 Participants

Sixteen children were recruited using the Neuronauts database of the Australian Research Council Centre of Excellence in Cognition and its Disorders. All participants were native English speakers and had non-verbal reasoning and verbal abilities within the normal range as measured by the matrices section of the Kaufman Brief Intelligence Test, Second Edition (K-BIT 2, Kaufman and Kaufman, 2004) and the Peabody Picture Vocabulary Test—4th Edition (PPVT – 4, Dunn and Dunn, 2007). Participants received $25 for their participation, as well as a sticker and a certificate. The data from one participant were excluded due to technical issues during recording. The final set of data thus came from 15 participants (age range: 6 to 12 years old, $M=9.2$, $SD=2.6$, 4 male and 11 female). This study was approved by the Macquarie University Human Research Ethics Committee (Reference number: 5201200658). Participants' parents or guardians provided written consent and the children provided verbal consent.

2.2.1.2 Stimuli

Stimuli comprised 63 pairs of normatively associated words. Following Cruse et al. (2014), we began with word pairs taken from the Nelson et al. (1998) free association norms database. These norms comprise a large number of cue-target pairings, developed by asking participants to produce the first meaningfully or strongly associated word that comes to mind when presented with a particular cue. We initially chose pairs from the normative database with a forward associative strength (cue to target) greater than 0.5, meaning that more than 50% of the participants in the Nelson et al. norm-development study produced this target word in response to the cue. We included only pairs where the target was a noun, and where the cue and target were one syllable in length. We also included only words that had an age of acquisition rating of 8-years or less (Kuperman et al., 2012), meaning that these words were typically known by children of 8-years and above. We excluded any pairs where either the cue or target had a homophone (according to the N-watch database; Davis, 2005) with an age of acquisition of less than or equal to 10 years, and where the cue or target was not applicable to the Australian context (e.g. FUEL-GAS).

To minimise repetition across the stimulus set, we allowed each target word to appear in a maximum of two word pairs. The cue words were only used once as cues, but they could also appear up to twice as targets. Thus, the maximum number of repetitions of particular words across the entire list of related items was three (i.e. once as a cue, and twice as a target – this was the case for 13 words). For word pairs with singular and plural forms (e.g. GIRL-BOY and GIRLS-BOYS), only the pair with the strongest association was included. The final set of 63 pairs had a mean forward associative strength of 0.676 (see supplementary Table S1 for the list of stimuli). These word pairs formed the *related* condition.

We created a list of 63 *unrelated* word pairs by recombining the cue and target words from the related condition. Constructing the unrelated list in this way ensured a fully balanced design in which the cues and targets in the related and unrelated lists were identical (and therefore matched for word frequency, familiarity, phoneme length, etc.). We ensured that target words in the unrelated condition did not start with the same sound, rhyme, or have any semantic or associative connections with the cue or related target. In addition, we respected the grammatical number structure of the related word pair when choosing an unrelated target. For example, in creating an unrelated combination for a plural-singular pair (e.g. SUDS-SOAP), another singular target word was chosen (e.g. SUDS-ART).

Stimuli were digitally recorded by a female native Australian-English speaker, and the best auditory tokens, where the voice had a natural intonation and was not raspy, were selected using Praat software (Boersma, 2001). We used the same target tokens in the related and unrelated conditions so that there were no auditory differences to drive a differential EEG response to the two conditions. For each target, the related and the unrelated cue were recorded close together in time, and were chosen to have approximately the same length, intensity, and voice quality as the target.

2.2.1.3 EEG Equipment

We recorded simultaneously from two EEG systems in an electrically-shielded room. The research EEG system Neuroscan SynAmps2 (Scan version 4.3) Ag-AgCl electrodes were fitted to an elastic cap (Easy Cap, Herrsching, Germany) at 33 locations (Figure 1), according to the international 10-20 system, including M1 (online reference), AFz (ground electrode), and M2. We measured vertical and horizontal eye movements with electrodes placed above and below the left eye and next to the outer canthus of each eye. Neuroscan was sampled at 1000 Hz (downsampled to 500 Hz during processing) with an online bandpass filter from 1 to 100 Hz. We marked the onset of each sentence and target word using parallel port events.

The EPOC+ is a wireless headset with flexible plastic arms holding 16 gold-plated sensors. In order to accommodate the concurrent setup with the Neuroscan system, we placed the EPOC+ sensors at the following scalp locations of the international 10-20 system (See Figure 1)[1]. M1 acted as the online reference, and M2 was a feed-forward reference that reduced external electrical interference. The signals from the other 14 channels were high-pass filtered

[1] Note that in the Emotiv software, TestBench 3.1.21, the electrodes at FC3/4 are labelled F3/4, F3/4 are labelled AF3/4 and FT7/8 are labelled FC5/6; this is because we adjusted the electrode placement to accommodate the concurrent setup. In this paper we refer to the electrodes according to their placement on the scalp when worn concurrent with the Neuroscan EasyCap, not the labels used in the Emotiv software.

online with a 0.16 Hz cut-off, pre-amplified and low-pass filtered at an 83 Hz cut-off. The analogue signals were then digitised at 2048 Hz. The digitised signal was filtered using a 5th-order sinc filter to notch out 50Hz and 60 Hz, low-pass filtered and down-sampled to 128 Hz (specifications taken from the EPOC+ system web forum). The effective bandwidth was 0.16–43 Hz.

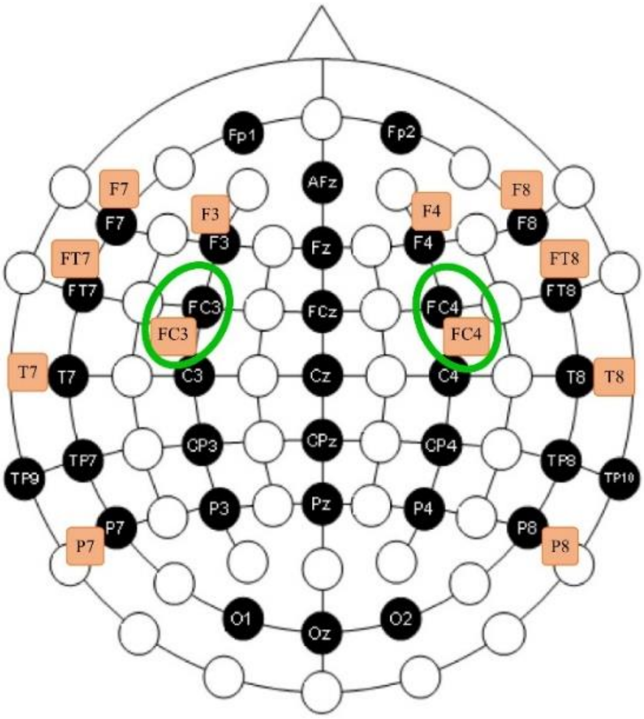


Figure 1. Electrode position on the scalp for the Neuroscan (black circles) and EPOC+ (orange rectangles) systems. The adjacent electrodes used to calculate correlations between the two systems are circled in green.

To accurately time-lock the ERPs to the onset of the target word, we modified the EPOC+ system to incorporate event markers, following Badcock et al. (2015). We did this using a custom-made transmitter unit communicating with a custom receiver unit through infrared light (Thie, 2013). The transmitter box was connected to the audio output of the presentation computer. At the onset of each sentence and each target word, a tone of particular frequency (2400Hz for the sentence onset, 600Hz for a related target onset, and 1600Hz for an unrelated target onset) was sent to the transmitter unit through a separate audio channel. This in turn activated the receiver unit, which generated an electrical pulse in the O1 and O2 channels (from which we did not acquire neural data).

2.2.1.4 Experimental Procedure

For each participant, we set up the Neuroscan system first and adjusted the impedances to under 5 kΩ. We then placed the EPOC+ system over the top, with cotton wool bridging scalp to sensor through custom slits in the EasyCap. EPOC+ impedances were adjusted to be below 220 kΩ in the TestBench software. Setup took up to 50 minutes, during which participants watched a DVD of their choice. Following setup, participants were seated in front of a 17-inch monitor, with speakers on both sides of the screen, at a viewing distance of about 1 metre. Before the main experiment, participants completed the PPVT 4, and after the EEG session, they completed the matrices section of the K-BIT 2.

Participants completed two EEG acquisition sessions of 20 minutes, separated by a 5-minute break. Each session included all 126 cue-target word pairs (63 related, 63 unrelated). The stimuli were presented using Psychophysics Toolbox 3 extensions (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) in MATLAB. The word pairs were presented in a pseudo-random order which was reversed in the second session. The order was optimised to minimise order bias in the sequence of related and unrelated trials, with the additional constraint that no more than 4 trials in a row were of either condition. The first session was divided into 17 blocks consisting of 5 to 15 trials, and the second section was divided into 11 blocks of 8 to 15 trials, as the participants were more familiar with the task during the second session.

To make the task more engaging for children, it was introduced in the context of a story. The child was told that they were listening to different aliens in an English speaking competition, and they had to rate each alien (1-5 stars) depending on its skill at producing pairs of words that went together. We asked participants to listen carefully to each pair and to decide whether the two words were related or unrelated, then give an overall judgment of the alien's performance (1 star if they produced mainly unrelated pairs, 5 stars if they said mainly related words) at the end of each block. This encouraged the children to pay attention to, and make a covert decision about, the relatedness of the words in each pair. The choice of this covert task was motivated by the results from Cruse et al. (2014), who found that a covert design was more sensitive than a passive design where participants were asked to simply listen to the stimuli (58% for the covert task versus 0% for the passive task). According to the same study, an overt task would have been more sensitive again, but we anticipated that our target population may not give reliable overt responses, whereas they may be able to perform the task covertly.

At the beginning of each block, an alien appeared on the screen, then moved behind a black "recording booth" in the middle of the screen. A light bulb was depicted on top of the box and lit up during each trial to encourage children to pay attention and to reduce eye movements during the trial. The light bulb lit up 500 ms before the cue word, remained lit until 1500 ms after the target word onset, then turned off. The interval between the cue and target was 1000 ms. After another 1500 ms, the next trial began. At the end of each block, the alien moved out of the box and participants were prompted to grade the alien on a five-point scale regarding its overall performance. At the end of the experiment, participants were shown the "winners" of the alien contest. The total time, including the EEG setup, was about 1hr 40 minutes.

2.2.1.5 Offline EEG processing

We processed all EEG signals in EEGLAB (v13.4.4b, Delorme and Makeig, 2004) in MATLAB (R2014b). We first applied a band-pass filter between 0.1 and 40Hz, then cut the data into epochs from -100 ms to 1000 ms around target word onset. We ran an Independent Component Analysis (ICA) on all the epochs. Components with scalp distribution, frequency, and timing that corresponded to eye movements and eye blinks were removed from the Neuroscan data. In line with a previous EPOC+ study with children (Badcock et al., 2015), we could not identify any eyeblink artefacts in the EPOC+ data. This may be because eye blinks were not consistent or strong enough to affect EPOC+ data. Alternatively, the ICA for the Neuroscan data could have benefited from the signal recorded by the Neuroscan electrodes recording eye movements. The EPOC+ did not have such electrodes so the ICA for the EPOC+ used only the scalp electrodes. The epochs were baseline corrected against the averaged signal from the 100 ms preceding the target onset. These data were then used for the decoding analyses. For univariate analyses, we further removed, from all electrodes, epochs with extreme values (±150μV) in any of the electrodes of interest (see below). An average of 6 epochs (SD = 4, min = 0, and max = 41) were rejected from the Neuroscan data, and of 15 epochs (SD = 9, min = 1, and max = 58) from the EPOC+ data.

2.2.1.6 Group ERP analyses

The N400 is typically recorded over the centroparietal regions of the brain. We therefore focused our univariate analyses on three electrodes sites of interest: Cz, as the N400 effect is reported to be the strongest in centro-parietal sites (Kutas and Federmeier, 2011), and FC3 and FC4, which are the closest channels to Cz that we can compare between the Neuroscan and EPOC+ systems.

We first verified that our paradigm evoked a classic N400 effect at a group level and tested whether this was detectable in both Neuroscan and EPOC+ systems. For trial-averaged waveforms, we ran group analyses with paired t-tests comparing the two conditions at each time point from 150 ms after stimulus onset for each of the five sensors of interest. We corrected for multiple-comparisons for each channel independently using a statistical temporal cluster extent threshold (Guthrie and Buchwald, 1991). Briefly, this method calculates the autocorrelation between consecutive time points of the ERP signal, for each channel. We can then determine the minimum number of consecutive time points that need to show a statistical difference in a one-tailed t-test between the related and unrelated conditions to be considered a significant cluster at $p < .05$ corrected (for details, see Guthrie and Buchwald, 1991). We used a one-tailed test because the direction of the N400 effect (a more negative response in the unrelated condition) was pre-specified. We restricted our statistical analyses to the time points after 150 ms to decrease the number of statistical tests computed, as the N400 is typically reported to occur later than 150 ms (Kutas and Federmeier, 2011).

We additionally illustrated the topographic distribution of the N400 effect, based on the Neuroscan grand average data, by subtracting activation in the related condition from the unrelated condition and averaging over sequential 200 ms time windows (200 to 400 ms, 400 to 600 ms, 600 to 800 ms and 800 to 1000 ms).

2.2.1.7 Single subject ERP analyses

Our next goal was to assess the sensitivity of our paradigm and EEG systems to detect N400 effects in individual children. For each individual and each system, we conducted first-level (single subject) analyses using independent samples t-tests between the two conditions at the electrodes Cz (for Neuroscan only), and FC3 and FC4 (for both systems), for each time point starting at +150 ms after the target onset. To correct for multiple comparisons, we again used the autocorrelation score to determine the temporal cluster extent threshold for each electrode in each participant independently (mean cluster length over participants and electrodes for Neuroscan: M = 132 ms, range = [70, 206], and for Emotiv: M = 58 ms, range = [47, 70]).

We illustrated the topographic distribution of the N400 effect in individuals based on the Neuroscan data by subtracting activation in the related from the unrelated condition and averaging over sequential 200 ms time windows (200 to 400 ms, 400 to 600 ms, 600 to 800 ms and 800 to 1000 ms).

2.2.1.8 Comparison of EEG systems

We next sought to validate the EPOC+ system for recording N400 ERPs. To compare the shape of Neuroscan and EPOC+ waveforms, we ran intraclass correlations (ICC) (Bishop et al., 2007), a global index of waveforms similarities and amplitude; and Spearman's correlations, which measure the rank correlation between two waveforms and is therefore less sensitive to amplitude. For this analysis, in order to have a fair comparison between the two systems, we re-did the pre-processing so that the Neuroscan and EPOC+ data were as comparable as possible and treated in the same way. For this, the processing proceeded as describe above except that we down sampled Neuroscan data to match EPOC+'s sampling rate of 128 Hz and we did not remove eye-blink components from either system. We then calculated the correlations for each condition, using the entire epoch, at our two locations of interest where the electrodes from the two systems lie in close proximity: the left and right frontocentral sites (FC3 and FC4 - see locations on Figure 2). We calculated the correlation for each condition, in each individual at these two locations. We examined whether correlations were significant by computing the 95% confidence interval of the group mean and checking if they overlapped with 0 (which would correspond to no correlation). Finally, we asked whether the amplitude of the N400 effect differed between the two systems. To this end, we compared the area under the difference curve (related – unrelated ERP), using a trapezoidal integration from 300 ms to 800 ms, between the two systems using a two-tailed, paired t-test across individuals. These time points correspond to the expected N400 effect time course (Kutas and Federmeier, 2011).

2.2.1.9 Group and single subject MVPA

In order to be sensitive to individual variation in the topology of N400 effects, without increasing multiple comparisons (i.e. without analysing every electrode separately), we used multivariate pattern analyses (MVPA). We analysed all the data using the CosmoMVPA toolbox (Oosterhof et al., 2016) in MATLAB. We used a linear discriminant analysis (LDA) classifier to analyse the group and the individual data obtained by Neuroscan and EPOC+ separately. At each time point, we divided our data into a training set and a testing set, using a leave-one-target-out cross-validation approach. The training set consisted of the activation pattern across all the electrodes for trials corresponding to all the targets but one, and the classifier was trained to find the decision boundary that best distinguished between the two categories (related versus unrelated). Since each pair was repeated once, the training set consisted of 248 (62 stimuli * 2 conditions * 2 repetitions) trials. We then tested the classifier's ability to classify the category of the remaining four trials (two related, two unrelated) corresponding to the remaining target. We repeated this procedure 63 times, each time leaving a different target out. Finally, we averaged the accuracy of the classifier for these 63 tests to yield an accuracy value at each time point and for each individual. The group average classifier accuracy, at each time point, was obtained by averaging each individual's accuracy. If the classifier performs significantly above chance (50%), we infer that there was information in the brain signals that differed between the two conditions (related versus unrelated words) at that time point.

For statistical inference we implemented a sign-permutation test (at the group level) or a label-permutation test (in individuals) at each time point (Maris and Oostenveld, 2007). The sign-permutation test consists of randomly swapping the sign (positive or negative after subtracting chance, 50%) of the decoding results of each of the participants. The label-permutation test consists of randomly permuting the condition label of the targets before classification to obtain classifier accuracies under the null-hypothesis. We performed 1000 permutations to obtain a null distribution at each time point. The observed (i.e. correctly labelled) decoding accuracies and permutation results were then transformed using Threshold Free Cluster Enhancement (TFCE; Smith and Nichols, 2009) which yields a statistic of cluster level support at each time point. To account for multiple comparisons across time, the maximum TFCE statistic of each permutation from across all time points was used to form a single corrected null-distribution. The observed (correctly labelled) TFCE statistic at each time point was considered significantly above chance if it was larger than 95% of the TFCE values in the corrected null-distribution.

2.2.2 Results

2.2.2.1 Behavioural results

All children had a standard score within or above the normal range (90-110) for non-verbal reasoning (K-BIT M= 123, 95% CI [114,131]) and receptive vocabulary (PPVT M= 120, 95% CI [115,126]). We asked children for a subjective rating (1-5 stars) of the performance of the aliens in each block but, as there was no 'correct' answer, we do not report accuracy. Children seemed to understand the instructions well and informally reported the task to be engaging.

2.2.2.2 Group ERP analyses

At the group level, we replicated the typical N400 effect using the Neuroscan system. We found significant N400 effects at all three of our regions of interest: Cz, FC3, and FC4 (Figure 2, left panels). For the central location (Cz), the N400 effect was significant for a cluster of time points from 272 – 1000 ms, post-stimulus onset (Figure 2, top panel). For FC3, the N400 effect was significant for a cluster from 292 – 1000 ms, and for FC4 the N400 effect was significant in a cluster from 302 – 1000 ms. The group-level topographic distribution of the effect (Figure 2, bottom panel) was initially centro-parietal, spreading frontally at later time points. We were also able to record N400 effects for the group using the EPOC+ system in FC3 (from 350 to 747 ms) and FC4 (from 469 to 596, and from 684 to 739 ms), our two locations of interest (Figure 2, right panels).

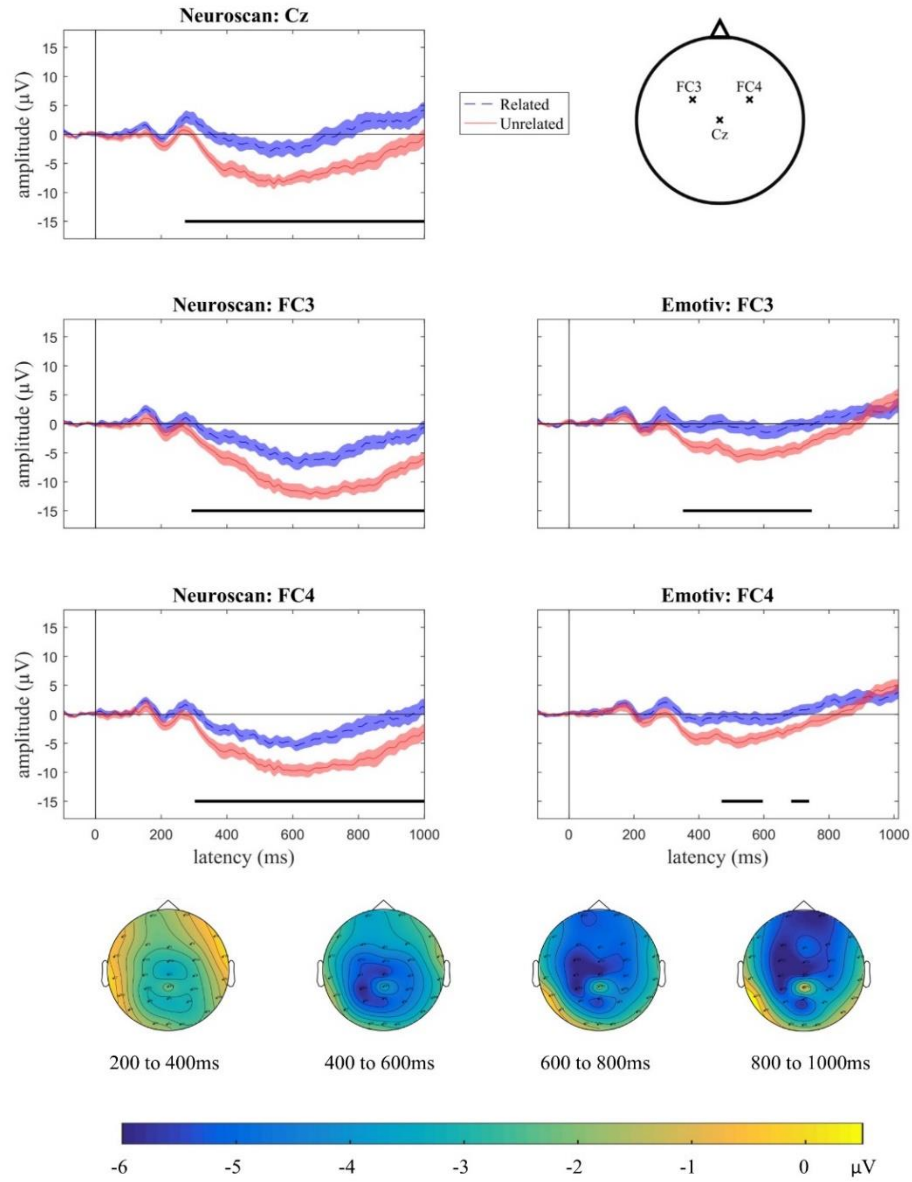


Figure 2. Group N400 effects for Experiment 1 (word pairs). Plots display grand average ERPs (n=16), with related (dashed blue) and unrelated (solid red) conditions for Neuroscan electrodes Cz (top left panel), FC3 (middle left) and FC4 (bottom left), and EPOC+ electrodes F3 (middle right) and F4 (bottom right). Shading indicates standard error of the mean. Time points at which

there was a statistical difference between the conditions are indicated with a solid black line under the plot ($p < .05$, after cluster correction for multiple comparisons). Locations are shown on the top right panel. The bottom panel illustrates the topographic map of the N400 effect (unrelated minus related condition) from 200 to 1000 ms after target onset in the group for the Neuroscan system. Yellow colours indicate no difference between the two conditions and blue colours indicate a more negative-going response for the unrelated condition. The N400 effect was distributed over central and centro-frontal regions.

2.2.2.3 Single subject ERP analyses

Our next goal was to assess the detection rate of N400 effects in individual subjects. We defined a significant N400 effect as the presence of a statistically larger N400 in the unrelated compared to related condition (corrected for multiple comparisons across time points) in one or both of the two locations of interest that were present in both systems (FC3 or FC4). Table 1 shows the percentage of participants with significant N400 effects ("detection rate") at each electrode. A significant N400 effect was found in 7 of the 15 (47%) participants' Neuroscan data, and in the same number of participants' EPOC+ data. Two participants showed an effect in the Neuroscan data but not the EPOC+ data, and vice versa. Additionally, we assessed whether inter-individual differences in the N400 effect could be explained by individual factors, such as age, vocabulary or non-verbal reasoning. We did not find any significant correlation between the amplitude of the N400, as measured by the area between the curves between 300 and 800ms recorded from Neuroscan at Cz, and age ($r = -.12$, $p = .68$), PPVT score ($r = -.19$, $p = .52$), or K-BIT score ($r = -.07$, $p = .52$).

Table 1. Experiment 1 (word pairs) detection rate (% of individuals) of statistically significant N400 effects in each of the three electrodes of interest, and the detection rate in a more lenient assessment where the effect was considered present if it occurred in either one or both of the two frontal electrodes.

Electrode	FC3	FC4	Total (FC3 and/or FC4)	Cz
Neuroscan	47%	33%	47%	47%
EPOC+	40%	40%	47%	N/A

We show the individual waveforms recorded by the Neuroscan in Cz, which is where N400 effects are typically recorded (Figure 3, first and fourth columns) and in FC3, where we have both Neuroscan and EPOC+ data (Figure 3, second, third, fifth and sixth columns); FC4 results were similar. In addition, we illustrate the topography of the effect over time (Figure 4). To summarise, the detection rate of individual N400 effects was less than 50% with either EEG system, and large variations in the topography of the effect were found, suggesting inter-individual variabilities in the recorded neural signals.

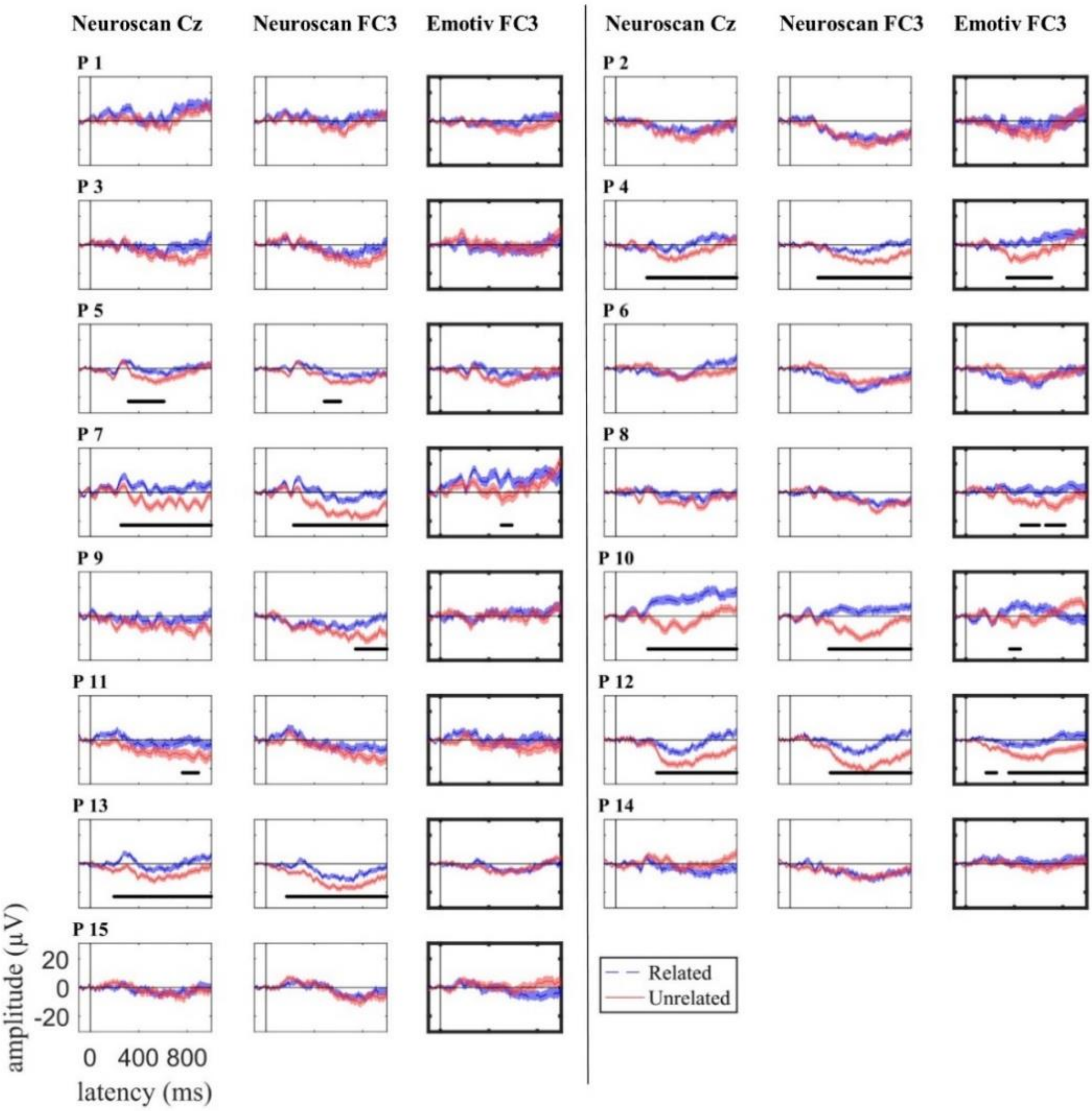


Figure 3. Experiment 1 (word pairs paradigm) individual participant event-related potentials to target words following a related (dashed blue) and unrelated (solid red) word for Neuroscan electrode Cz (first and fourth column), and adjacent Neuroscan and EPOC+ electrode FC3 (second, third, fifth, and sixth column), plotted ± standard error (shaded area). Time points where there was a statistically significant N400 effect in each participant and sensor are indicated with a solid, horizontal, black line. EPOC+ results are outlined in bold. P indicates participant.

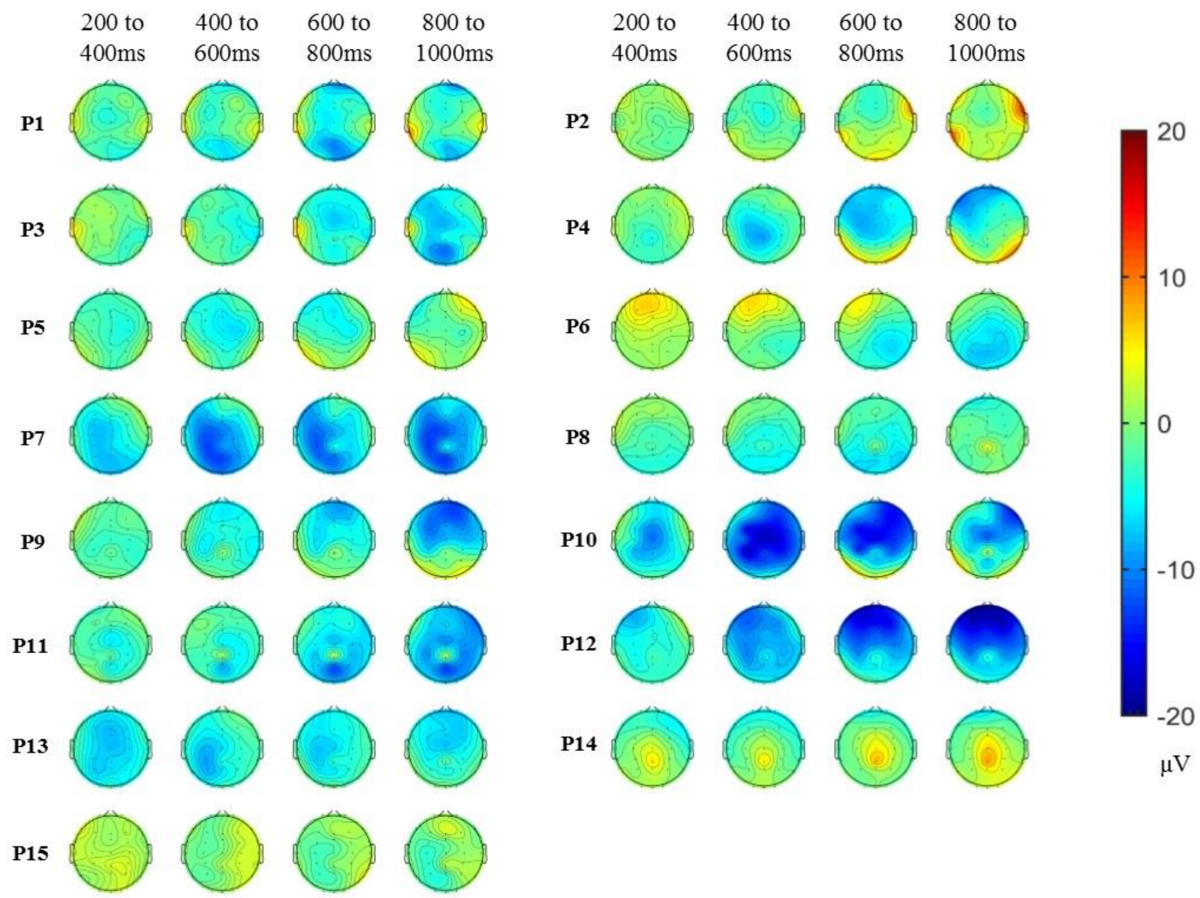


Fig 4. Experiment 1 (word pairs paradigm) individual topographic maps of the N400 effect (unrelated minus related condition) for 200 ms time windows from 200 to 1000 ms after target onset for Neuroscan. Red areas indicate more negative-going response for the related condition and blue areas indicate more negative-going response for the unrelated condition. The topography of the N400 effect varied across individuals. P indicates participant.

2.2.2.4 Comparison of EEG systems

We next compared the responses of the two EEG systems using ICC and Spearman's rank correlation (see Table 2) and by comparing the areas under the difference curves. Waveforms across the two systems were qualitatively similar in shape, and positively correlated (all Spearman's rho ≥ 0.49, 95% CI not including zero for any of the comparisons). Mean ICC values ranged from 0.19 to 0.63 across the different conditions and sites. The ICC was greater than 0 (CIs did not include 0) for the related condition on both sides and for the unrelated condition on the left side, but not for the unrelated condition on the right side. We tested whether the amplitude of the effect was larger for Neuroscan compared to EPOC+. The area between the related and unrelated curves was numerically larger for Neuroscan than for EPOC+ both in FC3 (140 for EPOC+ versus 294 for Neuroscan), and FC4 (82 for EPOC+ versus 251 for Neuroscan), but the difference was not significant (FC3: $t_{(14)} = 1.90$, $p = .0779$; FC4: $t_{(14)} = 1.74$, $p = .103$). Therefore, the ERPs recorded by the two systems were comparable in shape and not significantly different in amplitude.

Table 2: Experiment 1 (word pairs paradigm, bottom) mean Spearman's rho (ρ) and ICC (r), and 95% confidence intervals, between waveforms simultaneously recorded with the research (Neuroscan) and gaming (EPOC+) EEG systems for the left (FC3) and right (FC4) frontocentral locations, in the semantically related and unrelated conditions. We also present the difference in area between the two condition curves between Neuroscan and EPOC+ averaged across subjects, with 95% confidence intervals.

Condition	**Coeff.**	**Location**	
		Left frontocentral (FC3)	Right frontocentral (FC4)
Related	**ρ**	0.63 [0.51, 0.75]	0.53 [0.37, 0.70]
	r	0.46 [0.31, 0.60]	0.25 [0.03, 0.48]
Unrelated	**ρ**	0.52 [0.33, 0.72]	0.49 [0.27, 0.71]
	r	0.27 [0.02, 0.52]	0.19 [-0.1, 0.47]
Area between curves (µV)		EPOC+ : 140 [-45, 324] Neuroscan : 294 [139, 448]	EPOC+ : 82 [-113, 276] Neuroscan : 251 [118, 384]

2.2.2.5 Decoding analyses

Our univariate N400 analyses were restricted to three *a priori* sites of interest. However, individual topography plots (Figure 4) suggested that the topography of the effect was highly variable between individuals, with effect location ranging from centroparietal to frontal sites. Therefore, we used MVPA to integrate information from across all sensors to detect differences in the neural patterns of activity to related and unrelated targets. Group level decoding performance (average over subjects) for the Neuroscan and EPOC+ data is shown in Figure 5. For Neuroscan (Figure 5, purple), classifier accuracy was statistically above chance in a cluster from 402 ms after target onset until the end of the epoch, indicating a reliable difference between the two conditions. There was no significant coding of associative context in the EPOC+ EEG data (Figure 5, yellow).

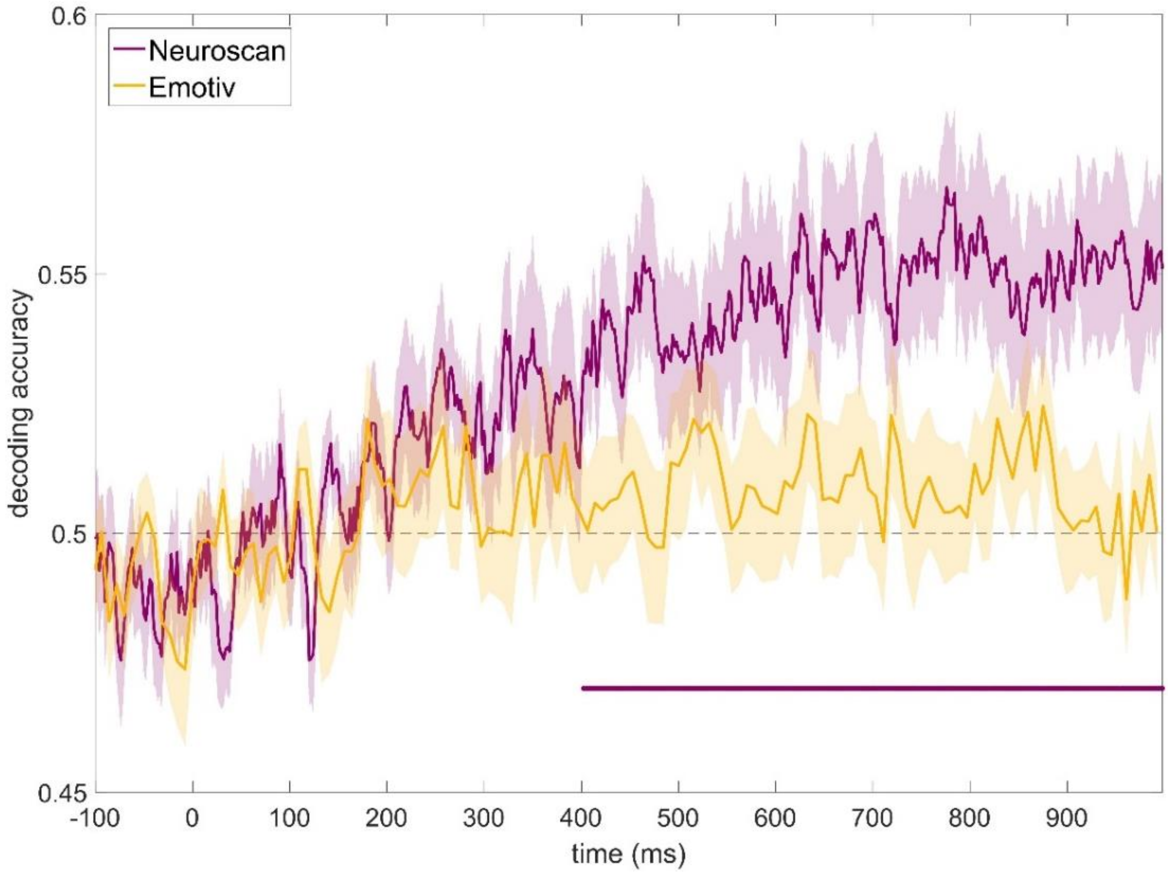


Figure 5. Experiment 1 (word pairs) grand average decoding accuracy for discriminating between congruent and incongruent conditions over time for Neuroscan (purple) and EPOC+ (yellow) data, shown with standard error of the mean. Time points with significant decoding for Neuroscan ($p < .05$, assessed with TFCE permutation tests corrected for multiple comparisons, see Methods) are shown by a purple horizontal line. Decoding accuracy was significantly above chance for Neuroscan from 402 ms but was not significant at any time point for EPOC+.

Individual decoding results are shown in Figure 6. In the Neuroscan data, decoding was significant in 8/15 participants (53% detection rate). However, for EPOC+, the classifier only detected a significant effect in 2 of the 15 participants (13% detection rate).

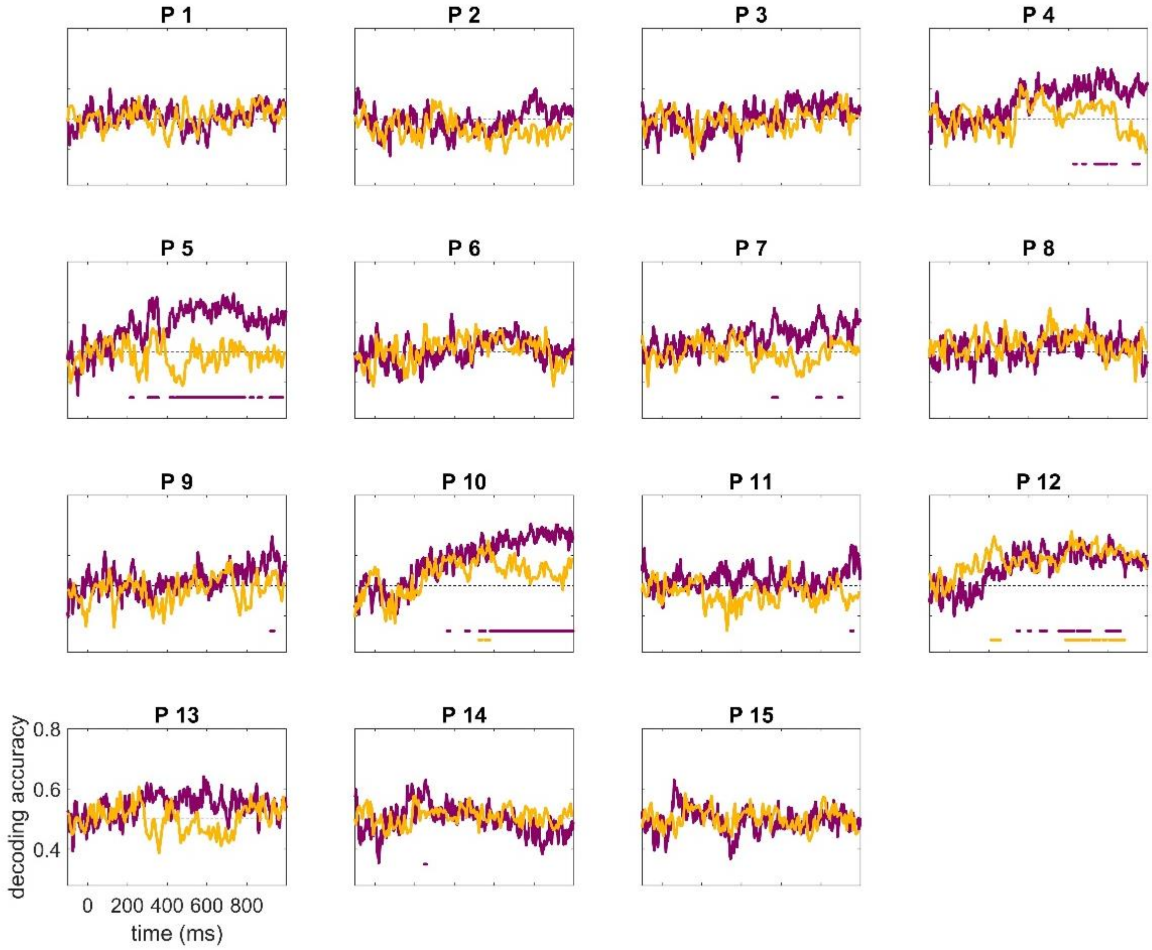


Figure 6. Experiment 1 (word pairs) individual participant decoding accuracy for classification of congruent versus incongruent conditions over time for Neuroscan (purple) and EPOC+ (yellow) data. Time points with accuracy significantly above chance ($p < .05$ assessed with TFCE permutation tests corrected for multiple comparisons, see Methods) are shown as solid horizontal lines (in purple for Neuroscan, and yellow for EPOC+). Chance (50%) is indicated by the horizontal dashed line. Semantic condition could be decoded in 8/15 (53%) of individuals' Neuroscan data and in 2/15 (13%) of individual's EPOC+ data. P indicates participant.

2.2.3 Experiment 1 summary

We examined whether differential neural responses were elicited to identical spoken words presented in different normatively-associative contexts using two different EEG systems, Neuroscan and EPOC+. We were able to elicit N400 effects in a group of children, as well as in some individual children. However, detection rates were low (47% of individuals using either the Neuroscan or EPOC+ system), and the topography and timing of the effect were variable across individuals. Multivariate pattern analyses returned similar (53%) or weaker (13%) detection rates for Neuroscan and EPOC+, respectively. Interestingly, we recorded similar waveforms using EPOC+, suggesting its potential utility for research purposes.

Since our overarching aim was to derive a sensitive measure of semantic language processing for use in individual children, we next considered another avenue to increase our detection rate of N400 effects. A common way to elicit N400 effects is to present words in the context of semantically congruent or incongruent sentences. This may yield larger N400 effects than the word pairs task, as sentences provide a stronger semantic context compared to a single probe word (Kutas, 1993). In addition, any deleterious effect of repeating stimuli (which is necessary to perfectly match the stimuli across conditions) may be attenuated in sentences since so many words are presented on each trial (e.g. Cruse et al. 2014). For this reason, our second experiment used words presented in sentences. We also modified the task that participants performed to make it more demanding and encourage greater attention to the stimuli than for the word pairs task.

2.3 Experiment 2: congruent and incongruent sentences

2.3.1 Methods

2.3.1.1 Participants

Eighteen participants, aged 6 to 12 years, were recruited as described for Experiment 1. The data from two participants were excluded due to excessive artefacts in the EEG data. The final set of data thus came from 16 participants (mean age =10.3, SD=2.4, 8 male and 8 female), five of whom had also participated in Experiment 1.

2.3.1.2 Stimuli

We created two conditions: (1) *congruent sentences*, which were semantically correct (e.g. “she wore a necklace around her *neck*”); and (2) *incongruent sentences,* which ended with an anomalous word (e.g. “There were candles on the birthday *neck*”). The set of congruent sentences were based on 56 high-close probability sentences from the norms of Block and Baldwin (2010) and were chosen according to suitability for children and such that target words were high-frequency words acquired by age five (Kuperman et al., 2012). We recombined sentence stems and target words to form the set of incongruent sentences. We ensured that the incongruent target word was unexpected but grammatically correct. It also did not begin with the same phoneme or rhyme with the corresponding congruent target word. Within one session, each sentence stem and target was used twice, once in the congruent condition and once in the incongruent condition. The final set of stimuli consisted of 56 sentences in each condition, 112 in total (see supplementary table S2 for the complete list).

Stimuli were digitally recorded by a female native British English speaker in a soundproof room and edited in Audacity®. To avoid co-articulation, the speaker recorded the sentence stems separately from the target words. This also introduced a lengthening in the final word of the sentence stem. Sentence stems and targets were combined online during stimulus presentation with a 100 ms silence between the sentence frame and the target word.

2.3.1.3 EEG Equipment

The equipment and experimental setup were the same as in Experiment 1, including the completion of the matrices section of the Kaufman Brief Intelligence Test, Second Edition (K-BIT), and the Peabody Picture Vocabulary Test, Fourth Edition (PPVT). Participants who completed Experiment 1 did not complete these tests a second time.

2.3.1.4 Experimental procedure

Participants completed two EEG recording sessions of 25 minutes, separated by a 5-minute break. Each session included all 112 sentences (56 congruent, 56 incongruent). We presented the sentences in a pseudo-random order that was reversed in the second session. We optimised the order to avoid bias in the sequence of related and unrelated trials as described above, with all sentences presented once before being repeated in the alternate condition, and to maximise the distance between repeated presentations of the same target word. We allocated the sentences to this trial order pseudo-randomly with the additional constraint that there were at least two sentences between any repetitions of semantic content in the sentence frame or target word. We presented an image of a satellite centrally on the screen to signal the onset of each trial, and kept this display on for the whole trial. This served as an alerting cue and encouraged children to fixate, reducing eye-movements. After 2 s, we presented the sentence through the speakers and the satellite remained onscreen for a further 1.5 s after the presentation of the target word (Figure 7). There was then a 2 s inter-trial-interval before the next trial. Each 20-minute session consisted of 16 blocks of 4 to 10 trials, after which children gave an answer to the experimenter (see below).

We designed a task that was strongly engaging for children while requiring minimal overt responses. It was embedded in the context of a story: an evil alien Lord had messed up some of the “messages” that we were trying to send to our extra-terrestrial friends. Participants were asked to pay attention to each sentence and to count how many did not make sense. Accurate responses, given at the end of each block, would help “catch” an evil alien’s henchman who appeared on the screen. This encouraged participants to pay attention and to make covert semantic judgments of sentences. Most of the children appeared to be highly engaged and motivated by the task, and reported that they found it entertaining. The whole experiment, including setup, took approximately 2 hours.

2.3.1.5 Offline EEG processing, ERP, and MVPA

The correlation scores, ERP analyses, and decoding analyses were performed as for Experiment 1.

2.3.2 Results

2.3.2.1 Behavioural results

All participants scored within or above the normal range for non-verbal reasoning (K-BIT score M= 111, 95% CI [101,120]) and receptive vocabulary (PPVT score M=117, 95% CI [111,122]).

Participants performed the behavioural task with a high degree of accuracy (mean percent correct: M=96.29%, SD=3.36%, range = [88.4%, 100%]), indicating that they understood the sentences, and were able to notice semantic anomalies.

2.3.2.2 Group ERP analyses

We recorded large N400 effects in the group using Neuroscan in all of the electrodes of interest (Figure 8, left panels). We also recorded N400-like effects using the EPOC+ system at our two locations of interest, FC3 and FC4. However, these effects only reached significance at FC4, possibly indicating a lesser sensitivity of the EPOC+ system (Figure 7, right panels).

For the central location (Cz), the N400 effect started at 171 ms and continued until 817 ms post-stimulus onset (Figure 7, top panel). For the frontal sites, the effect started later with a significant cluster from 409 – 697 ms for FC3 (Figure 7, middle left panel) and a significant cluster from 387 – 875 ms for FC4 (Figure 7, bottom left panel). For EPOC+, the N400 effect was significant in FC4 for a cluster from 418 – 752 ms. Potentials in both conditions and all three sensors also shifted in the positive direction over time, possibly corresponding to the closure positive shift, an ERP component reflecting the processing that occurs at a prosodic boundary (Steinhauer et al., 1999; Steinhauer and Friederici, 2001).

The topographic distribution of the N400 effect is shown in Figure 7, bottom panel. The distribution was centro-frontal with a slight right bias. This is in line with previous reports that found the N400 effect to be more frontal in children than in adults (Friedrich and Friederici, 2004; Henderson et al., 2011a), although we did not observe this in Experiment 1.

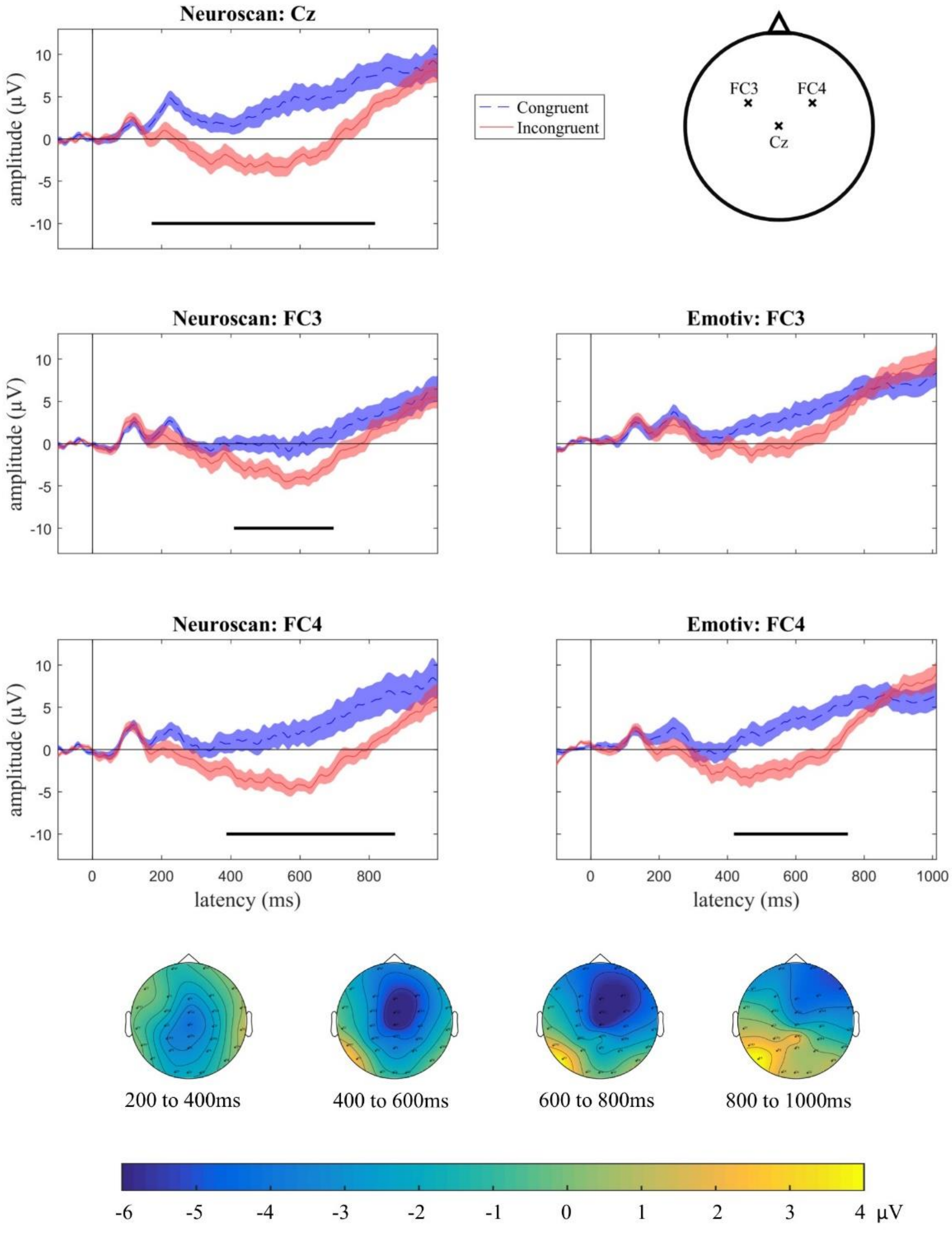


Figure 7. Group N400 effects for Experiment 2 (sentence paradigm). Plots show grand average ERPs (n=16), with congruent (dashed blue) and incongruent (solid red) conditions for Neuroscan electrodes Cz (top left panel), FC3 (middle left) and FC4 (bottom left), and EPOC+ electrodes F3 (middle right) and F4 (bottom right). Shading indicates standard error of the mean. Temporal clusters at which there was a statistical difference between the conditions are indicated with a solid black line ($p < .05$, after cluster correction for multiple comparisons). The bottom panel illustrates the topographic map of the N400 effect (incongruent minus congruent condition) from 200 to 1000 ms after target onset in the group (n=16). Yellow areas

indicate a more negative-going response for the congruent condition and blue areas indicate a more negative-going response for the incongruent condition. The N400 effect was mainly distributed over the centro-frontal region.

2.3.2.3 Single subject ERP analyses

We then assessed the detection rate of neural responses in individual subjects. We observed reliable N400 effects in one or both of FC3 and FC4 in 56% of the participants with the Neuroscan data and 50% in the EPOC+ data (detection rates are shown in Table 3). We again did not find any significant correlation between the amplitude of the N400 effect and age ($r = -.13$, $p = .63$), PPVT score ($r = .25$, $p = .37$), or K-BIT score ($r = .41$, $p = .37$). The topography of the effect was again variable between participants, ranging from frontal (P1, P11, P14) to parietal locations (P13).

Table 3: Experiment 2 (sentence paradigm) detection rate (% of individuals) of statistically significant N400 effects in each of the three electrodes of interest, and the detection rate in a more lenient assessment where the effect was considered present if it occurred in either one or both of the two frontal electrodes.

Electrode	**FC3**	**FC4**	**Total (FC3 and/or FC4)**	**Cz**
Neuroscan	38%	50%	56%	50%
EPOC+	38%	44%	50%	N/A

Figure 8 shows individual participant waveforms for Neuroscan electrode Cz (Figure 8, first and fourth column), and for Neuroscan (Figure 8, second and fifth column) and EPOC+ electrode FC3 (Figure 8, third and sixth column), which was the site with the highest detection rate. We again found that the topographical distribution of the effect was variable across individuals (Figure 9).

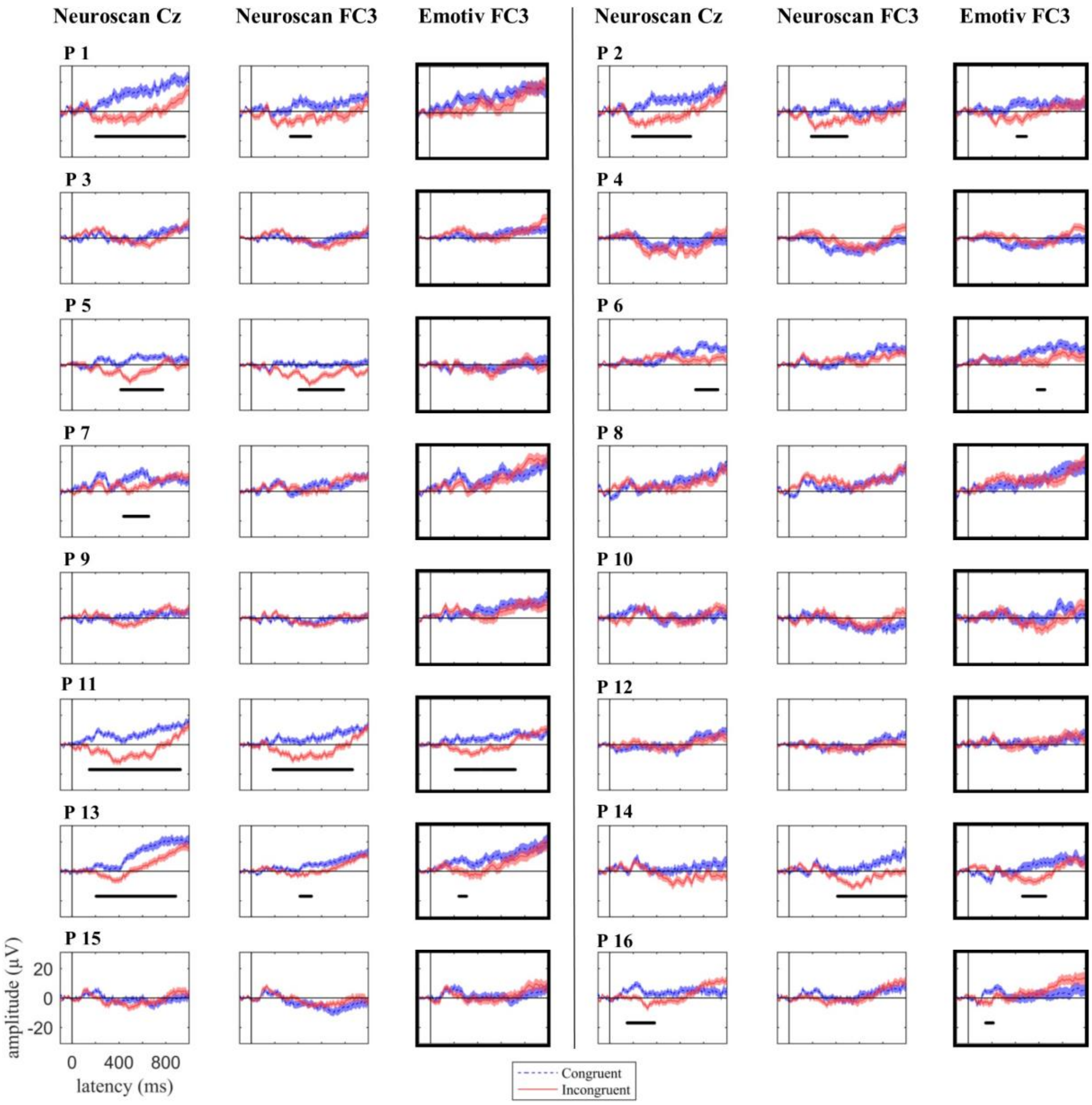


Figure 8. Experiment 2 (sentence paradigm) individual responses to target words in congruent (dashed blue) and incongruent (solid red) sentences for Neuroscan electrode Cz (first and fourth column), and adjacent Neuroscan and EPOC+ electrode FC4, plotted ± standard error (second, third, fifth, and sixth column). Statistical N400 effect is shown as a solid black line. EPOC+ results are outlined in bold. P indicates participant.

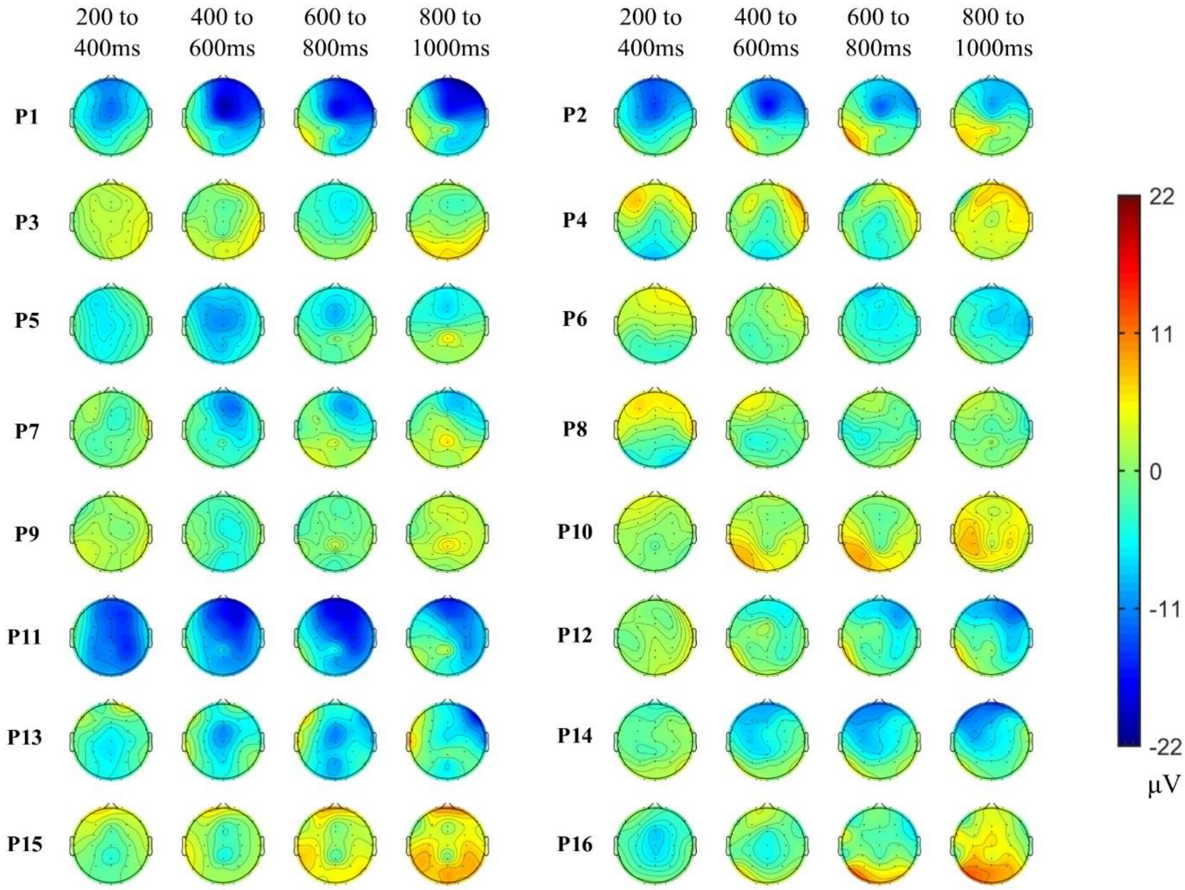


Fig 9. Experiment 2 (sentence paradigm) individual participant topographic maps of the N400 effect (incongruent minus congruent condition) for 200 ms time windows from 200 to 1000 ms after target onset for Neuroscan. Red areas indicate more negative-going response for the congruent condition and blue areas indicate more negative-going response for the incongruent condition. The N400 effect location was variable across individuals. P indicates participant.

2.3.2.4 Comparison of EEG systems

We again compared the responses of the two systems directly using ICC and Spearman's rank correlation and by comparing the area under the difference curve at our two locations of interest (see Table 4). Waveforms for the sentences task across the two systems were qualitatively similar in shape and amplitude, as indicated by positive correlations (Spearman's rho $\geq$ 0.54, ICC $\geq$ 0.23 CI not including zero for any of the comparisons) for both sites and conditions. The area between the related and unrelated curves was numerically larger for Neuroscan than for EPOC+ both in FC3 (125 arbitrary units for EPOC+ versus 208 for

Neuroscan), and FC4 (196 for EPOC+ versus 340 for Neuroscan), but the difference was not significant for either site (FC3: $t_{(15)} = 0.82$, p= .43, FC4: $t_{(15)}=1.44$, p= .17).

Table 4: Experiment 2 (sentence paradigm) mean Spearman' (ρ) and ICC (r) and 95% confidence intervals between waveforms simultaneously recorded with the research (Neuroscan) and gaming (EPOC+) EEG systems for the left (FC3) and right (FC4) frontocentral locations, in the semantically congruent and incongruent conditions. We also present the difference in area between the two conditions for Neuroscan and EPOC+ averaged across subjects with 95% confidence intervals.

Condition	**Coeff.**	**Location**	
		Left frontocentral (FC3)	Right frontocentral (FC4)
Congruent	ρ	0.57 [0.38, 0.76]	0.54 [0.32, 0.77]
	r	0.23 [0.01, 0.46]	0.30 [0.02, 0.58]
Incongruent	ρ	0.71 [0.58, 0.85]	0.72 [0.58, 0.87]
	r	0.53 [0.34, 0.72]	0.62 [0.44, 0.80]
Area between curves (μV)		EPOC+ : 125 [-39, 290] Neuroscan : 208 [42, 373]	EPOC+ : 196 [52, 339] Neuroscan : 340 [138, 541]

2.3.2.5 Decoding analyses

We again tested whether our detection rate would improve by combining data across sensors using MVPA. At the group level, we saw significant decoding of semantic context (congruent or incongruent sentence frames) in the Neuroscan data from 162 ms to 1000 ms (Fig 10, top panel). We also decoded the semantic category from the EPOC+ group level data (Figure 10, bottom panel), at several clusters of time points between 414 ms and 860 ms. Therefore, MVPA decoding matched the time course of univariate decoding seen at the group level in Neuroscan and EPOC+ data (Fig 7).

At the individual level, we decoded semantic category from the Neuroscan data in all but two participants (88% detection, Fig 11, purple traces). This was a marked improvement relative to the univariate detection rate at individual channels (38-50%, above). However, for EPOC+, the classifier only detected a significant effect in 4 of the 16 participants (25% detection rate, Fig 11, yellow traces). We examined whether the superior decoding of Neuroscan could be attributed to the larger number of electrodes (33 in Neuroscan versus 12 in EPOC+) by performing an additional MVPA analysis on the Neuroscan data using only the 12 electrodes closest to the EPOC+ ones. However, in these conditions, the MVPA again performed well for the Neuroscan data, identifying statistical differences in the signal for 88% of participants. Thus, the Neuroscan data appear to be more suitable for decoding than the EPOC+ data and the difference cannot be attributed to the difference in the number or location of electrodes.

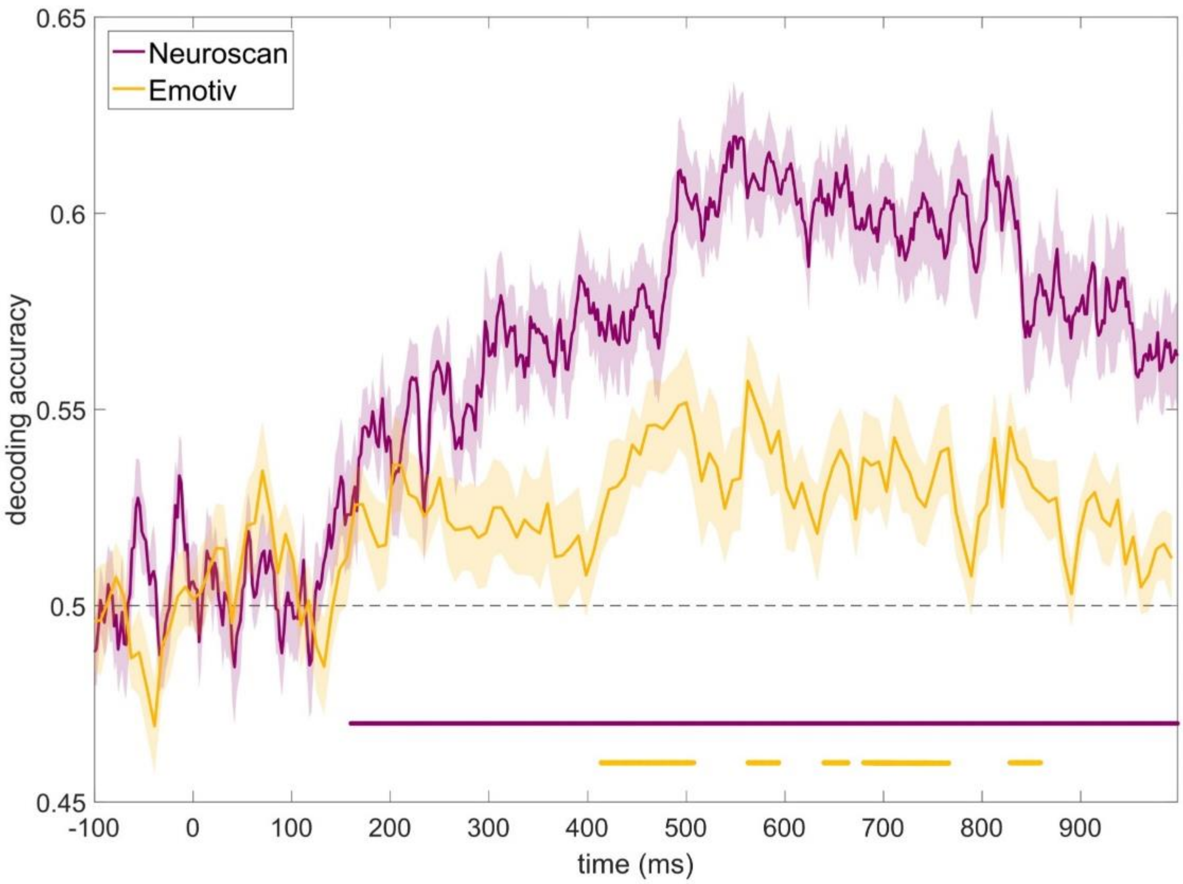


Figure 10. Experiment 2 (sentence paradigm) grand average decoding accuracy for discriminating between identical target words presented in congruent and incongruent conditions at each time point for Neuroscan (purple) and EPOC+ (yellow) data. Shading indicates standard error of the mean. Clusters of significant decoding are shown by a purple (Neuroscan) and yellow (EPOC+) horizontal line. Decoding accuracy was significantly above chance for Neuroscan in a cluster from 162 ms to 1000 ms, and for EPOC+ at several clusters of time points between 414 ms and 860 ms.

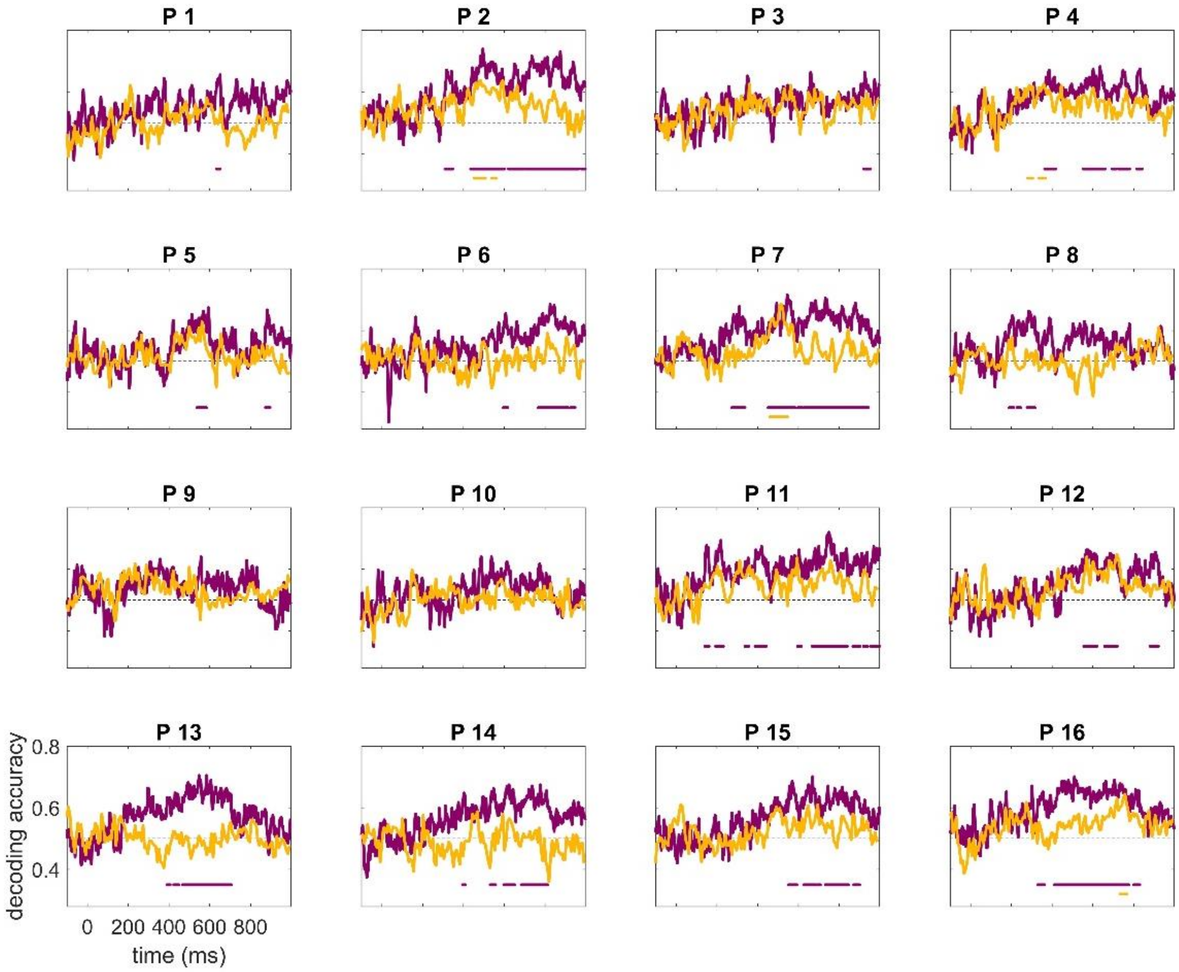


Figure 11. Experiment 2 (sentence paradigm) individual decoding accuracy for discriminating between congruent and incongruent conditions over time for Neuroscan (purple) and EPOC+ (yellow) data. Time points with accuracy significantly above chance (temporal cluster correction, $p < .05$) are shown as solid horizontal lines (in purple for Neuroscan, and yellow for EPOC+). Semantic condition could be decoded in 88% of the participants' Neuroscan data and in 25% of the participants' EPOC+ data. P indicates participant.

2.3.3 Experiment 2 summary

Using a paradigm that contrasted congruent and incongruent sentences, we elicited an N400 effect in a group of children, and in up to 50% of the participants using Neuroscan and 44% using EPOC+. Using multivariate analyses, we decoded the semantic condition in all but two individuals (88%) using the Neuroscan data, and in four individuals (25%) using the EPOC+ data. This suggests that multivariate analyses, which take the pattern across all electrodes into account, may be a sensitive way to detect effects in individuals, accounting for the topographic variability of effects across individuals. However, decoding was, if anything, slightly worse than univariate analyses for the EPOC+ system (25% with MVPA, as opposed to 44% with univariate analyses), suggesting that MVPA may be more sensitive to the differences in data quality from research-grade and gaming EEG systems than traditional analyses.

2.4 General Discussion

Motivated by a recent report of preserved spoken word processing in minimally-verbal autistic children (Cantiani et al., 2014; DiStefano et al., 2019), we examined the suitability of two N400 paradigms to detect lexico-semantic processing in individuals. We asked whether neural responses to identical spoken words differed when they were presented in the context of a congruent versus incongruent single word prime (Experiment 1) or sentence frame (Experiment 2). We compared the signal recorded by a low-cost gaming EEG system, Emotiv EPOC+, to that recorded by a traditional research-grade EEG system, Neuroscan SynAmps2. We further used two approaches to evaluate the electrophysiological response: univariate analysis of the N400 effect and multivariate decoding. We found that the N400 effect could be observed at the group level and detected at an individual level in about half of the children, using either paradigm (word pairs or sentences) and using either EEG system. However, the sentences paradigm, in which children counted the incongruent sentences, was more promising, with the best individual participant detection rate, 88%, given by multivariate analyses of the Neuroscan data.

Despite an extensive body of literature on the N400 effect, only a few studies have carried out statistical testing at the individual subject level or report individual participants' waveforms. In our data, statistically significant traditional N400 effects were present in about half of our participants using either normatively-associated word pairs or sentences to induce a violation of lexical-semantic predictions. This may seem low, given that the N400 is widely reported to be a large and robust effect (e.g. Kutas and Federmeier, 2011), but it is in fact similar to statistical detection rates in adults using similar tasks (Cruse et al., 2014). This raises an important question about how robust and prevalent individual N400 effects actually are. At minimum, it points to large inter-individual variability in response. There are several possible explanations. It is possible that we simply lacked sensitivity to detect consistent neural effects within testing constraints (i.e. we needed more data to detect an effect in every individual). This is difficult to quantify because traditional power analyses fail to account for the type of cluster-based multiple-comparisons correction we used. It is also possible that lexico-semantic abnormalities were not treated in similar cognitive ways in different individuals. For example, individuals may vary in the extent to which they attempt to integrate the unpredicted target word into the semantic context or be differentially sensitive to the wider experiment in which predictions were violated 50% of the time. In addition, or instead, it is possible that individuals may have similar cognitive reactions but the neural substrates supporting this processing vary. Previous studies have reported that N400 effects vary with age (e.g. Atchley et al., 2006; Holcomb et al., 1992; Juottonen et al., 1996) and non-verbal intelligence (Jaušovec and Jaušovec, 2000) but we did not find any systematic effects of this in our small sample. We also did not find any association with vocabulary knowledge, in line with previous reports (Henderson et al., 2011b). Our findings thus do not provide a simple explanation for inter-individual variability in the N400.

The details of the stimuli and the participants' task seem likely to affect the size of the N400 (e.g. Bentin et al., 1993; Chwilla et al., 1995; Ortega et al., 2008; Perrin and García-Larrea, 2003) and therefore, presumably, individual-subject detection rate. For example, Cruse et al. (2014) reported superior individual-subject detection rates for active and overt tasks relative to passive ones. Cruse et al. also reported superior rates for normatively associated word pairs (50%) than for sentences (17%) when heard passively. However, we found that the response to the two stimulus sets were comparable, or slightly better for sentences, when we considered an effect recorded from either of our sites of interest (47% in word pairs, and 56% in sentences). Our relatively high detection rate, particularly for sentences, is potentially attributable to our engaging covert tasks which required participants to attend to the sematic relatedness of the sentence and target word and encouraged participants to remain attentive throughout the experiment. Although for future applications to clinical populations a completely passive task would be preferable, here we considered a trade-off between the likely sensitivity of the task with how demanding it was. We considered that minimally-verbal autistic individuals with good receptive language may be able to follow instructions covertly, even if they cannot do so overtly. We do acknowledge, however, that this approach risks potentially missing effects in children that are be unable to follow these instructions despite preserved linguistic processing.

Our individual-participant data also emphasise the variability in topography and time course of individual N400 responses (previously suggested by Henderson et al. (2011)), which may be particularly important to consider when testing children and special populations, and might necessitate different analysis approaches. For example, in those participants not showing a reliable N400 effect, it is possible that univariate analyses failed to detect more subtle changes in the neural responses or that N400-like responses occurred in EEG channels that were not analysed. To overcome this, we employed MVPA to integrate information from across all the sensors. MVPA provides a sensitive method of assessing effects occurring at any location or combination of locations without increasing type I error (only one statistical test is performed on the pattern of response across all sensors) at the cost of specificity regarding where the effect arises from (e.g. Grootswagers et al., 2017). We ran our individual-subject multivariate analyses at every time point (with correction for multiple comparisons) allowing us to be sensitive to individual variability in both topology and time course.

With this approach, semantic condition could be decoded from the Neuroscan data in all but two individuals (88% detection rate, Experiment 2). In our pursuit of an individualised neural marker of language comprehension, this is encouraging. It also confirms the intuition that considering the signal recorded by all electrodes can be more powerful than restricting analysis to one or a few. Even though the N400 effect is well established as having a centro-parietal topography at the group level, our data suggest that there is variation in topology at the individual subject level, and we anticipate that the variation may be even greater amongst minimally-verbal autistic children. Moreover, we do not want to be too restrictive *a priori* because ultimately, any statistically reliable decodable difference between the brain responses to identical auditory tokens presented in different semantic contexts will be meaningful for our purpose, independently of *where* (and to a certain extent *when*) this effect occurs. Following this logic, it is also possible that the higher MVPA detection rate obtained with our sentence paradigm compared to the word pairs paradigm reflects neural processes in response to the task (participants counting the incongruent sentences). In this case, we would not be detecting the brain responses to semantic condition *per se* but rather the brain responses to the task. For our purposes of detecting intact receptive language however, this would still be meaningful as it indicates differential brain processing between the conditions.

Finally, we assessed the sensitivity of low-cost portable EEG technology, Emotiv EPOC+ for detecting N400 effects. EPOC+ was able to record N400 effects at the group level and showed a similar detection rate to the research system at the individual subject level. Moreover, when we formally compared the data between systems, we found that the waveforms recorded by EPOC+ were similar in shape and amplitude to those recorded by the research system Neuroscan. This makes it a promising avenue for research, particularly where full EEG setups are unfeasible. Our result adds to the previous literature demonstrating the usability of the EPOC+ to record ERPs (Badcock et al., 2015, 2013; de Lissa et al., 2015; Duvinage et al., 2013) and shows that late ERPs such as the N400 can also be recorded with a low-cost, portable system. Nonetheless the effects recorded with EPOC+ tended to be numerically lower amplitude than for Neuroscan, and detection rate did not improve with MVPA meaning that the best detection rate for Neuroscan (88%, MVPA, Exp 2) was far above the best with EPOC+ (50%, univariate, Exp 2).

A few limitations of the EPOC+ system could be mitigated in future research. First, due to a limitation of the software (Testbench software), the impedance of the EPOC+ electrodes was not assessed precisely and we were only able to ensure that it remained lower than 220 kΩ. It is thus likely that impedances were much higher for EPOC+ than for Neuroscan (for which impedances were adjusted to < 5 kΩ), although the differences between the systems (e.g. active versus passive electrodes respectively) makes differences in impedance difficult to interpret. Use of a more precise measure of impedance to ensure the best possible connection in every participant would be beneficial. Second, the N400 effect was centrally distributed on the scalp, at least at the group level, in accordance with previous literature (Friedrich and Friederici, 2004; Kutas and Federmeier, 2011), and the EPOC+ headset does not have any central sensors. In future research, researchers should consider wiring an additional electrode in a centro-parietal location where the effects were the largest. It is also important to consider the trade-off between the sensitivity we want to achieve (in which case Neuroscan may be more suitable), and the level of portability and accessibility (in which case EPOC+ is more suitable). In particular, when testing children or special populations, the possibility of recording EEG outside of the lab, with an easy and fast setup procedure, may outweigh EPOC+'s apparently lower sensitivity in the context of multivariate analyses, or may motivate using EPOC+ in cases where it would be impossible to obtain data on a research-grade system.

Taken together, our results indicate that it may be possible to index lexical-semantic processing in individual children using EEG. However, contrasting the results we obtained from different paradigms, EEG systems, and analysis methods, several trade-offs have to be considered. The best individual-subject detection rate (88%) was yielded from MVPA of research-grade EEG data in response to identical spoken words presented in the context of congruent and incongruent sentences and a sentence counting task (Experiment 2). Our variable individual subject data emphasise the importance of analysing and reporting individual ERP results, in addition to the grand average data, to illustrate the variability in the presence, location, and timing of ERPs.

2.5 Conclusion

In this study, we set out to establish the reliability of discriminative brain signals in response to lexico-semantic violations in individual children. We devised two paradigms that used identical spoken language tokens presented in congruent or incongruent lexical-semantic contexts. Additionally, with clinical applications in mind, we tested whether we could use a portable and low-cost device, Emotiv EPOC+, to measure these neural signals. Our results suggest that large inter-individual variability in the neural signatures of lexico-semantic processing exist, even in the neurotypical population, and (in our small sample) this variability was not explained by age, vocabulary knowledge, or non-verbal intelligence. Despite this variability, we were able to replicate group-level N400 effects in neurotypical children using both the EPOC+ and the research-grade Neuroscan system. At an individual level, an N400 effect was evoked in about half of neurotypical children using either Neuroscan or EPOC+ systems. MVPA analyses allowed us to reach near-perfect detection rate with Neuroscan EEG, with only two participants not showing a reliable electrophysiological signature to semantic anomalies in sentences. These results give us a basis for future research in which we can test for receptive language processing in people who are unable to communicate.

2.6 References

Audacity(R) software is copyright (c) 1999-2014 Audacity Team. The name Audacity(R) is a registered trademark of Dominic Mazzoni.

Atchley, R.A., Rice, M.L., Betz, S.K., Kwasny, K.M., Sereno, J.A., Jongman, A., 2006. A comparison of semantic and syntactic event related potentials generated by children and adults. Brain Lang., Language Comprehension across the Life Span 99, 236–246. https://doi.org/10.1016/j.bandl.2005.08.005

Badcock, N.A., Mousikou, P., Mahajan, Y., de Lissa, P., Thie, J., McArthur, G., 2013. Validation of the Emotiv EPOC® EEG gaming system for measuring research quality auditory ERPs. PeerJ 1, e38. https://doi.org/10.7717/peerj.38

Badcock, N.A., Preece, K.A., Wit, B. de, Glenn, K., Fieder, N., Thie, J., McArthur, G., 2015. Validation of the Emotiv EPOC EEG system for research quality auditory event-related potentials in children. PeerJ 3, e907. https://doi.org/10.7717/peerj.907

Barham, M.P., Clark, G.M., Hayden, M.J., Enticott, P.G., Conduit, R., Lum, J.A.G., 2017. Acquiring research-grade ERPs on a shoestring budget: A comparison of a modified Emotiv and commercial SynAmps EEG system. Psychophysiology 54, 1393–1404. https://doi.org/10.1111/psyp.12888

Bentin, S., Kutas, M., Hillyard, S.A., 1993. Electrophysiological evidence for task effects on semantic priming in auditory word processing. Psychophysiology 30, 161–169. https://doi.org/10.1111/j.1469-8986.1993.tb01729.x

Beukema, S., Gonzalez-Lara, L.E., Finoia, P., Kamau, E., Allanson, J., Chennu, S., Gibson, R.M., Pickard, J.D., Owen, A.M., Cruse, D., 2016. A hierarchy of event-related potential markers of auditory processing in disorders of consciousness. NeuroImage Clin. 12, 359–371. https://doi.org/10.1016/j.nicl.2016.08.003

Block, C.K., Baldwin, C.L., 2010. Cloze probability and completion norms for 498 sentences: Behavioral and neural validation using event-related potentials. Behav. Res. Methods 42, 665–670. https://doi.org/10.3758/BRM.42.3.665

Boersma, P., 2001. Praat, a system for doing phonetics by computer. Glot Int.

Borovsky, A., Elman, J.L., Kutas, M., 2012. Once is Enough: N400 Indexes Semantic Integration of Novel Word Meanings from a Single Exposure in Context. Lang. Learn. Dev. Off. J. Soc. Lang. Dev. 8, 278–302. https://doi.org/10.1080/15475441.2011.614893

Brainard, D.H., 1997. The Psychophysics Toolbox. Spat. Vis. 10, 433–436.

Cantiani, C., Choudhury, N.A., Yu, Y.H., Shafer, V.L., Schwartz, R.G., Benasich, A.A., 2016. From Sensory Perception to Lexical-Semantic Processing: An ERP Study in Non-Verbal Children with Autism. PLOS ONE 11, e0161637. https://doi.org/10.1371/journal.pone.0161637

Chwilla, D.J., Brown, C.M., Hagoort, P., 1995. The N400 as a function of the level of processing. Psychophysiology 32, 274–285. https://doi.org/10.1111/j.1469-8986.1995.tb02956.x

Coderre, E.L., Chernenok, M., O'Grady, J., Bosley, L., Gordon, B., Ledoux, K., 2019. Implicit Measures of Receptive Vocabulary Knowledge in Individuals With Level 3 Autism. Cogn. Behav. Neurol. 32, 95. https://doi.org/10.1097/WNN.0000000000000194

Cruse, D., Beukema, S., Chennu, S., Malins, J.G., Owen, A.M., McRae, K., 2014. The reliability of the N400 in single subjects: Implications for patients with disorders of consciousness. NeuroImage Clin. 4, 788–799. https://doi.org/10.1016/j.nicl.2014.05.001

Davis, C.J., 2005. N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics. Behav. Res. Methods 37, 65–70. https://doi.org/10.3758/BF03206399

de Lissa, P., Sörensen, S., Badcock, N., Thie, J., McArthur, G., 2015. Measuring the face-sensitive N170 with a gaming EEG system: A validation study. J. Neurosci. Methods 253, 47–54. https://doi.org/10.1016/j.jneumeth.2015.05.025

DiStefano, C., Senturk, D., Spurling Jeste, S., 2019. ERP Evidence of Semantic Processing in Children with ASD. Dev. Cogn. Neurosci. 100640. https://doi.org/10.1016/j.dcn.2019.100640

Dunn, L.M., Dunn, D.M., 2007. PPVT-4: Peabody picture vocabulary test. Pearson Assess.

Duvinage, M., Castermans, T., Petieau, M., Hoellinger, T., Cheron, G., Dutoit, T., 2013. Performance of the Emotiv Epoc headset for P300-based applications. Biomed. Eng. OnLine 12, 56. https://doi.org/10.1186/1475-925X-12-56

Elsawy, A.S., Eldawlatly, S., Taher, M., Aly, G.M., 2014. Performance analysis of a Principal Component Analysis ensemble classifier for Emotiv headset P300 spellers, in: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Presented at the 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 5032–5035. https://doi.org/10.1109/EMBC.2014.6944755

Friedrich, M., Friederici, A.D., 2005. Semantic sentence processing reflected in the event-related potentials of one- and two-year-old children. Neuroreport 16, 1801–1804.

Friedrich, M., Friederici, A.D., 2004. N400-like Semantic Incongruity Effect in 19-Month-Olds: Processing Known Words in Picture Contexts. J. Cogn. Neurosci. 16, 1465–1477. https://doi.org/10.1162/0898929042304705

Giacino, J.T., Smart, C.M., 2007. Recent advances in behavioral assessment of individuals with disorders of consciousness: Curr. Opin. Neurol. 20, 614–619. https://doi.org/10.1097/WCO.0b013e3282f189ef

Grootswagers, T., Wardle, S.G., Carlson, T.A., 2017. Decoding dynamic brain patterns from evoked responses: A tutorial on multivariate pattern analysis applied to time-series neuroimaging data. J. Cogn. Neurosci. 29, 677–697. https://doi.org/10.1162/jocn_a_01068

Guthrie, D., Buchwald, J.S., 1991. Significance testing of difference potentials. Psychophysiology 28, 240–244.

Harrison, A.H., Connolly, J.F., 2013. Finding a way in: a review and practical evaluation of fMRI and EEG for detection and assessment in disorders of consciousness. Neurosci. Biobehav. Rev. 37, 1403–1419. https://doi.org/10.1016/j.neubiorev.2013.05.004

Haynes, J.-D., 2015. A Primer on Pattern-Based Approaches to fMRI: Principles, Pitfalls, and Perspectives. Neuron 87, 257–270. https://doi.org/10.1016/j.neuron.2015.05.025

Hebart, M.N., Baker, C.I., 2018. Deconstructing multivariate decoding for the study of brain function. NeuroImage 180, 4–18. https://doi.org/10.1016/j.neuroimage.2017.08.005

Henderson, L.M., Baseler, H.A., Clarke, P.J., Watson, S., Snowling, M.J., 2011a. The N400 effect in children: Relationships with comprehension, vocabulary and decoding. Brain Lang. 117, 88–99. https://doi.org/10.1016/j.bandl.2010.12.003

Henderson, L.M., Baseler, H.A., Clarke, P.J., Watson, S., Snowling, M.J., 2011b. The N400 effect in children: Relationships with comprehension, vocabulary and decoding. Brain Lang. 117, 88–99. https://doi.org/10.1016/j.bandl.2010.12.003

Holcomb, P.J., Coffey, S.A., Neville, H.J., 1992. Visual and auditory sentence processing: A developmental analysis using event-related brain potentials. Dev. Neuropsychol. 8, 203–241. https://doi.org/10.1080/87565649209540525

Jaušovec, N., Jaušovec, K., 2000. Correlations between ERP parameters and intelligence: a reconsideration. Biol. Psychol. 55, 137–154. https://doi.org/10.1016/S0301-0511(00)00076-4

Juottonen, K., Revonsuo, A., Lang, H., 1996. Dissimilar age influences on two ERP waveforms (LPC and N400) reflecting semantic context effect. Cogn. Brain Res. 4, 99–107. https://doi.org/10.1016/0926-6410(96)00022-5

Kaufman, A.S., Kaufman, N.L., 2004. Kaufman brief intelligence test KBIT 2 ; manual.

Kiang, M., Patriciu, I., Roy, C., Christensen, B.K., Zipursky, R.B., 2013. Test-retest reliability and stability of N400 effects in a word-pair semantic priming paradigm. Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol. 124, 667–674. https://doi.org/10.1016/j.clinph.2012.09.029

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., Broussard, C., 2007. What's new in psychtoolbox-3. Perception 36, 1–16.

Kuperman, V., Stadthagen-Gonzalez, H., Brysbaert, M., 2012. Age-of-acquisition ratings for 30,000 English words. Behav. Res. Methods 44, 978–990. https://doi.org/10.3758/s13428-012-0210-4

Kutas, M., 1993. In the company of other words: Electrophysiological evidence for single-word and sentence context effects. Lang. Cogn. Process. 8, 533–572. https://doi.org/10.1080/01690969308407587

Kutas, M., A. Hillyard, S., 1980. Kutas, M. & Hillyard, S. A. Reading senseless sentences: brain potentials reflect semantic incongruity. Science 207, 203-205. Science 207, 203–5. https://doi.org/10.1126/science.7350657

Kutas, M., Federmeier, K.D., 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). Annu. Rev. Psychol. 62, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods 164, 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Oosterhof, N.N., Connolly, A.C., Haxby, J.V., 2016. CoSMoMVPA: Multi-Modal Multivariate Pattern Analysis of Neuroimaging Data in Matlab/GNU Octave. Front. Neuroinformatics 10, 27. https://doi.org/10.3389/fninf.2016.00027

Ortega, R., López, V., Aboitiz, F., 2008. Voluntary modulations of attention in a semantic auditory-visual matching Task: an ERP study. Biol. Res. 41, 453–460. https://doi.org/10.4067/S0716-97602008000400010

Pelli, D.G., 1997. The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat. Vis. 10, 437–442.

Perrin, F., García-Larrea, L., 2003. Modulation of the N400 potential during auditory phonological/semantic interaction. Cogn. Brain Res. 17, 36–47. https://doi.org/10.1016/S0926-6410(03)00078-8

Rämä, P., Sirri, L., Serres, J., 2013. Development of lexical–semantic language system: N400 priming effect for spoken words in 18- and 24-month old children. Brain Lang. 125, 1–10. https://doi.org/10.1016/j.bandl.2013.01.009

Rohaut, B., Faugeras, F., Chausson, N., King, J.-R., Karoui, I.E. Cohen, L., Naccache, L., 2015. Probing ERP correlates of verbal semantic processing in patients with impaired consciousness. Neuropsychologia 66, 279–292. https://doi.org/10.1016/j.neuropsychologia.2014.10.014

Smith, S.M., Nichols, T.E., 2009. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. NeuroImage 44, 83–98. https://doi.org/10.1016/j.neuroimage.2008.03.061

Steinhauer, K., Alter, K., Friederici, A.D., 1999. Brain potentials indicate immediate use of prosodic cues in natural speech processing. Nat. Neurosci. 2, 191–196. https://doi.org/10.1038/5757

Steinhauer, K., Friederici, A.D., 2001. Prosodic boundaries, comma rules, and brain responses: the closure positive shift in ERPs as a universal marker for prosodic phrasing in listeners and readers. J. Psycholinguist. Res. 30, 267–295.

Tager-Flusberg, H., Kasari, C., 2013. Minimally Verbal School-Aged Children with Autism Spectrum Disorder: The Neglected End of the Spectrum. Autism Res. 6, 468–478. https://doi.org/10.1002/aur.1329

Thie, J., 2013. A wireless marker system to enable evoked potential recordings using a wireless EEG system (EPOC) and a portable computer (No. e32v1). PeerJ PrePrints. https://doi.org/10.7287/peerj.preprints.32v1

Torkildsen, J. von K., Syversen, G., Simonsen, H.G., Moen, I., Lindgren, M., 2007. Electrophysiological correlates of auditory semantic priming in 24-month-olds. J. Neurolinguistics 20, 332–351. https://doi.org/10.1016/j.jneuroling.2007.02.003

Vos, M.D., Kroesen, M., Emkes, R., Debener, S., 2014. P300 speller BCI with a mobile EEG system: comparison to a traditional amplifier. J. Neural Eng. 11, 036008. https://doi.org/10.1088/1741-2560/11/3/036008

Yau, S.H., McArthur, G., Badcock, N.A., Brock, J., 2015. Case study: auditory brain responses in a minimally verbal child with autism and cerebral palsy. Front. Neurosci. 9. https://doi.org/10.3389/fnins.2015.00208

2.7 Supplementary materials

Table S1: Stimuli used in Experiment 1: normatively associated word pairs

Unrelated prime	Related prime	Target
House	Leg	Arm
Suds	Crafts	Art
Knob	Front	Back
Sap	Spine	Back
Meow	Bounce	Ball
Pork	Sand	Beach
Spoon	Chirp	Bird
Seat	Nest	Bird
Quack	Row	Boat
Arm	Ship	Boat
Filth	Text	Book
Cat	Scout	Boy
Soil	Girl	Boy
Oak	Groom	Bride
Stone	Comb	Brush
South	Meow	Cat
Job	Dog	Cat
East	Seat	Chair
Rake	Wear	Clothes
Nest	Cob	Corn
Lamp	Moo	Cow
Sand	Calf	Cow
Front	Mum	Dad
King	Soil	Dirt

Quiz	Filth	Dirt
Bulb	Pet	Dog
Ship	Cat	Dog
Calf	Hinge	Door
Bride	Knob	Door
Girl	Quack	Duck
Groom	Bait	Fish
Boy	Tile	Floor
Queen	Spoon	Fork
Moo	Back	Front
Clock	Boy	Girl
Bed	Vine	Grape
Text	Bride	Groom
Tile	Glove	Hand
Chirp	Cap	Hat
Cob	House	Home
Scout	Honk	Horn
Thigh	Queen	King
Wear	Rake	Leaves
Cap	Arm	Leg
Nap	Thigh	Leg
Mum	Bulb	Light
Pet	Lamp	Light
Honk	Dad	Mum
Black	South	North
Row	Ink	Pen
Glove	Pork	Pig
Dad	Stone	Rock

Gums	Socks	Shoes
Ink	Bed	Sleep
Vine	Nap	Sleep
Crafts	Suds	Soap
Comb	North	South
Socks	Gums	Teeth
Spine	Quiz	Test
Leg	Clock	Time
Bait	Sap	Tree
Hinge	Oak	Tree
Dog	East	West

Table S2: Stimuli used in Experiment 2: Congruent and incongruent sentences

Incongruent sentence frame	**Congruent sentence frame**	**Target**
At night the old woman locked the	She plays the guitar so she joined the	Band
I roasted marshmallows over the	She saved some money in her piggy-	Bank
In the quiet cinema the man's phone	When the thief entered the house the dog	Barked
The children ran outside to	In spring flowers	Bloom
She jumped in the pool and made a big	The student went to the library to read a	Book
His father is the coach of the football	She went to the bakery for a loaf of	Bread
He washed his hands with water and	For his wife's birthday he baked a	Cake
In autumn leaves fall off the	There were candles on the birthday	Cake
Mother hung the painting up on the	The farmer spent the morning milking the	Cows
She is afraid of the sea because she cannot	The movie was so sad it made me	Cry
In spring flowers	When babies are hungry they often	Cry

I put the kettle on to make a hot cup of	The dentist says to brush your teeth twice a	Day
She washed the dirty dishes in the	I prefer cats rather than	Dogs
The clouds are high up in the	I had no key to open the	Door
Most cats see very well at	At night the old woman locked the	Door
Today the baby spoke his first	She got out of the car and closed the	Door
The letter was sent by	For breakfast he ate scrambled	Eggs
There were candles on the birthday	She put on her sunglasses to protect her	Eyes
She went to the bakery for a loaf of	They raised pigs on their	Farm
The hungry baby wanted to drink	I roasted marshmallows over the	Fire
I had no key to open the	He went to the lake to catch	Fish
She waited for the phone to	The baby bird was ready to learn to	Fly
She accidentally tripped and fell down the	It was windy enough to fly a	Kite
He took his dog out for a	At dinner he cut his steak with a	Knife
She wore a necklace around her	The hungry baby wanted to drink	Milk
He looked up at night to see a sky full of	He was tired and in a bad	Mood
When driving you should keep your eyes on the	To hang the picture you need a hammer and	Nail
She saved some money in her piggy-	I could not remember his	Name
The princess may someday become a	She wore a necklace around her	Neck
He cannot post the letter without attaching a	The baby birds were in the	Nest
The student went to the library to read a	Most cats see very well at	Night
I had no umbrella as it began to	The children ran outside to	Play
For breakfast he ate scrambled	The letter was sent by	Post
At dinner he cut his steak with a	The princess can only marry a	Prince
The baby birds were in the	The princess may someday become a	Queen
The movie was so sad it made me	I had no umbrella as it began to	Rain
When the thief entered the house the dog	In the quiet cinema the man's phone	Rang

The baby bird was ready to learn to	She waited for the phone to	Ring
The dentist says to brush your teeth twice a	When driving you should keep your eyes on the	Road
It was windy enough to fly a	She washed the dirty dishes in the	Sink
I prefer cats rather than	The clouds are high up in the	Sky
She put on her sunglasses to protect her	He washed his hands with water and	Soap
She forgot her watch so she asked for the	She jumped in the pool and made a big	Splash
He went to the lake to catch	She accidentally tripped and fell down the	Stairs
THE princess can only marry a	He cannot post the letter without attaching a	Stamp
She got out of the car and closed the	He looked up at night to see a sky full of	Stars
When babies are hungry they often	She is afraid of the sea because she cannot	Swim
The genie promised to grant the man one	I put the kettle on to make a hot cup of	Tea
They raised pigs on their	His father is the coach of the football	Team
She plays the guitar so she joined the	The dentist opened his mouth to check his	Teeth
I could not remember his	She forgot her watch so she asked for the	Time
The farmer spent the morning milking the	In autumn leaves fall off the	Trees
The dentist opened his mouth to check his	He took his dog out for a	Walk
He was tired and in a bad	Mother hung the painting up on the	Wall
To hang the picture you need a hammer and	The genie promised to grant the man one	Wish
For his wife's birthday he baked a	Today the baby spoke his first	Word

Chapter 3

Even neurotypical children are heterogeneous: using multivariate decoding to improve individual subject analysis of lexico-semantic EEG data

Selene Petit[*,1], Nicholas A. Badcock[1], Tijl Grootswagers[1,2,] and Alexandra Woolgar[1,3]

[1] Department of Cognitive Science, Macquarie University, Australia

[2] School of Psychology, University of Sydney, Australia

[3] Medical Research Council (UK), Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, UK

* Corresponding author at: Australian Hearing Hub, Level 3, 16 University Avenue, Macquarie University, NSW 2109, Australia.

This chapter is currently under review at *scientific reports*:

Petit S., Badcock N. A., Grootswagers T., Woolgar A. (submitted). Even neurotypical children are heterogeneous: using multivariate decoding to improve individual subject analysis of lexico-semantic EEG data.

Abstract

In conditions such as minimally-verbal autism, standard behavioural assessments of language comprehension are often unreliable, so recent efforts have focused on developing a neural test of language processing. These tests have focused on the N400, a robust marker of lexical-semantic violation at the group level. However, homogeneity of the N400 response in individual neurotypical children has not been established, and, given the known heterogeneity within the autistic population, it is crucial that any test is sensitive in individuals. Here, we presented 20 children with visual animations that were paired with spoken sentences ending in either a congruent or an incongruent final word. We measured their neural response to the final word using electroencephalography (EEG). Despite robust group-level responses, we found high inter-individual variability in their response to lexico-semantic anomalies. To overcome this, we analysed our data using temporally and spatially unconstrained multivariate pattern analyses (MVPA), supplemented by descriptive analyses to examine the time course, topography, and strength of the effect. We conclude that even neurotypical children exhibit heterogenous responses to lexical-semantic violation, implying that any application to heterogenous disorders such as autism will require individual-subject analyses that are robust to variation in topology and time course of neural responses.

3.1 Introduction

In a variety of conditions, the absence of spoken language does not necessarily reflect an absence of comprehension. For populations that are unable to overtly communicate, such as people in a vegetative state or with cerebral palsy or non-verbal autism (Geytenbeek et al., 2010; Giacino & Smart, 2007; Tager-Flusberg & Kasari, 2013), an accurate assessment of cognitive abilities is essential to provide appropriate support and enhance wellbeing. Yet in those cases, traditional testing materials often fail to capture the full cognitive and language abilities of these individuals. This is due to the various constraints imposed by the standardized testing environment and materials, and the social constraints associated with the examiner-examinee interactions (Kasari, Brady, Lord, & Tager-Flusberg, 2013; Plesa Skwerer, Jordan, Brukilacchio, & Tager-Flusberg, 2016). To assess cognitive abilities of these populations, we need to develop objective, fast, and reliable measures.

One possible avenue, which bypasses the need for overt communication, is to measure neural responses using event-related potentials (ERPs) measured with electroencephalography (EEG). Recent work has used this approach to assess covert cognitive abilities in non-communicative populations (see a recent review by Harrison & Connolly, 2013). In particular, language processing has been studied using EEG including in patients with disorder of consciousness (Hinterberger, Birbaumer, & Flor, 2005; Kotchoubey et al., 2005), schizophrenia (Mathalon, Faustman, & Ford, 2002; Sharma et al., 2017), and autistic individuals (Coderre, Chernenok, Gordon, & Ledoux, 2017; Pijnacker, Geurts, van Lambalgen, Buitelaar, & Hagoort, 2010; Wang, Yang, Liu, Shao, & Jackson, 2017). However, in these cases, EEG research has focused on the population-level, with minimal data reported on an individual-participant basis. Yet, in order to design a clinical test of language comprehension, and particularly given the known heterogeneity in developmental disorders such as autism, it is critical to use paradigms and methods that reliably elicit meaningful brain signals in *individuals*. In this study, we developed and assessed a new EEG paradigm to measure language comprehension in individual children. We report the heterogeneity of neural responses to semantic anomalies of speech using different data analysis techniques.

The N400 ERP is a well-studied component that indexes the semantic integration of words into their context. The N400 effect is a more negative-going electrical potential when participants read or listen to words presented in an incongruent semantic context, relative to a congruent semantic context (Kutas & Federmeier, 2011). As such, it seems to be a well-suited candidate for the assessment of lexico-semantic processing. The N400 has been widely reported in groups of typical adults, children, and special populations (see Kutas & Federmeier, 2011 for a review), including recently in autistic children (Cantiani et al., 2016; Coderre et al., 2019; DiStefano, Senturk, & Spurling Jeste, 2019). However, despite the recent interest in using N400 as an assessment tool, individual subject level analyses are rarely reported. Additionally, studies that do report individual-subject data are often limited by methodological or conceptual factors, such as a lack of proper statistical analyses or lack of a control group.

Recently, two studies have looked at the individual sensitivity of auditory N400 paradigms, one in healthy adults (Cruse et al., 2014), and the other in healthy children (Petit et al., 2019; Chapter 2 of this thesis). These studies each used two paradigms contrasting auditory words presented in congruent or incongruent semantic contexts. Using a task in which participants were asked to make covert violation decisions, these studies reported a statistically significant N400 effect in a maximum of 58% of neurotypical adults (Cruse et al., 2014) and 56% of neurotypical children (Petit et al., 2019). This moderate sensitivity reflects the challenges associated with assessing ERPs at the individual level. However, using multivariate pattern analyses (MVPA) on these EEG data, we increased the individual-subject detection rate to 88% (Petit et al., 2019). In the current study, we aimed to build on this work by improving the stimulus-delivery paradigm and by using a more robust and sensitive analysis framework. Our goal was to derive a highly sensitive measure of language comprehension in individual children.

We set out to extend the paradigm developed in Petit et al. (2019) and to measure neural responses to auditory words presented in congruent and incongruent sentence frames (i.e. "The squirrel stored nuts in the *tree/door*"). In order to increase children's engagement to the task, and build up strong semantic contexts, the spoken sentences were accompanied by short video-animated cartoons (*e.g.* an animation of a squirrel storing nuts in a tree). With clinical applications in mind, we used a semi-covert task, in which participants were asked to silently judge the semantic congruency of each sentence in their head, with occasional button press requests to check for compliance. We recorded EEG using the active BioSemi system, which allows a shorter setup time with less requirement to prepare the skin, which is preferable in children. We explored several analyses of the EEG data.

With traditional within-individual ERP analyses, it is necessary to *a priori* choose time windows and electrodes of interest in order to reduce the number of comparisons and thus increase the statistical power. However, we have previously demonstrated that for individual subject analysis, *a priori* assumptions about spatio-temporal location should be avoided as there is substantial inter-individual variability in the location and timing of N400-like effects (Petit et al., 2019). For clinical populations, for whom inter-individual variability may be even higher, it is essential to allow some inter-individual variability in the location and timing of the effect of interest. To allow for this without increasing multiple comparisons, we previously used multivariate pattern analyses (MVPA). This approach uses the signal recorded across all EEG channels to detect patterns of brain activation that reliably distinguish between two conditions, in this case, between identical words presented in different lexico-semantic contexts. Using MVPA at each time point, we retain information about when an effect occurs, but allow it to arise from any spatial location. In the current study, we again used MVPA but additionally dropped the requirement for time-resolved results, allowing the classifier to detect an effect with any temporal profile and any spatial configuration. This allowed us to detect differences in the brain’s pattern of activity irrespective of the location and the timing of the difference and was done in an effort to increase our detection rate of effect occurring. Having detected a statistical difference using this approach, we then used time-resolved MVPA and univariate approaches to qualitatively describe the temporal and topographic distribution of the effect.

Using an unconstrained MVPA approach we reached a medium detection rate (65%, 13/20 participants). Univariate analyses were less sensitive (45%, 9/20 participants), possibly due to restricting the analyses to a predefined brain region for appropriate statistical power. Accordingly, we again observed substantial inter-individual variability in the location and timing of the discriminative neural signals. Comparing, informally, to the data from Chapter 2, there was no evidence to suggest than including visual animations increased the sensitivity of our approach.

3.2 Methods

All presentation scripts, stimuli, and analyses scripts are available at https://osf.io/bv2dy/.

3.2.1 Stimulus development and validation

3.2.1.1 Congruent stimuli

The stimuli consisted of 94 congruent and 94 incongruent sentences. The congruent sentences were adapted from the norms of Block and Baldwin (2010) who reported the cloze-probability of 498 sentences for adults in the USA. We started by selecting 450 sentences with a cloze probability higher than 50%. (i.e. more than 50% of the participants completed a given incomplete sentence with the same target word) and for which the target word was a noun. From this set, we then selected 242 sentences in which the target word and of all the keywords were of high frequency (Zipf Log10 frequency > 3.5) according to the children section of the SUBTLEX-UK word database (van Heuven, Mandera, Keuleers, & Brysbaert, 2014). This removed sentences that were not suitable for children.

To facilitate the splicing of the target word from the audio recording, we retained only the 105 sentences in which the boundary between the incomplete sentences and the target was a plosive sound (/t/, /d/, /k/, /p/, /b/, or /g/) and the 35 sentences in which it was possible to add an adjective ending in a plosive sound before the target without disrupting the meaning of the sentence (i.e. To cut the chicken Sue needed a *sharp* knife – see Table 1). The remaining 140 sentences were recorded by a female, native Australian-English speaker. In order to make sure that the congruent sentences were highly congruent for children, 18 native English speaker children, aged 8- to 12-years old (M = 9:11, range = [8:4 to 11:11]) participated in a validation experiment. We presented participants with each of the 140 incomplete sentences and asked them to say the word they thought would best complete each sentence. We then selected only the sentences with a target cloze probability of over 60%, leaving a full stimulus set of 94 sentences (see Supplementary Table S1 for a complete list of the stimuli).

3.2.1.2 Incongruent stimuli

In order to generate an incongruent condition in which the target words and sentence frames were perfectly matched with the congruent condition, we swapped the target words of pairs of congruent sentences (see Table 1). Each target and each incomplete sentence was thus presented once in the congruent condition and once in the incongruent condition. When recombining targets with incongruent sentences frames, we ensured that the target did not violate syntactic structure of the sentence, including matching for plurality. Furthermore, we ensured that the incongruent target did not start with the same sound nor rhyme with the congruent target.

Table 1. Example of congruent and incongruent sentences. The conditions are matched by pairing the targets between the two conditions. In some cases (last two rows), we added an adjective (italics) ending with a plosive sound before the target word to facilitate audio splicing.

Incongruent sentence	Congruent sentence	Target
In autumn, leaves fall off the	After dinner, they washed the	dishes
After dinner, they washed the	In autumn, leaves fall off the	trees
They raised pigs on their *private*	To cut the chicken Sue needed a *sharp*	knife
To cut the chicken Sue needed a *sharp*	They raised pigs on their *private*	farm

3.2.2 Animations

For each congruent sentence, we designed a short, colourful, animated cartoon that matched the meaning of the sentence (using Adobe Photoshop CC 2017). The animation corresponded to the congruent version of the sentence and each animation was presented twice, once in each condition (*e.g.* the animation of a leaves falling off autumnal trees was presented with the sentences “in autumn, leaves fall off the *trees*” versus “in autumn, leaves fall of the *dishes*”).

3.2.3 EEG experiment

3.2.3.1 Participants

Twenty children aged 9 to 12 years (M = 10:6, SD = 0:11, 12 males and 8 females) were recruited through the Neuronauts database of the Australian Research Council Centre of Excellence in Cognition and its Disorders. All participants were native English speakers and received $25 for their participation. This study was approved by the Macquarie University Human Research Ethics Committee (Reference number: 5201200658) and all methods were performed in accordance with the relevant guidelines and regulations. Informed consent was obtained from a parent and/or legal guardian for each participant.

3.2.3.2 Experimental procedure

Participants were seated in front of a computer screen in a slightly dimmed room. The auditory sentences were presented via speakers on both sides of the screen and the corresponding animations played on the screen at a visual angle of approximately 4° of height and width.

All 188 sentences were presented once within a single recording session in a pseudo-random order that was optimised to minimise bias in the sequence of congruent and incongruent trials. Additionally, all of the sentence frames were presented once in either condition before they could be presented again in the other condition. We chose to present each complete sentence only once to minimise repetition effects and to limit the length of the recording session. For our main analysis (below) we calculated that 94 trials per condition would give us 96% power to detect a medium effect (Cohen's d = 0.5) at the individual-subject level.

Each trial started with a central fixation cross, displayed for two seconds, followed by the presentation of a sentence. The animation started with the beginning of the auditory sentence, and included a 500 ms gradual fade-in and a 500 ms fade-out to minimise abrupt onsets and reduce eye-movements. The animation disappeared one second before the target word was presented to ensure that EEG responses to visual information did not contaminate the response to the target word (see Figure 1).

We presented the participants with six practice trials at the start of the experiment to ensure they understood the instructions. Participants were instructed to listen carefully to each sentence and covertly decide whether the sentence was correct or incorrect. Additionally, on some trials, participants were asked to quickly indicate their judgement by pressing a button with their right hand. On these trials a response screen was presented with a green "tick" ("correct meaning") and a red cross ("incorrect meaning") on either side of the display. A central vertical bar decreasing in size indicated the time left to answer (4 s in total). Question trials occurred approximately every 7 trials, with a jitter of ± 3 trials so they were unpredictable. The red cross and the green tick appeared randomly on either side of the screen, so participants could not anticipate which button they would have to press. If they answered correctly, the participant was told that they had caught an "evil alien". If they answered incorrectly, they were shown a semi-masked picture of the alien that they had "failed to catch". After each question trial they were told how many evil aliens they had left to catch.

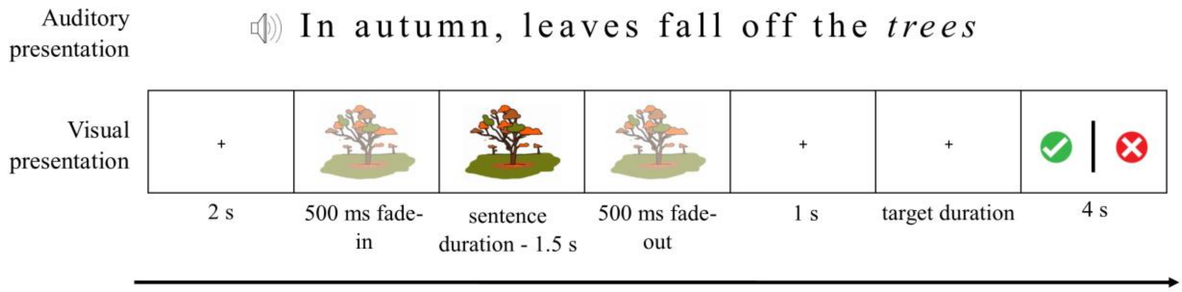


Figure 1. Trial sequence on a question trial. Note that the same animation would be presented for the incongruent version of this sentence ("in autumn, leaves fall of the dishes"). On non-question trials the final display was not shown and the 2 s inter-trial interval started immediately.

3.2.3.3 EEG recording and pre-processing procedures

We acquired EEG data from a 64-channel ActiveTwo BioSemi (BioSemi, Amsterdam, Netherlands). The electrodes were organized according to the 10-20 system, with two electrodes placed on the left and right mastoids for offline referencing. Electro-oculogram generated from eye movements and eyeblinks was recorded using two facial electrodes, located at the outer canthus of and under the right eye, respectively. The data were digitized at 512 Hz with an online reference to the common mode sense (CMS) and all impedances were kept below 30kΩ.

We processed the data off-line using the EEGLAB toolbox in MATLAB (Delorme & Makeig, 2004). We first re-referenced the data to the average of the left and right mastoids, then used a bandpass filter between 0.1 and 40 Hz. We then segmented the data into 1100 ms epochs time-locked to the onset of the auditory target word (100 ms pre-stimulus and 1000 ms post-stimulus). To correct for eye-blinks and eye-movements, we ran an Independent Component Analysis (ICA). We removed the components with scalp distributions, time-courses and spectral contents indicative of eye blinks and eye movements. Finally, we applied a baseline correction to the epochs, and either saved all the epochs for subsequent MVPA analyses, or removed epochs with extreme values (±200 mV) for subsequent univariate analyses. For the purpose of illustrating the topography of the ERP across the scalp, we removed epochs that had extreme values in any channel. At this point, any channel that contributed to the rejection of more than 10% of the trials at this point was interpolated using spherical interpolation, and the pre-processing was done again using the interpolated channel(s) from the filtering stage. An average of 1.9 channels were interpolated per participant (range = 0 - 8). None of the channels in the region of interest (see below) had to be interpolated. For the purpose of analysing the N400 within our region of interest, we removed epochs that had extreme values in any of the 9 channels of interest, in order to preserve data. An average of 4.4 trials were rejected from the region of interest across participants.

3.2.3.4 Multivariate Pattern Analyses

We performed MVPA using the CoSMoMVPA toolbox (Oosterhof, Connolly, & Haxby, 2016) in MATLAB (R2017b). For each participant, a support vector machine (SVM) was trained to discriminate between the brain patterns evoked by the two semantic conditions, using a leave-one-target-out cross-validation approach. Data from all electrodes and all time points in the epoch were used for training. We trained the classifier using a training set consisting of all the trials except the two trials corresponding to a single target word (one from each condition). We then tested the classifier's correct categorization of the remaining two trials. This was repeated 94 times, leaving a different target out each time. We then averaged the accuracy obtained for each target to obtain classifier accuracy for each participant. This yielded a single decoding accuracy score per participant. If this accuracy was significantly above chance (i.e. 50%), we conclude that there was information in the brain signals that was different between the two semantic conditions. To test for significance, we used a permutation test (Maris & Oostenveld, 2007). This test consists of randomly attributing every trial to one of the two conditions, then running the above classification procedure on these permuted data. We repeated this 1000 times to estimate a null distribution of accuracies. Accuracy was considered significant when the observed accuracy was higher than 95% of the null distributions' accuracies (α = .05). We additionally estimated effect sizes for each participant using their mean decoding accuracy and the null distribution, according to the formula:

$$d = \frac{(participant's\ accuracy - mean\ of\ null\ distribution)}{standard\ deviation\ of\ null\ distribution}$$

Finally, we examined the effect of age on the decoding accuracy using Spearman's correlation.

Having established the presence of a difference in the brain signals, we then performed additional analyses to describe the time course and the topology of the effect. We first computed the time course and the topography of the MVPA results, then examined the ERP using univariate analyses of the two conditions.

3.2.3.5 Time-resolved MVPA

In order to quantify the time course of the discriminating brain signals, we ran a follow-up analysis where we trained the classifier to distinguish between the two conditions over time. For each time point, we trained the classifier on the data from all electrodes at that time point and the 10 neighbouring time points (5 on each side). We then repeated this analysis across time points. To test for significance, we used a permutation test and threshold-free cluster enhancement (TFCE), as described by Smith and Nichols (2009), on all time points excluding the baseline. This approach allows for extraction of a statistic of cluster-level support at each time point for the observed accuracy and the permutation results. The maximum TFCE statistic across time of each permutation was used to create a corrected null-distribution. The observed TFCE statistic at each time point was considered significant if it was larger than 95% of the null-distribution. This allowed us to observe the evolution of decoding accuracy over time, thus indicating *when* the classifier could find brain information that discriminated between conditions.

3.2.3.6 Time-space-resolved MVPA

Finally, in order to illustrate both the time course and topography of the decoding, we ran a second follow-up decoding analysis allowing the classifier to use the data coming from a subset of neighbouring electrodes (5 electrodes per neighbourhood) and from a subset of neighbouring time points (11 time points per neighbourhood). We repeated this analysis for each electrode and each time point. This yielded a topographic map of decoding accuracy over time for each participant. For visualisation, we illustrate this topography for decoding accuracy averaged over time within four time-windows spanning 200 to 1000 ms post-stimulus onset.

3.2.3.7 Univariate analyses

We defined a region of interest centred around Cz and the eight surrounding electrodes (Fc1, FCz, FC2, C1, C2, CP1, CPz, and CP2), based on Cruse et al. (2014) and Petit et al. (Chapter 2). We restricted our analyses to the data from 150 ms post-stimulus based on previous results (Kutas & Federmeier, 2011). We averaged the data from the nine electrodes of interest and analysed the difference between the two conditions by running t-tests at each time point from 150 ms onward. We corrected for multiple comparisons by using a temporal cluster threshold calculated using Guthrie and Buschwald's (1991) method. The rationale for this method is that consecutive time points of an ERP reflect continuous neural processes and thus are not statistically independent but correlated. We first calculated the autocorrelation value of our ERP waveform. We then generated 1000 random ERP series with the same autocorrelation value as our original ERP waveform. We calculated the t-test statistics for the difference between the congruent and the incongruent condition at each time point for our original waveform and each of the random series. For each random series, we determined the longest run of t-values below .05. A cluster was considered significant in our original waveform if it was longer than 95% of the random series' longest cluster. These analyses were carried out at the group level and each individual separately.

3.3 Results

We examined children's brain responses to semantically congruent and incongruent spoken and visual sentences using EEG. During the experiment, children simultaneously watched and listened to sentences, and were occasionally prompted to press a button to indicate whether the sentence they just heard was correct (e.g. "the squirrel stored nuts in the *tree*") or incorrect (e.g. "the squirrel stored nuts in the *door*").

3.3.1 Behavioural accuracy

Participants (n = 20) performed the button-press task with a high degree of accuracy (mean percent correct: M = 97.17%, SD = 3.29%, range = [90%, 100%]), indicating that they understood the meaning of the sentences and were able to notice semantic anomalies. They also responded within the required time on 99.3% of trials (mean reaction time: M = 1.63 s, SD = 0.41, range = [0.51, 4.02]).

3.3.2 Multivariate Pattern Analyses results

Using temporally- and spatially-unconstrained multivariate classification analysis, we could decode whether the target word was semantically congruent or incongruent with the sentence in 65% of participants (13/20 participants, Figure 2). Individual decoding accuracies ranged from 68 to 49%, with effect sizes ranging from 4.08 to -.27. Using all of the data available allowed us to detect statistical effects of semantic violation in two thirds of individuals. We did not find a significant correlation between the decoding accuracy and the participants' age (Spearman correlation r = .136, p = .568).

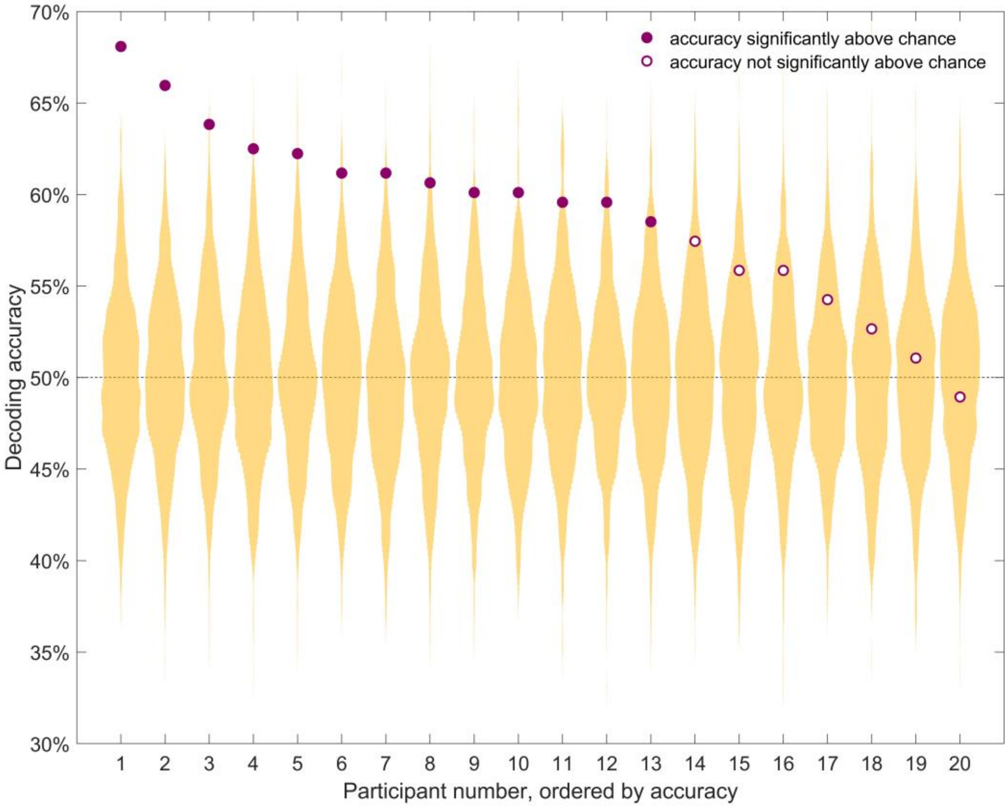


Figure 2. Individual decoding accuracy for classification of identical target words in congruent and incongruent contexts. Purple circles indicate decoding accuracy for each participant. The yellow distribution shows the null distribution obtained by the permutation test for that participant. Thirteen out of 20 participants had a decoding accuracy significantly above chance (full purple circles). Chance (50%) is indicated by the horizontal dashed line. Participants are ordered by decoding accuracy.

3.3.3 Time-resolved MVPA

As our main result yields optimal statistical power by trading off spatial and temporal resolution, we conducted a series of follow-up analysis to qualify when and where the effect of interest arose. First, we used time-constrained multivariate classification analyses to extract the temporal evolution of decoding accuracy for each participant (Figure 3). Restricting the classifier to short time windows, and repeating this over time, shows the time course with which information is decodable from the spatial pattern of activity across the scalp. This analysis revealed substantial variability in the time course of the effect across participants. While three participants (P5, P6, P12) showed decodable information about the semantic condition at around 400 ms, in line with classic N400 effects, others showed decodable semantic information earlier (200 ms, P9) or later (600 ms or later, P1, P3, P9). As expected, this approach was less sensitive than the main analysis because the classifier was not given the entire length of the epoch to distinguish between the conditions, and because the multiple comparisons inherent in this approach necessitated a more stringent alpha level for inference. Of the 13 participants with significant decoding over time and space, only 7 retained enough spatial information to be decoded using this approach.

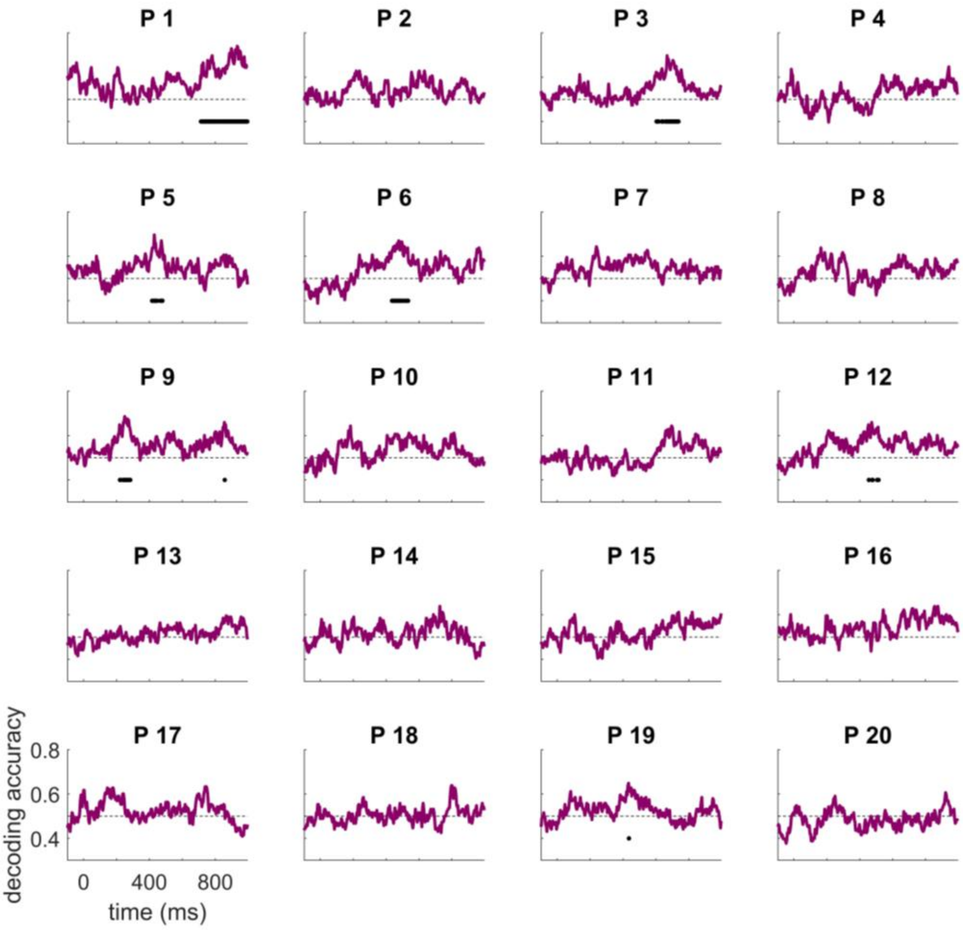


Figure 3. Individual time-resolved decoding accuracies. Chance (50%) is indicted by the horizontal dashed line. Accuracy significantly above chance is indicated by a solid black

horizontal line for each participant (after multiple-comparison correction). Participant numbering is as in Figure 2. Six participants had decoding accuracy significantly above chance.

3.3.4 Time-space resolved MVPA

Next, we used a time and space resolved decoding approach to illustrate when and where there was discriminative information for each individual (Figure 4). There was substantial inter-subject variability in the regions and times that are informative for the classifier. While some participants showed high accuracy at times and regions that correspond to the typical univariate N400 effect (i.e. a centroparietal effect around 400ms, e.g. P5, P6, P8), others showed high accuracies at unexpected times (e.g. late, P11) and/or at unexpected locations (e.g. left lateralised, P9; late and occipitotemporal, P12). To further investigate these results and examine their mapping onto univariate differential responses between conditions, we additionally extracted the N400 ERP in response to the two conditions.

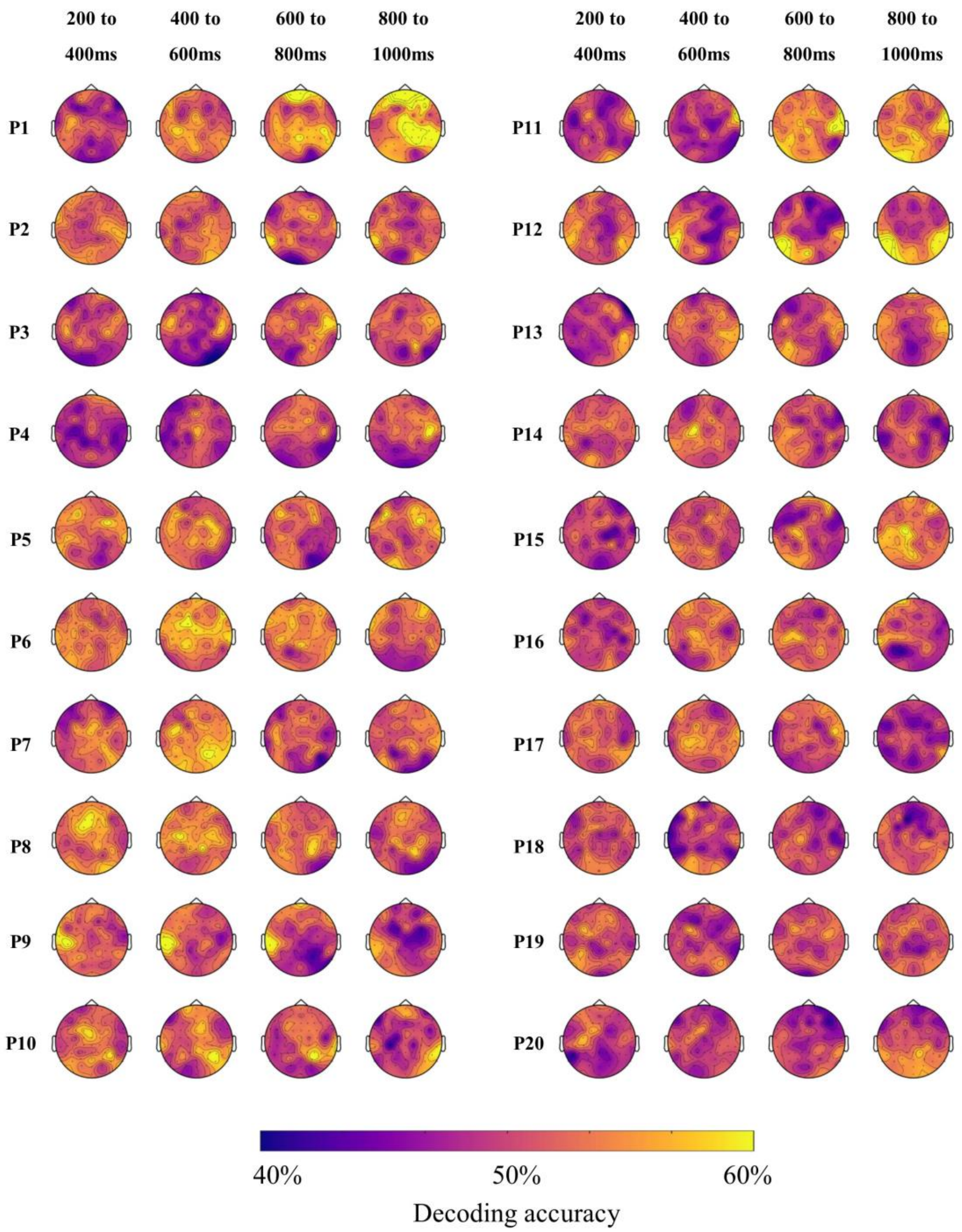


Figure 4. Individual topographic distribution of decoding accuracy for 200 ms time windows from 200 to 1000 ms after target onset. Yellow areas indicate higher decoding accuracy and blue areas indicate lower decoding accuracy. Participant numbering is as in Figure 2, participants 1-13 showed a significant effect in the main analysis. Videos of the decoding accuracy evolution over time are available at https://osf.io/bv2dy/.

3.3.5 Univariate analyses

Finally, we examined the N400 univariate effect by computing the voltage changes over time for the two experimental conditions (congruent and incongruent words) in a pre-specified centroparietal region of interest. This analysis was included for comparison with the wider literature and to determine the extent to which the results of our main analysis could be attributed to classic N400 effects. At the group level, we found a significant N400 effect for a cluster of time points from 289 – 873 ms (Figure 5, top panel). The effect was maximal at central locations from 200 – 400 ms, then extended to frontal regions at later time points (Figure 5, bottom panel).

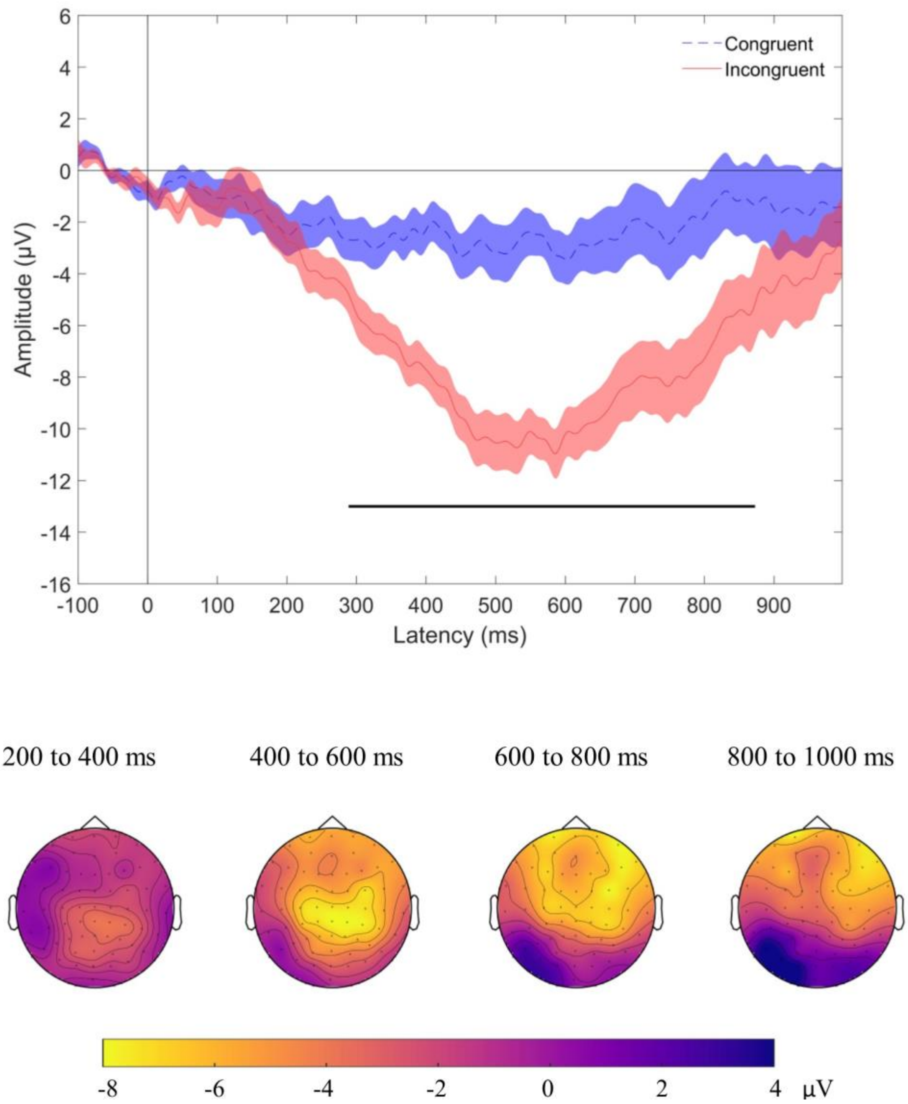


Figure 5. Group N400 effect. Top panel displays grand average ERPs (n = 20), with congruent (dashed blue) and incongruent (solid red) conditions for the signal averaged over our region of interest. Shading indicates standard error of the mean. Time points at which there was a statistical difference between conditions are indicated with a solid black line under the plot ($p < .05$, after cluster correction for multiple comparisons). Bottom panel illustrates the group-level topographic map of the N400 effect (unrelated minus related condition) in 200 ms time windows from 200 to 1000 ms after target onset. Yellow colours indicate a negative difference between the two conditions (incongruent more negative than congruent), while blue colours indicate a positive difference. The N400 effect was distributed over central and centro-frontal regions.

At the individual level, 45% of participants showed a significant univariate N400 effect for at least one significant cluster of time points in our region of interest (9/20 participant, Figure 6). To summarise across analyses, 7 of the 13 participants with significant decoding results also showed a detectable N400 effect, while the remaining 5 did not. Only one participant (P17) showed a detectable N400 effect in the absence of significant MVPA classification.

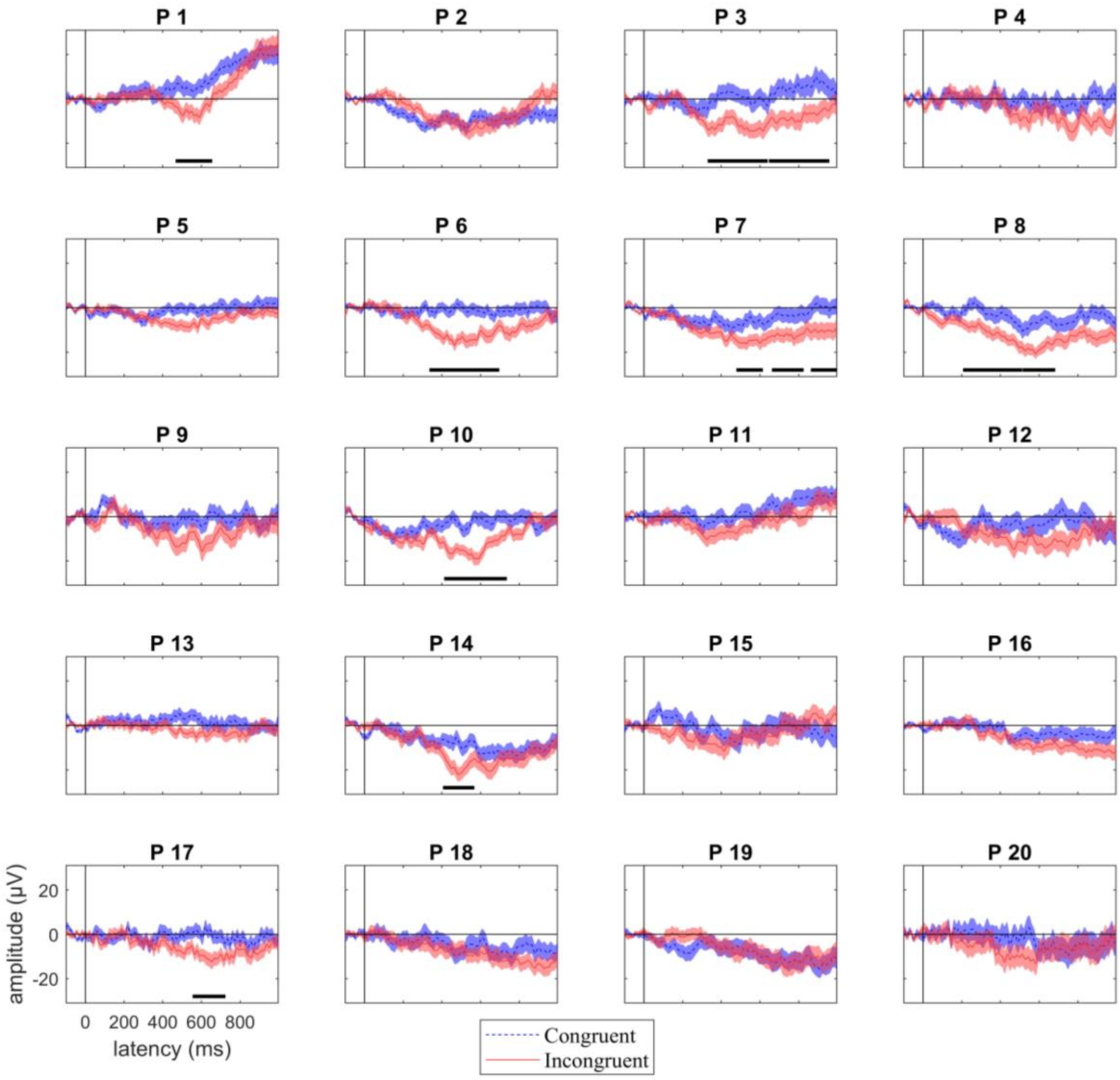


Figure 6. Individual participant responses to target words following congruent (dashed blue) and incongruent (solid red) sentences in our region of interest. Shading indicates standard error of the mean. Time points where there was a statistically significant N400 effect are indicated with a solid, horizontal, black line. P indicates participant, with participant numbering as in Figure 2. 45% (9/20) of participants showed a statistically significant univariate N400 effect.

The topology of the N400 effect was highly variable across participants (Figure 7) in line with the time-space-resolved MVPA analysis (above) and our previous work (Petit et al, 2019).

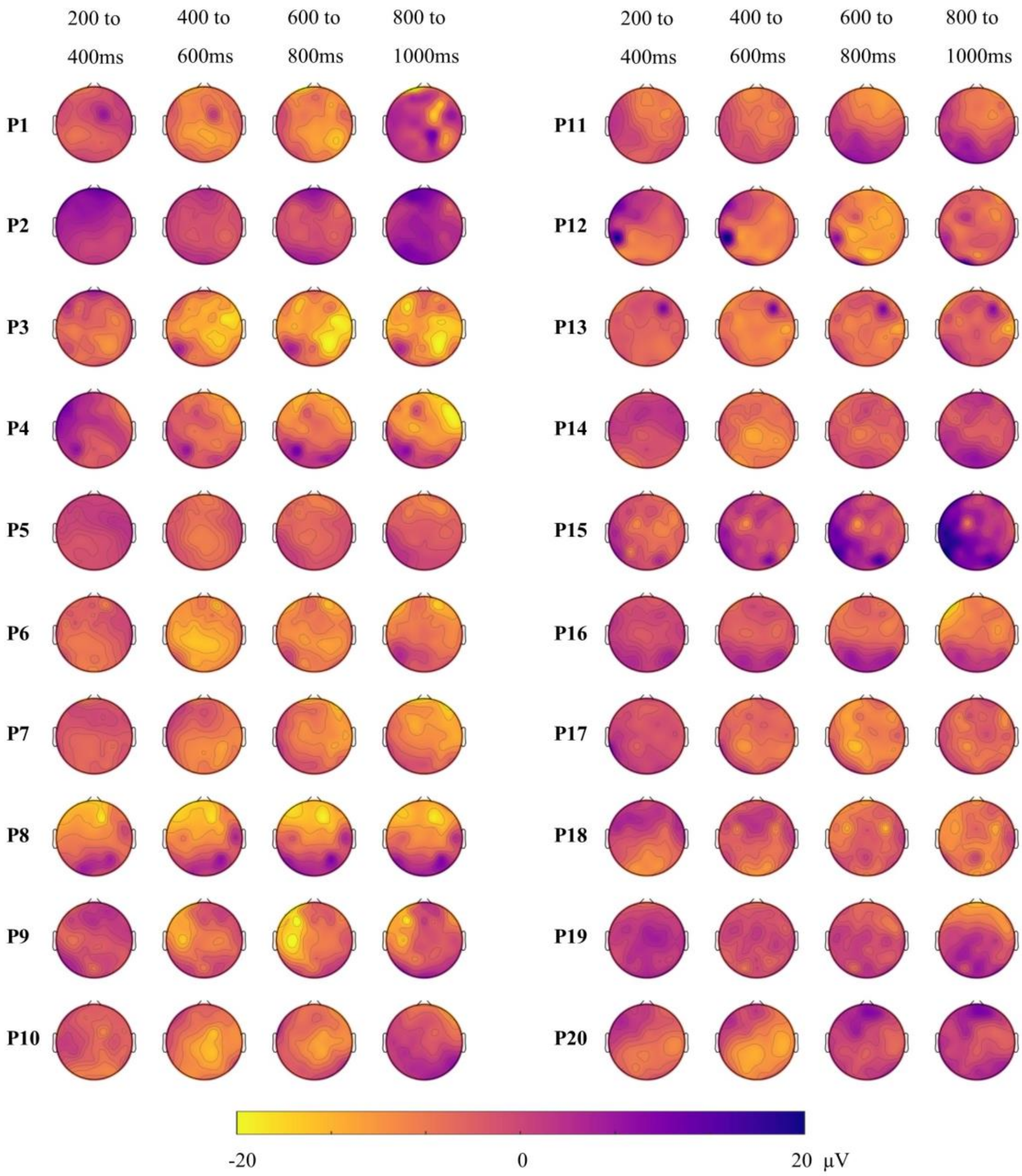


Figure 7. Individual topographic maps of the N400 effect (unrelated minus related condition) for 200 ms time windows from 200 to 1000 ms after target onset. Yellow colours indicate a negative difference between the two conditions (incongruent more negative than congruent), while blue colours indicate a positive difference. The topography of the N400 effect varied across individuals. P indicates participant, with participant numbering as in Figure 2.

3.4 Discussion

Despite the proficiency of studies of the N400, heterogeneity in the neural signals in individual children is yet to be addressed. This study was a validation of a multi-modal N400 paradigm designed to assess lexico-semantic processing from electrophysiological activity, and reports on the sensitivity of neural signals across individual neurotypical children. We recorded EEG from children while they watched video-animated sentences with matched congruent and incongruent endings, and used two complementary approaches to analyse their brain data. Using Multivariate Pattern Analyses (MVPA) to pool information over both space and time, we detected patterns of brain activity that discriminated between the two semantic conditions in 65% of individual children. Further investigation revealed that the patterns of discriminative activity were variable across individuals, ranging both in topography and in time. We additionally analysed the N400 ERP using a univariate approach and found a robust N400 effect in the central location at the group level, as well as in 45% of individual participants. We present a summary of these analyses in Table 2.

Table 2. Summary of individuals' effect in different analyses. 'Yes' indicates that we found a significant effect for this participant.

Participant number	Time- and space-constrained MVPA	Time-resolved MVPA	Univariate N400
1	Yes	Yes	Yes
2	Yes	No	No
3	Yes	Yes	Yes
4	Yes	No	No
5	Yes	Yes	Yes
6	Yes	Yes	Yes
7	Yes	No	Yes
8	Yes	No	Yes
9	Yes	Yes	No
10	Yes	No	Yes
11	Yes	No	No
12	Yes	Yes	No
13	Yes	No	Yes
14	No	No	No
15	No	No	No
16	No	No	No
17	No	No	Yes
18	No	No	No
19	No	Yes	No
20	No	No	No

Our data replicate recent findings that statistically significant univariate differences in neural responses to auditory tokens presented in different semantic contexts can be observed in about half of the participants (Beukema et al., 2016; Cruse et al., 2014; Petit et al., 2019; Rohaut et al., 2015). For the participants not showing statistical univariate N400s, it is unclear whether our methods were not sensitive enough to detect N400 effects to semantic violations or whether N400 effects were truly absent in some participants. In the latter case, it remains unclear whether the absence of an N400 reflects a difference in cognitive processing across individuals. For example, Osterhout (1997) found that in response to syntactic violations, some participants showed an N400 effect while others showed a later P600 effect instead. This would explain why MVPA outperforms univariate analyses here, as it may be capturing discriminatory signals from any electrode or time point. It is also possible that, for some participants, the incongruent sentences did not elicit strong violations of semantic expectations, or that these expectations became weaker over the course of the experiment. Future research should aim to explore individual differences in detail and elucidate whether our methods lack sensitivity to detect individual subject effects or whether the effects genuinely vary between participants. We did not find any evidence for an impact of age on detectable neural effects, although previous work has (Atchley et al., 2006; Friederici & Hahne, 2001; Hahne, Eckstein, & Friederici, 2004; Holcomb, Coffey, & Neville, 1992; Juottonen, Revonsuo, & Lang, 1996). Here, it would be interesting to assess other potential sources of inter-individual differences, such as language ability or lateralisation.

Previous work using purely auditory paradigms have found similar (Cruse et al, 2014; Petit et al, 2019, Exp 1) or better (Petit et al, 2019, Exp 2) detection rates of brain responses to semantic violations than the current study. There was thus no suggestion that adding a visual component made the N400 effect stronger. We hypothesized that creating a semantic context using both visual and auditory information would increase the participants' expectation for the final word, making an incongruent ending even less predictable. This assumption was partly based on previous findings that the N400 reflects lexico-semantic access irrespective of the modality. For example, Nigam et al. (1992) found similar N400 effects for final words of written sentences, whether the final words were in written or picture format. Similarly, when testing the N400 in response to auditory words versus written words, Holcomb and Neville found overlapping neural processes between these two modalities (1990). However, in our study, it is possible that adding the visual animation did little to augment the semantic context that was already unambiguously provided by the sentence. Moreover, it is also possible that adding a visual component had a detrimental effect, for example, by shifting attention away from the auditory modality to the visual one. The high accuracy of participants on the task does suggest that participants correctly classified the congruency of the sentences when probed but clearly this is not a sensitive measure of attention. In the future, presenting the animations before the auditory sentences, instead of concurrently, may alleviate this concern.

Using MVPA allowed us to detect differential brain responses to congruent and incongruent sentences in 65% of individuals (13/20 participants). Of these 13, only 8 (62%) also showed a significant N400 effect. Of the 7 participants who did not have a significant classification result, only one showed a significant univariate N400 effect. This highlights the higher sensitivity of MVPA to detect differences in the pattern of brain activity that are relevant to the semantic condition but may occur at variable times or locations, or as subtle changes in the pattern of brain activation, across individuals. In our implementation, MVPA takes data from all sensors and time points into account without introducing multiple comparisons (only one statistical test is performed per participant). In our univariate approach we minimised multiple comparisons by restricting our analyses to a pre-defined region of interest, with the trade-off of not being sensitive to potential differences occurring outside of this region. Only two studies have investigated decoding accuracy to classify individuals' EEG data. Geuze et al. (2013) used a similar approach as the current study but with adult participants who were presented with words that were normatively related or unrelated to a previously-heard probe word. They found above-chance decoding for all of their participants. It is likely that their higher accuracy reflects the better data quality recorded in adults, who are better able to sit still and engage with the task. More recently, Tanaka et al. (2019) reported significant decoding of single-trial semantic anomalies, with an overall accuracy of 59.5%. They did not report individual-participant accuracies, however, making it hard to compare their results to the current findings. The current study adds to this body of literature by illustrating the capacity of classifiers to detect processing of semantic congruency in individual children performing a semi-covert task, using rigorous and robust EEG data analyses.

Although MVPA appears to be more sensitive for detecting differences in brain activation between conditions, it does not perform better than univariate analyses for all the participants. For one participant in our study (P 17), decoding accuracy was not significantly different from chance, while the N400 univariate effect was statistically significant. Although this may not reflect a meaningful difference in the sensitivity of the two approaches in general, it does prompt some discussion of their relative strengths and weaknesses. At least two accounts may explain a poorer performance by MVPA in some situations. First, there was a difference in the pre-processing pipeline. MVPA was performed using data from all the trials, while univariate analyses used data that we had cleaned during pre-processing (we rejected epochs with extreme amplitude values to remove noisy trials). It is possible that the classifier could not overlook the additional noise in the data, or that the univariate effect was driven by a few trials and was not consistent enough to be detected as a pattern by the classifier. We chose not to clean the data before performing MVPA as our leave-one-target out approach requires the dataset to be balanced, with the same trials present in both conditions. Removing noisy trials would have involved removing their counterpart in the other condition, as well as the trials in both conditions that had the same sentence frame as the noisy trials. Thus, for each noisy trial, we would have to reject four trials in order to keep a balanced design, which would leave too few trials to use for classification. The second explanation is that since the N400 effect occurred at the expected location and time in that individual, adding data from the other channels in MVPA would have mainly added noise, whereas the univariate approach was perfectly aligned with the data. However, for the purpose of a neural test of language comprehension in non-communicative populations, this typicality cannot be assumed, so it may nonetheless be preferable to use methods that can detect differences in brain activity at unexpected timing and locations. This is reflected in our data, where five participants had significant decoding but did not show an N400 effect at our pre-specified region of interest. Participant 4, for example, had a strong decoding result but showed an N400-like effect only at antero-frontal locations (see Figure 7), which would not be detected with conventional N400 analyses.

3.5 Conclusion

This study aimed to assess the reliability of neural signals in response to semantic anomalies in individual neurotypical children. We used a paradigm contrasting congruent and incongruent sentences that were presented with video animations, and analysed the EEG response using different methods to illustrate the heterogeneity of children's brain responses. Despite the challenges associated with analysing individuals' EEG data, we were able to reliably detect neural responses to lexico-semantic anomalies in 65% of individual children using MVPA. Of these, only a subset (62%) also showed a typical N400 ERP at the central location, indicating substantial individual variability in the neural basis of lexical-semantic processing. For the purpose of assessing language comprehension in special populations, such as minimally-verbal autistic children, this paradigm yields only medium sensitivity. Nonetheless, the logic of our approach, using first unconstrained MVPA for maximal statistical power, followed by time and space resolved MVPA and univariate analyses to interrogate the nature of the effect, seems a promising avenue for individualised assessment in the future.

3.6 References

Atchley, R. A., Rice, M. L., Betz, S. K., Kwasny, K. M., Sereno, J. A., & Jongman, A. (2006). A comparison of semantic and syntactic event related potentials generated by children and adults. *Brain and Language*, *99*(3), 236–246. https://doi.org/10.1016/j.bandl.2005.08.005

Beukema, S., Gonzalez-Lara, L. E., Finoia, P., Kamau, E., Allanson, J., Chennu, S., … Cruse, D. (2016). A hierarchy of event-related potential markers of auditory processing in disorders of consciousness. *NeuroImage : Clinical*, *12*, 359–371. https://doi.org/10.1016/j.nicl.2016.08.003

Block, C. K., & Baldwin, C. L. (2010). Cloze probability and completion norms for 498 sentences: Behavioral and neural validation using event-related potentials. *Behavior Research Methods*, *42*(3), 665–670. https://doi.org/10.3758/BRM.42.3.665

Cantiani, C., Choudhury, N. A., Yu, Y. H., Shafer, V. L., Schwartz, R. G., & Benasich, A. A. (2016). From Sensory Perception to Lexical-Semantic Processing: An ERP Study in Non-Verbal Children with Autism. *PLOS ONE*, *11*(8), e0161637. https://doi.org/10.1371/journal.pone.0161637

Coderre, E. L., Chernenok, M., Gordon, B., & Ledoux, K. (2017). Linguistic and Non-Linguistic Semantic Processing in Individuals with Autism Spectrum Disorders: An ERP Study. *Journal of Autism and Developmental Disorders*, *47*(3), 795–812. https://doi.org/10.1007/s10803-016-2985-0

Coderre, E. L., Chernenok, M., O’Grady, J., Bosley, L., Gordon, B., & Ledoux, K. (2019). Implicit Measures of Receptive Vocabulary Knowledge in Individuals With Level 3 Autism. *Cognitive and Behavioral Neurology*, *32*(2), 95. https://doi.org/10.1097/WNN.0000000000000194

Cruse, D., Beukema, S., Chennu, S., Malins, J. G., Owen, A. M., & McRae, K. (2014). The reliability of the N400 in single subjects: Implications for patients with disorders of consciousness. *NeuroImage : Clinical*, *4*, 788–799. https://doi.org/10.1016/j.nicl.2014.05.001

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

DiStefano, C., Senturk, D., & Spurling Jeste, S. (2019). ERP Evidence of Semantic Processing in Children with ASD. *Developmental Cognitive Neuroscience*, 100640. https://doi.org/10.1016/j.dcn.2019.100640

Friederici, A., & Hahne, A. (2001). *Development Patterns of Brain Activity Reflecting Semantic and Syntactic Processes*. https://doi.org/10.1075/lald.24.11fri

Geuze, J., Gerven, M. A. J. van, Farquhar, J., & Desain, P. (2013). Detecting Semantic Priming at the Single-Trial Level. *PLOS ONE*, *8*(4), e60377. https://doi.org/10.1371/journal.pone.0060377

Geytenbeek, J., Harlaar, L., Stam, M., Ket, H., Becher, J. G., Oostrom, K., & Vermeulen, J. (2010). Utility of language comprehension tests for unintelligible or non-speaking children with cerebral palsy: A systematic review. *Developmental Medicine and Child Neurology*, *52*(12), e267-277. https://doi.org/10.1111/j.1469-8749.2010.03807.x

Giacino, J. T., & Smart, C. M. (2007). Recent advances in behavioral assessment of individuals with disorders of consciousness: *Current Opinion in Neurology*, *20*(6), 614–619. https://doi.org/10.1097/WCO.0b013e3282f189ef

Guthrie, D., & Buchwald, J. S. (1991). Significance testing of difference potentials. *Psychophysiology*, *28*(2), 240–244.

Hahne, A., Eckstein, K., & Friederici, A. D. (2004). Brain signatures of syntactic and semantic processes during children's language development. *Journal of Cognitive Neuroscience*, *16*(7), 1302–1318. https://doi.org/10.1162/0898929041920504

Harrison, A. H., & Connolly, J. F. (2013). Finding a way in: A review and practical evaluation of fMRI and EEG for detection and assessment in disorders of consciousness. *Neuroscience and Biobehavioral Reviews*, *37*(8), 1403–1419. https://doi.org/10.1016/j.neubiorev.2013.05.004

Hinterberger, T., Birbaumer, N., & Flor, H. (2005). Assessment of cognitive function and communication ability in a completely locked-in patient. *Neurology*, *64*(7), 1307. https://doi.org/10.1212/01.WNL.0000156910.32995.F4

Holcomb, P. J., Coffey, S. A., & Neville, H. J. (1992). Visual and auditory sentence processing: A developmental analysis using event-related brain potentials. *Developmental Neuropsychology*, *8*(2–3), 203–241. https://doi.org/10.1080/87565649209540525

Holcomb, P. J., & Neville, H. J. (1990). Auditory and Visual Semantic Priming in Lexical Decision: A Comparison Using Event-related Brain Potentials. *Language and Cognitive Processes*, *5*(4), 281–312. https://doi.org/10.1080/01690969008407065

Juottonen, K., Revonsuo, A., & Lang, H. (1996). Dissimilar age influences on two ERP waveforms (LPC and N400) reflecting semantic context effect. *Cognitive Brain Research*, *4*(2), 99–107. https://doi.org/10.1016/0926-6410(96)00022-5

Kasari, C., Brady, N., Lord, C., & Tager-Flusberg, H. (2013). Assessing the minimally verbal school-aged child with autism spectrum disorder. *Autism Research: Official Journal of the International Society for Autism Research*, *6*(6), 479–493. https://doi.org/10.1002/aur.1334

Kotchoubey, B., Lang, S., Mezger, G., Schmalohr, D., Schneck, M., Semmler, A., … Birbaumer, N. (2005). Information processing in severe disorders of consciousness: Vegetative state and minimally conscious state. *Clinical Neurophysiology*, *116*(10), 2441–2453. https://doi.org/10.1016/j.clinph.2005.03.028

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Mathalon, D. H., Faustman, W. O., & Ford, J. M. (2002). N400 and automatic semantic processing abnormalities in patients with schizophrenia. *Archives of General Psychiatry*, *59*(7), 641–648.

Nigam, A., Hoffman, J. E., & Simons, R. F. (1992). N400 to Semantically Anomalous Pictures and Words. *J. Cognitive Neuroscience*, *4*(1), 15–22. https://doi.org/10.1162/jocn.1992.4.1.15

Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (2016). CoSMoMVPA: Multi-Modal Multivariate Pattern Analysis of Neuroimaging Data in Matlab/GNU Octave. *Frontiers in Neuroinformatics*, *10*, 27. https://doi.org/10.3389/fninf.2016.00027

Osterhout, L. (1997). On the brain response to syntactic anomalies: Manipulations of word position and word class reveal individual differences. *Brain and Language*, *59*(3), 494–522. https://doi.org/10.1006/brln.1997.1793

Petit, S., Badcock, N. A., Grootswagers, T., Rich, A. N., Brock, J., Nickels, L., … Woolgar, A. (2019). Towards an individualised neural assessment of receptive language in children. *BioRxiv*, 566752. https://doi.org/10.1101/566752

Pijnacker, J., Geurts, B., van Lambalgen, M., Buitelaar, J., & Hagoort, P. (2010). Exceptions and anomalies: An ERP study on context sensitivity in autism. *Neuropsychologia*, *48*(10), 2940–2951. https://doi.org/10.1016/j.neuropsychologia.2010.06.003

Plesa Skwerer, D., Jordan, S. E., Brukilacchio, B. H., & Tager-Flusberg, H. (2016). Comparing methods for assessing receptive language skills in minimally verbal children and adolescents with autism spectrum disorders. *Autism*, *20*(5), 591–604. https://doi.org/10.1177/1362361315600146

Rohaut, B., Faugeras, F., Chausson, N., King, J.-R., Karoui, I. E., Cohen, L., & Naccache, L. (2015). Probing ERP correlates of verbal semantic processing in patients with impaired consciousness. *Neuropsychologia*, *66*, 279–292. https://doi.org/10.1016/j.neuropsychologia.2014.10.014

Sharma, A., Sauer, H., Hill, H., Kaufmann, C., Bender, S., & Weisbrod, M. (2017). Abnormal N400 Semantic Priming Effect May Reflect Psychopathological Processes in Schizophrenia: A Twin Study. *Schizophrenia Research and Treatment*, *2017*.

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, *44*(1), 83–98. https://doi.org/10.1016/j.neuroimage.2008.03.061

Tager-Flusberg, H., & Kasari, C. (2013). Minimally Verbal School-Aged Children with Autism Spectrum Disorder: The Neglected End of the Spectrum. *Autism Research*, *6*(6), 468–478. https://doi.org/10.1002/aur.1329

Tanaka, H., Watanabe, H., Maki, H., Sakriani, S., & Nakamura, S. (2019). Electroencephalogram-Based Single-Trial Detection of Language Expectation Violations in Listening to Speech. *Frontiers in Computational Neuroscience*, *13*. https://doi.org/10.3389/fncom.2019.00015

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology (2006)*, *67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

Wang, S., Yang, C., Liu, Y., Shao, Z., & Jackson, T. (2017). Early and late stage processing abnormalities in autism spectrum disorders: An ERP study. *PloS One*, *12*(5), e0178542.

3.7 Supplementary materials

Supplementary table 1: List of stimuli used in the auditory-visual N400 experiment

Sentence frame	Congruent target	Incongruent target
While eating Steve accidentally bit his	tongue	kite
It was windy enough to fly a	kite	tongue
The directions did not match any roads on the old	map	pace
Jessie ran the race at a slower	pace	map
At the royal wedding the princess married the	prince	chair
The man sat down on the comfortable	chair	prince
For her school dance Becky wore a new	dress	bridge
The boat passed easily under the	bridge	dress
The little girl was very afraid of the	dark	bread
She went to the bakery to buy a loaf of	bread	dark
The dentist says to brush your teeth twice a	day	band
She played the guitar so she joined the	band	day
The student went to the library to read a	book	paint
He wanted colour in the room so he bought a tin of	paint	book
The squirrel stored nuts in the	tree	bike
A good way to exercise is to ride a	bike	tree
During the lecture Jen kept checking the	time	queen
The princess may someday become a	queen	time
After dinner they washed the	dishes	trees
In autumn leaves fall off the	trees	dishes
Most students prefer to work during the	day	tea
I put the kettle on to make a hot cup of	tea	day
After pulling him over, the police officer told the man to get out of his	car	book
He turned the page of his favourite	book	car
She forgot her watch so she asked for the	time	book
He read a chapter of the	book	time
She got out of the car and closed the	door	cake
There were candles on the birthday	cake	door
I prefer cats rather than	dogs	teeth

The dentist opened her mouth to check her	teeth	dogs
For his wife's birthday he baked a	cake	time
Her job was easy most of the	time	cake
At night the old woman locked the	door	break
Several hours into his shift Lyle was ready for a	break	door
During the winter holidays people tend to eat a lot of good	food	door
I had no key to open the	door	food
His father was the coach of the football	team	door
After coming inside Bob locked the	door	team
The indoor plant was growing bigger and needed a new	pot	broom
John swept the floor with a	broom	pot
The moon was shining in the night	sky	vase
Katie put the flowers in a big	vase	sky
She put on her sunglasses to protect her tired	eyes	marks
Due to his hard work he received excellent	marks	eyes
The surfer is scared of being bitten by a great white	shark	ink
The cheap pen quickly ran out of black	ink	shark
The dragon was slain by the valiant	knight	stick
To help her walk around the girl used a walking	stick	knight
Father carved the turkey with a sharp	knife	ink
The pen in his pocket had unfortunately leaked	ink	knife
He washed his hands with water and	soap	knife
At dinner he cut his steak with a sharp	knife	soap
He had a long day and was in a bad	mood	splash
She jumped in the pool and made a big	splash	mood
To keep the dogs out of the yard he built a six foot	fence	snack
Before practice he decided to eat a quick	snack	fence
She made him a sandwich for his packed	lunch	cell
The prisoner fell asleep inside his dark	cell	lunch
She looked up at night to see a sky full of bright	stars	aim
Tom’s arrows missed the target due to his bad	aim	stars
For breakfast he ate scrambled	eggs	socks
Derek’s feet were cold so he put on some thick	socks	eggs
Cid needed a belt to hold up his	pants	teeth

After every meal it's good to brush your	teeth	pants
His shirt was so worn it had a big	hole	walk
The old man has to use a cane to go on a short	walk	hole
The genie granted the man his first	wish	sky
The stars are high up in the night	sky	wish
There were no extra seats so she sat on the concrete	floor	milk
The little girl left Santa a plate of cookies and	milk	floor
She could tell he was mad by the tone of his loud	voice	milk
The hungry baby wanted to drink	milk	voice
He went to the lake to catch a big	fish	wall
She hung the painting up on the white	wall	fish
He took his dog out for a short	walk	sun
The baseball player's cap protected him from the hot	sun	walk
I roasted the marshmallow over the hot	fire	ring
Bob proposed and gave her a diamond	ring	fire
It was dark in the room so she turned on the bright	light	ring
The man presented his new fiancée with an expensive diamond	ring	light
Sarah saw animals from around the world at the country's best	zoo	sword
The knight got ready for battle and drew his magnificent	sword	zoo
To cut the chicken Sue needed a sharp	knife	farm
They raised pigs on their private	farm	knife
Tim joined the air force because he always wanted to fly a	plane	desk
To do her homework she sat down at her	desk	plane
The real estate agent sold the big	house	word
Today the baby spoke his first	word	house
Most cats see very well at	night	floor
I accidentally spilled some water on the concrete	floor	night
After driving for hours Kevin rested at the next	stop	hand
When he touched the hot stove, the child burned his right	hand	stop
Pete broke his arm and needed to wear a	cast	pan
Jane fried some bacon in the	pan	cast

Chapter 4

Finding hidden treasures: a child-friendly neural test of task-following in individuals using functional transcranial Doppler ultrasound

Selene Petit[*,1], Alexandra Woolgar[1,2], and Nicholas A. Badcock[1]

[1] Department of Cognitive Science, Macquarie University, Australia

[2] Medical Research Council (UK), Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, UK

* Corresponding author at: Australian Hearing Hub, Level 3, 16 University Avenue, Macquarie University, NSW 2109, Australia.

This chapter is currently under review at *Neuropsychologia*:

Petit S., Badcock N. A., Woolgar A. (under review). Finding hidden treasures: a child-friendly neural test of task-following in individuals using functional transcranial doppler ultrasound

Abstract

Despite growing interest in the mental life of individuals who cannot communicate verbally, objective and non-invasive tests of covert cognition are still sparse. In this study, we assessed the ability of neurotypical children to understand and follow task instructions by measuring neural responses through functional transcranial Doppler ultrasound (fTCD). We used fTCD to record the blood flow velocity to the two brain hemispheres of twenty children (aged 9 to 12) while they performed either a language task or a visuospatial-memory task on identical visual stimuli. We extracted measures of cerebral lateralisation for the two tasks separately to investigate lateralisation and we compared the left-right pattern of activation across tasks to assess task-following. At the group level, we found that neural responses were significantly left-lateralised when children performed the language task, as expected. However, when children performed the visuospatial task, we did not find the expected right-lateralisation of neural responses. Furthermore, with unbiased analyses and controlled paradigms, lateralisation was lower than expected from the literature in individual children. Nonetheless, the pattern of hemispheric activation for the two tasks allowed us to confirm task-following in the group of participants, as well as in over half of the individuals separately. This provides an avenue for a covert and inexpensive test of children's ability to covertly follow task instructions and perform different mental tasks on identical stimuli.

4.1 Introduction

Modern neuroscience is taking a growing interest in the mental life of individuals who may not be able to overtly display the extent of their cognitive abilities. In the case of vegetative patients or minimally verbal autistic individuals, for example, recent evidence has shown intact consciousness and language comprehension despite an absence of communicative behaviour (Bekinschtein, Manes, Villarreal, Owen, & Della Maggiore, 2011; Cantiani et al., 2016; Owen et al., 2006). For these populations, it appears that cognitive abilities may be under-estimated by standard assessments, which may lead to inadequate care and support. For this reason, it is crucial to develop a reliable test of cognitive abilities that does not rely on behavioural responses. In this study, we developed a paradigm to examine typically-developing children's task-following abilities directly from their brain activity using a portable and easy-to-setup neuroimaging technology.

Previous research has used functional Magnetic Resonance Imaging (fMRI) to study cognitive abilities in non-communicative patients. In a seminal study, Owen et al. (2006) instructed a patient in vegetative state to modulate her brain activity by alternatively performing one of two mental imagery tasks (imagining playing tennis or imagining walking around her house). The patient's brain responses were similar to a group of neurotypical adults, suggesting that the patient was able to understand the instructions and wilfully modulate her brain response. These results have been replicated and expanded in several studies requiring patients to follow different instructions, such as imagining moving their right versus left hand, counting versus listening to words, or naming pictures (Bekinschtein et al., 2009; Monti, Coleman, & Owen, 2009; Rodriguez Moreno, Schiff, Giacino, Kalmar, & Hirsch, 2010). These studies all found preserved task-following abilities in a number of patients who were previously thought to lack consciousness. However, these promising initial results are mitigated by the inherent limitations of fMRI. First, the high cost of MRI make large-sample studies complex and expensive and may be a barrier for routine clinical use. Second, the requirement to lie still in the scanner, and the noise associated with the scanning procedure, make this method inaccessible to some populations such as young children and some autistic individuals.

Recently, research teams have begun to use functional transcranial Doppler ultrasonography (fTCD) as a non-invasive and inexpensive alternative to fMRI. Being relatively insensitive to movements, fTCD allows for testing a wider range of populations, including those with difficulties staying still such as children (Lohmann, Ringelstein, & Knecht, 2006) and even infants (Kohler et al., 2015). It is also portable allowing it to be used outside the laboratory. FTCD uses two probes placed on participants' left and right temples to measure the blood flow velocity through the left and right middle cerebral arteries. It is inferred that faster blood flow to one hemisphere results from higher neural activity in that region (Lohmann et al., 2006). Thus, fTCD allows for an indirect measure of brain activation in the two hemispheres and can be used to examine the lateralisation of neural responses associated with different cognitive processes.

Our interest was in deriving an implicit measure of language comprehension in non-speaking populations such as minimally-verbal autistic children. We combined the logic of task-following paradigms, in which evidence for wilful modulation of neural activity must reflect comprehension of verbal instructions, with the accessible technology of fTCD. In particular, we sought to use fTCD to provide a measure of differential brain activation in response to different tasks. We designed two tasks to invoke activity that was lateralised to the left or the right hemispheres: a word generation task and a visuospatial memory task, respectively. During both tasks, participants were presented with a spatial array in which a single letter was presented in several locations. In the word generation task, participants were asked to silently generate as many words as possible starting with this letter. During the visuospatial memory task, participants were asked to study the location of letters and remember their location after the letters disappear. Typically, these two types of tasks dominantly activate the left and the right hemisphere, respectively (Bishop, Watt, & Papadatou-Pastou, 2009; Groen, Whitehouse, Badcock, & Bishop, 2012; Rosch, Bishop, & Badcock, 2012a). Here, we examined task-related fTCD responses in a group of 20 participants aged 9-12 years.

In developing our approach, we sought to redress three limitations that would otherwise prevent the clinical application of this method as a test of task-following. First, an issue in extant fTCD research is the heterogeneity in paradigms used to measure cognitive processes. For instance, most researchers estimate language lateralisation using word generation paradigms that involve generating language after viewing a letter on a screen, or describing short animated stories (Badcock, Nye, & Bishop, 2012; Rosch et al., 2012a; Woodhead, Rutherford, & Bishop, 2018). On the other hand, to estimate visuospatial lateralisation, researchers have used tasks with complex visual displays, such as finding rabbits hiding in different holes or lines masked by complex visual dynamic masks (Groen, Whitehouse, Badcock, & Bishop, 2011; Rosch et al., 2012a). As such, the difference in lateralisation between tasks may correspond to changes in the visual and auditory stimuli instead of differences in language and visuospatial processes. Our paradigm overcomes this limitation by presenting the same stimuli in both language and visuospatial tasks.

Second, we highlight a statistical flaw in that the way that lateralisation indices (LIs) are often analysed, which systematically over-estimates laterality. Typically, LIs are calculated by finding the peak in the left-right blood flow velocity difference, then averaging the velocity values over a time-window (usually 2 s) centred on that peak (e.g. Badcock, Nye, et al., 2012; Deppe, Knecht, Lohmann, & Ringelstein, 2004; Groen et al., 2011; Kohler et al., 2015). This quantifies the *maximum* difference of the waveform, which can then be compared between tasks groups. However, because of the way it is derived (taking a maximum from a continuous waveform), it is statistically flawed to compare this difference to zero, for example, to infer that the group or individuals have “significant” lateralisation. We show through simulation that this method, which is unfortunately common practice in fTCD research, will inflate type I (false positive) error. This flaw, also known as ‘double dipping’, is well known in fMRI and electroencephalography research (Kilner, 2013; Kriegeskorte, Simmons, Bellgowan, & Baker, 2009) and applies equally to fTCD data. To avoid this issue, we compute the mean velocity difference over a pre-defined period of interest that is independent of the collected data, and can therefore legitimately be compared to zero to infer significance of lateralisation. Simulations confirm that this method does not produce a bias towards false positives.

A final consideration is that the typical pattern of left-sided language dominance and right-sided visuospatial dominance is only seen in the majority of children, not all of them. Current estimates are that around 60% to 80% of the population have the typical left representation of language while 7.5% to 25% of the population have right hemisphere language and around 10% to 15% have bilateral representation of language functions(Knecht et al., 2000; Lust, Geuze, Groothuis, & Bouma, 2011; Whitehouse & Bishop, 2009). Similarly, visuospatial functions are not supported by the right hemisphere in every individual (Groen et al., 2012; Whitehouse & Bishop, 2009). In their respective studies, Whitehouse and Bishop (2009) found that 25% of adults had either a bilateral or a left-hemisphere dominance for visuospatial memory, while Groen et al. (2012) found this pattern in 29% of children. Thus, for our purpose of measuring task-following on an individual-subject basis, it is not sufficient to rely exclusively on the lateralisation of a single cognitive task. Instead, it is much more powerful to compare lateralisation between tasks. Moreover, this will allow us to be sensitive to differences in the lateralisation of response to each task, and to differences in the time course with which the lateralisation occurs. For example, the difference may manifest as subtle changes of left and right hemisphere activation across time, even if both tasks are lateralized to the same hemisphere.

Using our controlled stimuli and unbiased analyses, we found that the lateralisation of both language and visuospatial processing was weaker than typically reported. Nonetheless, when we compared activation between the two tasks, we found robust evidence that they relied on different brain regions, as indicated by a difference in the pattern of hemispheric activation between tasks in the group. At the individual level, we could detect task-following abilities in

55% of children using this approach. This provides a possible avenue for a covert and inexpensive assessment of task-following in children.

4.2 Methods

All presentation scripts, stimuli, and analysis scripts are available at https://osf.io/xygjv/.

4.2.1 Participants

Twenty-two children were recruited using the Neuronauts database of the Australian Research Council Centre of Excellence in Cognition and its Disorders. All participants were native English speakers and they received $20 for their participation. The data from two participants were excluded due to failing to record data (one participant) and computer crashing (one participant). The final set of data thus came from 20 participants (age range: 9 to 12 years old, *M=10:7, SD=1:1,* 10 male and 10 female). Seventeen of the participants were right-handed, and three were left-handed, based upon parent reports. This study was approved by the Macquarie University Human Research Ethics Committee (Reference number: 5201500074). Participants' parents or guardians provided written consent and the children provided verbal consent.

4.2.2 Apparatus

We acquired blood flow velocity data using a Doppler ultrasonography device (Delica EMS-9UA, SMT medical technology GmbH&Co Wuerzburg, Germany), with probes held in place bilaterally over the left and right temporal windows via a headset. We adjusted the probes until we obtained a good signal of the blood flow through the left and right middle cerebral arteries. The experimental paradigm was presented using Psychtoolbox version 3 (Brainard, 1997; Kleiner et al., 2007) on MATLAB (R2017b), on a 27-inch monitor screen located at 80 cm from the participants. Responses to the trials were given via a button box (Cedrus RB-830).

4.2.3 Paradigm

In order to engage children with the task, we presented the paradigm as a game in which the children collected treasure. A male and a female “pirate” gave auditory instructions and feedback on each trial. The pirates’ voices were recorded by actors who were native Australian English speakers. Male and female voices were included for diversity and were not related to the two tasks (both voices instructed both tasks with equal probability). Each participant completed 40, one-minute trials, switching tasks every 10 trials. In total, they completed 20 trials of the word generation task (task 1) and 20 trials of the visuospatial-memory task (task 2). The order of the task was counterbalanced across participants.

Each trial started with a baseline period of 10 s during which the participants fixated on a black cross at the centre of a white screen. Then either the male (first half of the experiment) or the female (second half of the experiment) pirate was presented on screen and greeted the participant. The pirate asked the participant to get ready and gave the instructions for the task. The instructions were “Think of words that begin with this letter” for task 1 and “remember my treasure map” for task 2. Then a treasure map appeared with 8 repetitions of a letter randomly distributed on the screen (see Figure 1). The letters were displayed in black, presented at a visual angle of approximately 1° in height and width. The treasure map remained visible for 5 s during which children silently generated words (task 1) or studied the position of the letters (task 2). A white screen was then displayed for 10 s, during which the children continues to generate words (task 1) or remember the position of the letters (task 2). Finally, the map reappeared with the letters either exactly at the same location as the first map (in half of the trials) or with one letter displaced from its original location. The pirate would then ask “Did you think of lots of words?” (task 1) or “Is this the same treasure map?” (task 2), and the children would answer “yes” or “no” by pressing either a right or left button on a button box in front of them. The buttons’ position was counterbalanced across participants. A 5 s animation was then presented, showing the treasure that the pirate collected during the trial, with the pirate giving encouraging auditory feedback (e.g. “you’re winning!”, “blow me down, that was brilliant!”, “Pieces of eight, you’re doing great!”). Then the pirate’s voice would indicate that the child should take a break by saying, e.g. “time for a rest”, accompanied by a short animation showed a relaxing situation in which the pirate was yawning or sailing away for the night. This was included to encourage participants to stop performing the tasks, with the intention of encouraging task-related activation to return to baseline. Finally, a blank white screen was presented for 10 s of normalisation, then the next trial began. Each trial featured a different letter, with all the letters of the Roman alphabet being presented once in each task, except for the letters K, Q, W, X, Y, and Z, which were not used as words starting with these letters are rare (Badcock, Nye, & Bishop, 2012). The order of the letters was randomized for each participant. Each letter appeared once in a single task before being repeated in the other task, and the order of the letters was reversed in the second task for each participant. Thus, the paradigm was designed so that the two tasks consisted of identical visual stimuli and identical structure, and differed only in auditory instructions. Any difference in the hemispheric activation could therefore be attributed to the subtly different auditory stimulation or to the clearer difference in the mental task.

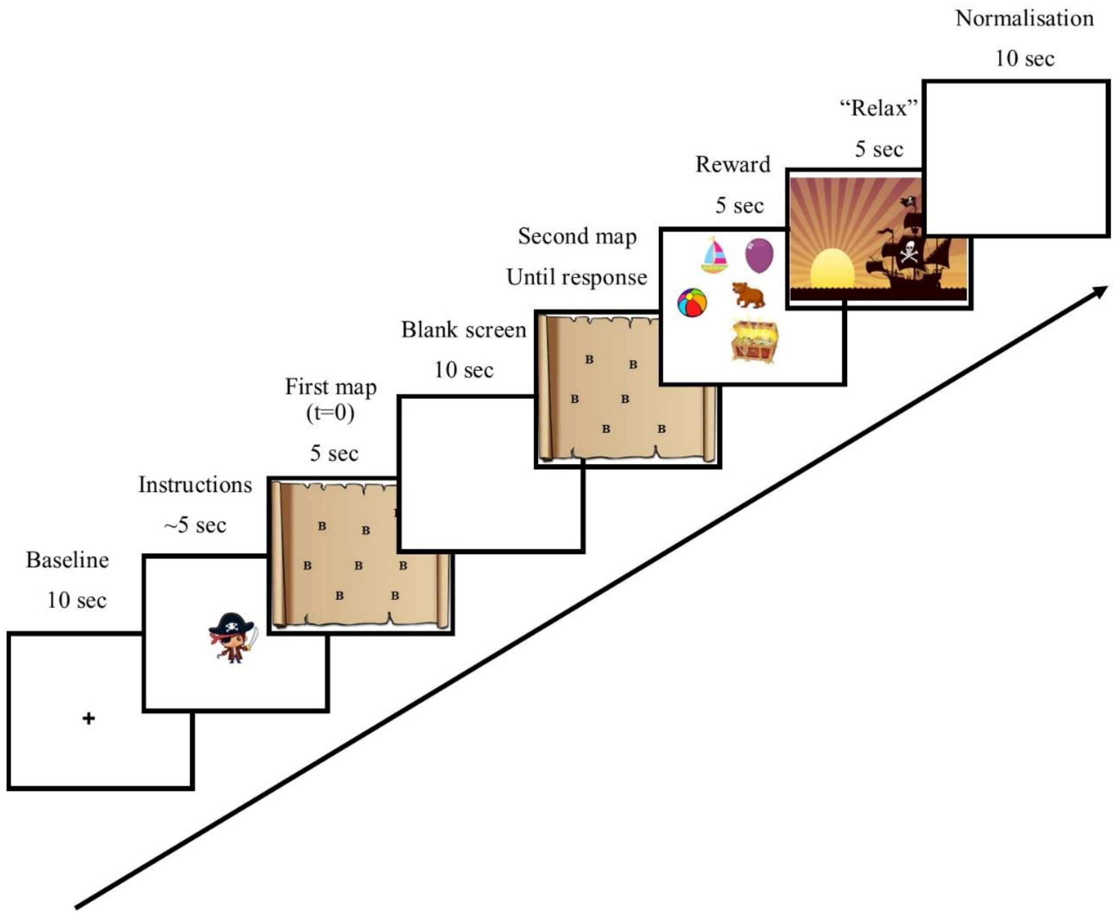


Figure 1: Trial structure. After a baseline period, a pirate was presented and introduced the task. Then a treasure map with letters appeared and children started generating words (language task) or remembering the location of the letters (visuospatial-memory task). The map then disappeared, reappearing after a delay with all letters at the same location or with one letter that changed location. A reward screen then appeared, followed by an animation instructing children to relax, then a normalisation period.

4.2.4 Data preprocessing

We pre-processed the fTCD data using DOPOSCCI (Badcock, Holt, Holden, & Bishop, 2012; Badcock et al., 2018) with MATLAB version R2017B (Mathworks Inc., Sherborn, MA, USA). We first down sampled the raw data to 25Hz, then removed the heart cycle by determining local peaks and using linear heart-cycle correction based on previous work (Badcock et al., 2018). To correct for overall differences in the strength of the signal from the left and right probe (e.g. due to a difference in the alignment of the probes), we normalised the signals to a mean of 100% on an epoch-by-epoch basis. Epochs were -15 to 40 s relative to the onset of the first visual map display (see Figure 1). At this stage, we rejected epochs with extreme values (beyond ± 50% of the mean signal), corresponding to poor insonation or excessive movement. Finally, we performed a baseline correction for each epoch by removing the averaged value of the signal from -15 to -10 s before stimulus onset.

4.2.5 Lateralisation Index

First, for comparison with previous literature, and to test for differences in lateralisation between two tasks using similar stimuli, we calculated lateralisation indices (LIs). Standard practice is to find the peak of the left-minus-right signal within a period of interest (POI) and calculate the average of a 2-s window surrounding this peak. The next step is to compare this LI with zero for categorisation purposes. This approach provides a metric that can be compared between groups or conditions, but because of the way it is derived, it cannot be compared to chance at the individual or group level. To demonstrate that deriving the LI in this was increases the risk of reporting false positives, we ran two simulations, one computing a hypothetical LI using the traditional peak method, and one calculating LIs by averaging over the entire signal in a pre-defined POI. As an approximation of fTCD data, we generated twenty noisy time-series of 1200 time points (approximately the size of our current epoch) by selecting a random number between -20 and 20 at each time point. Next, we chose an *a priori* POI spanning 200 time points (approximately 10 seconds of data, which corresponds to what is typically used in fTCD literature). We then either (a) found the peak in the data within the POI, then averaged the signal over time within a 2-second (50 time points) time window centred on the peak or (b) averaged the signal over the entire POI. For both simulations, we computed a t-test of this value against zero (no information in the signal), equivalent to testing the LI to zero in an individual. We ran these steps 10,000 times and obtained the percentage of simulations for which the LI was significantly different from zero, at alpha = 5%, i.e. the false positive error rate. The peak method yielded a false positive error rate of 6.53%, which represents an inflation of false positives relative to the desired error rate of 5% (as there was no signal in the data). This underestimates the problem in real fTCD data, which will increase with smoothness in the timeseries data. The averaging method yielded a slightly conservative false positive error rate of 4.75%. If we further entered the LI for each individual into a group level test against chance, our simulation returned 5.93% error rate for the peak method and only 5% (chance) for the mean method. Therefore, we used the mean method for LI calculation in individuals and at the group level.

For each task, we first averaged the signal from all the accepted epochs, for the left and right probes. We then calculated the difference between the left and right signals over time. We defined a language POI as 4 to 14 s after the first-map onset, in accordance with previous research that found the highest left-hemisphere activation during this POI for the word generation task (Bishop et al., 2009; Groen et al., 2011, 2012). We defined a visuospatial POI as 20 to 35 s after the first-map onset based on previous findings of highest right-lateralisation during this time-window (Groen et al., 2011, 2012; Rosch et al., 2012a). For each task, we assessed the left-minus-right signal difference within the corresponding POI using a grand average within the POI and performing a one-sampled t-test between this difference and zero. This was done at the group level (across participants) and at the individual level (across trials within participants). We additionally calculated Cohen's d (effect size).

4.2.6 Split-half reliability

Next, we estimated the reliability of the LI by calculating the split-half reliability for each task. This was done using Pearson's correlation between the LI of each participant for the odd and the even trials. We found good reliability for the word generation task ($r = .58$, $p = .0068$) and for the visuospatial memory task ($r = .75$, $p < .001$).

4.2.7 Hemispheric differences analyses

Finally, to address our main question of whether the two tasks yielded different patterns of hemispheric activation, we compared the left-minus-right difference between tasks. We performed a two-tailed, paired-sample t-test for the average blood flow velocity within the language task POI. This POI was chosen as it has showed the strongest lateralisation for language and no lateralisation for visuospatial processing, as in previous research (Badcock et al., 2012; Groen et al., 2012; Whitehouse & Bishop, 2009), so we expected the lateralisation for the two tasks to be maximally different during this period. By only analysing the left-minus-right difference once (in just the language POI), we avoid the need to correct for multiple comparisons thus maximising our statistical power. We performed this analysis at the group level, with a paired t-test across participants, and at the individual level, with a paired t-test across trials (pairing the letters in each condition) within participants. At the group level, with 20 participants, we had .56 power to detect a medium effect size (Cohen's $d = .50$), and .92 power to detect a large effect size ($d = .80$) for alpha = 5%. Similarly, at the individual level with 20 trials, we had .56 power to detect a medium effect size ($d = .50$), and .92 power to detect a large effect size ($d = .80$).

4.3 Results

We examined children's hemispheric activation upon performing two mental tasks, word generation or visuospatial memory. For each task, after every trial, participants had to press a button to indicate whether they could generate many words or whether the visual display was modified, respectively.

4.3.1 Behavioural responses

We considered the participants to have behaviourally performed the tasks based upon the high percentage of trials for which they reported that they thought of many words (M = 77%, range = [45%, 100%]) and the high accuracy for the visuospatial-memory task (mean accuracy = 88%, range = [70%, 100%]).

4.3.2 Group LI

Group level blood flow velocities for the two tasks are shown in Figure 2. We first illustrate the time course of the left and right hemispheres' blood flow velocity for each task (Figure 2A and 2B). We then subtracted the right from the left activation, for each task, within their respective POI (Figure 2C and 2D). We calculated the significance of the LI for both tasks by comparing the left-minus-right difference to 0. The LI for the word generation task was positive (M = 1.78 cm/s, SD = 3.33, 95% confidence interval (CI) = [.66, 2.90], Figure 2A and 2C) and significantly different from 0 ($t_{(19)} = 3.33$, $p = 0.0035$, Cohen's d = .744), indicating left lateralisation at the group level. The LI for the visuospatial memory task was negative (M = -0.62 cm/s, SD = 1.57, 95% CI = [-1.36, 0.11], Figure 2B and 2D) but was not significantly different from zero in the time window of interest ($t_{(19)} = -1.78$, $p = 0.0908$, Cohen's d = -.3983).

4.3.3 Group hemispheric differences

At the group level, blood flow velocity during word generation was significantly different from during the visuospatial memory task. The word generation task was significantly more left-lateralized than the visuospatial memory task (word generation minus visuospatial memory = 2.21 cm/s, $t_{(19)} = 4.11$, $p < .001$, Figure 2E).

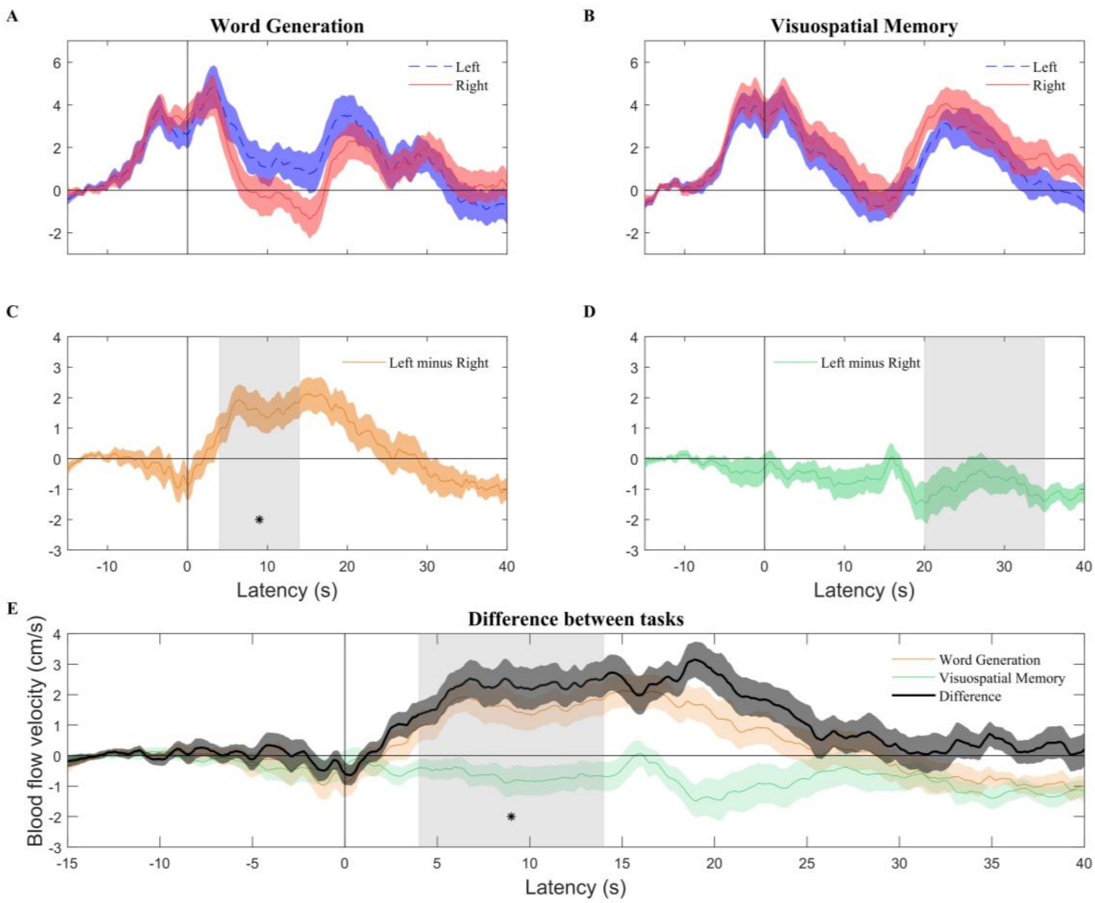


Figure 2: Grand average blood flow velocity for the left (dotted blue line) and right (solid red line) channels (**A**, **B**), and the left-minus-right difference (**C**, **D**) over time for the word generation (**A**, **C**) and visuospatial memory (**B**, **D**) task. **E** shows the left-minus-right difference (i.e. same as middle panels) for the word generation task (orange line) and the visuospatial memory task (green line), and the difference of these differences (black line). Shaded areas depict the standard error and the light-grey areas indicate the periods of interest. Black asterisks indicate significant effects.

4.3.4 Individual LIs

We then examined the significance of LIs in individuals by comparing left-minus-right differences to 0 in the language POI (4 to 14 s) for the language task and the visuospatial POI (20 to 35 s) for the visuospatial memory task. At the individual level (Figure 3), language was significantly lateralised to the left hemisphere for 50% of children (10/20) and to the right hemisphere for 5% of children (1/20). Visuospatial memory was significantly lateralised to the right hemisphere for 20% of children (4/20) and to the left hemisphere for 10% of children (2/20). The remaining participants did not show evidence of significant lateralisation. In addition, we examined the association between lateralisation for language and visuospatial memory. Although 9 of the 20 participants fell into the quadrant where they were numerically left lateralised for language and right lateralised for visuospatial memory, we found a significant correlation between the two functions (ρ = .48, p = 0.034). This indicated that participants with stronger left lateralisation for language also tend to have more leftwards lateralisation for visuospatial memory, and vice versa.

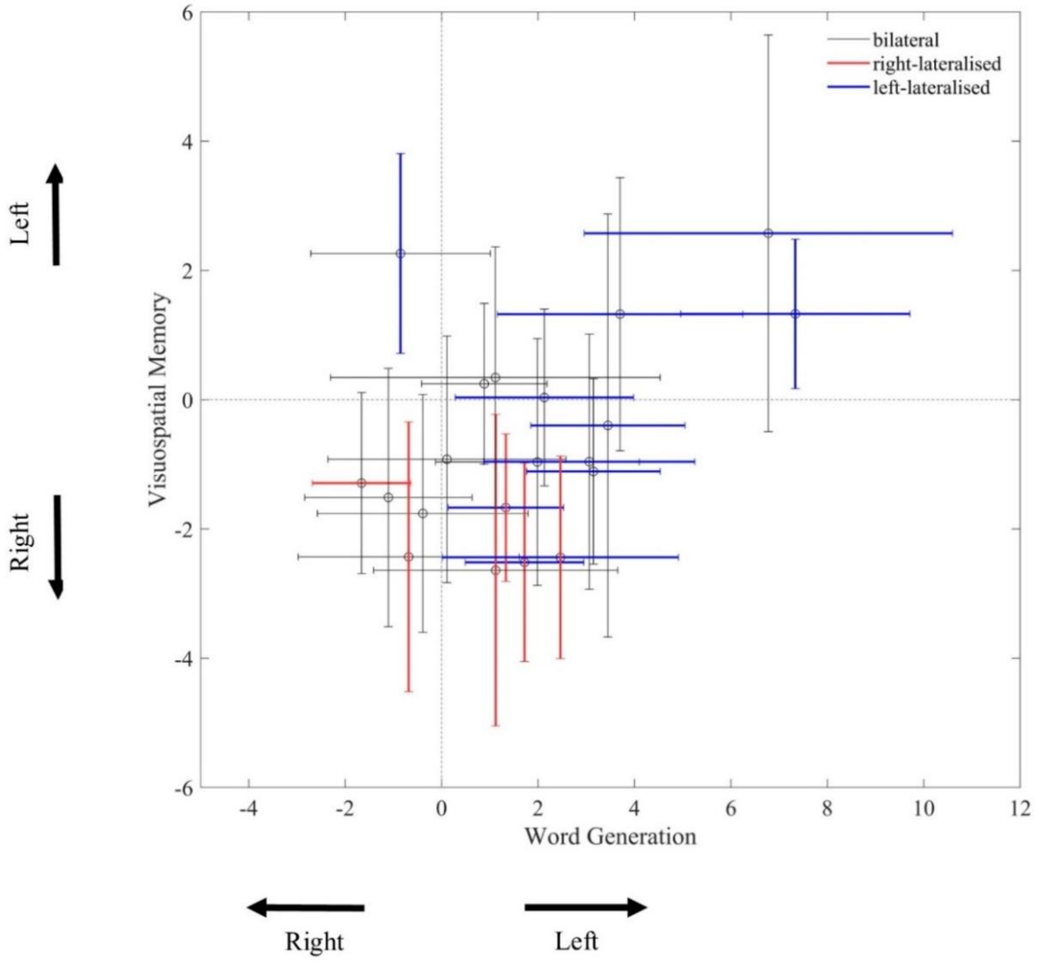


Figure 3. Scatterplot of laterality indices (LIs) of each participant for the word generation (POI = 4 to 14 s) and the visuospatial memory tasks (POI = 20 to 35 s), with 95% confidence interval for each participant (across trials). Participants with confidence intervals (CIs) overlapping

zero are not considered to be lateralised (grey errors bars). Participants with CIs strictly < 0 are right-lateralised (red error bars), and participants with CIs strictly > 0 are left-lateralised (blue error bars).

4.3.5 Individual hemispheric differences

Finally, our main question was whether we could use our fTCD paradigm as an implicit measure of task-following in individual children. We assessed the sensitivity of detecting task-related hemispheric activation in individuals by comparing the left-minus-right differences between tasks. A significant effect of task was found in 55% (11/20) of our participants (see Figure 4), indicating clear evidence for task-following in just over half of individuals. In all the individuals with a significant difference between tasks, blood flow was more left-lateralised in the language task than in the visuospatial task. We did not find significant correlations between the size of the left-right difference and children's performance on the behavioural tasks (word generation task: $\rho = -.28$, $p = .24$, spatial memory task: $\rho = .10$, $p = .66$).

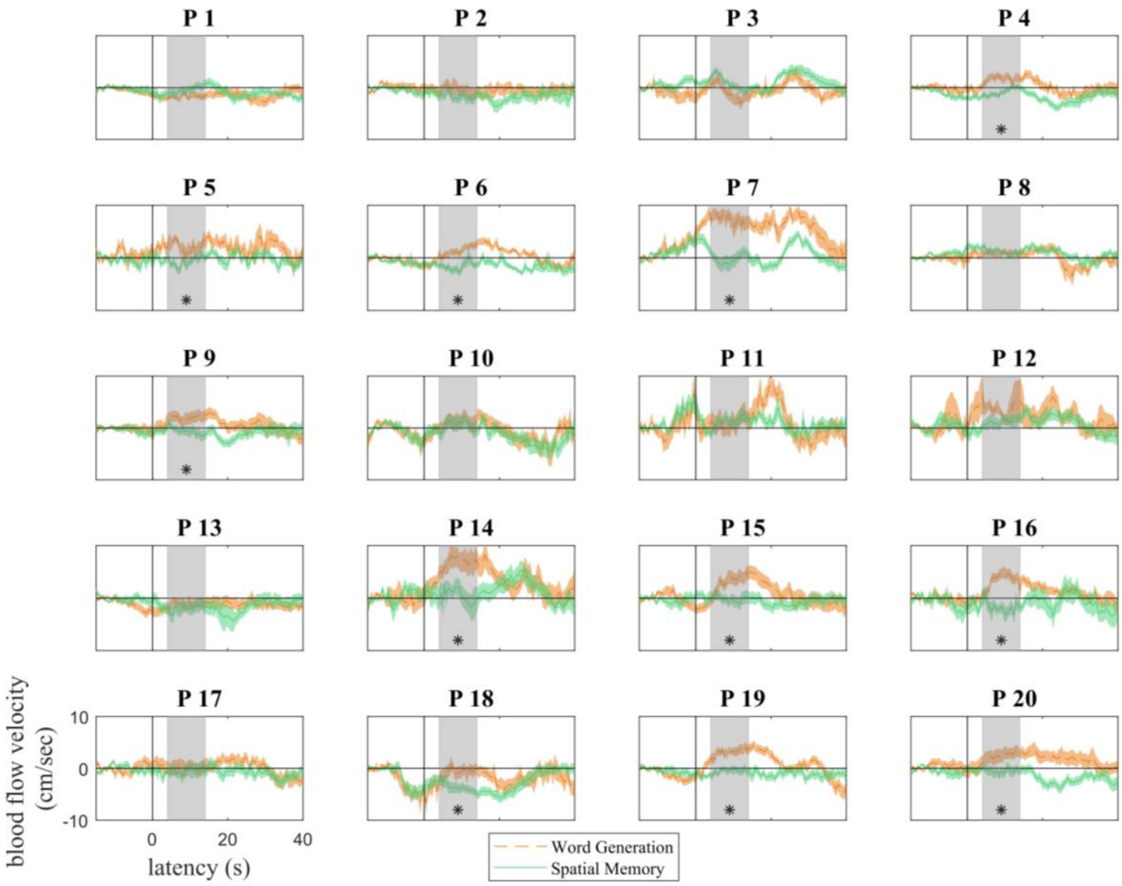


Figure 4. Individual participants pattern of activation (left-minus-right) for the word generation (orange line) and visuospatial-memory (green line) tasks, ± standard error of the mean (shaded area). Grey area indicates the period of interest (POI). Black asterisks indicate a significant difference in the POI. Eleven participants showed a statistically significant difference between the two tasks.

4.4 Discussion

In this study we proposed a rigorous method for evaluating task-following in children based on the lateralisation of brain functions using functional transcranial Doppler ultrasound (fTCD). We designed a controlled, child-friendly paradigm in which children either silently generated words beginning with a particular letter (language task) or remembered the spatial location of letters on a screen (visuospatial task). We computed the left and right hemispheric blood flow velocity while performing the tasks and inferred task-following from the difference in velocity between the two tasks. At the group level, we replicated previous literature in finding a significant left lateralisation for the language task in children (Bishop et al., 2009; Groen et al., 2011, 2012). However, we did not observe the expected right lateralisation for the visuospatial memory task. We also found less marked lateralisation of language and visuospatial memory in individuals, compared to what was expected from the literature. Upon comparing lateralisation between the two tasks, we could infer task-following at the group level, as well as in 55% of our individual participants.

As expected from the literature, we found significant left lateralisation of language at the group level. However, at the individual level, the lateralisation was not as pronounced as expected. Only 50% of children (10/20) had significant left lateralisation for language, and 25% (5/20) had significant right lateralisation for visuospatial memory. This rate is lower than previously reported, i.e. around 70% of people being left lateralised for language (e.g. Knecht et al., 2000; Lust, Geuze, Groothuis, & Bouma, 2011; Whitehouse & Bishop, 2009), and 70% of people being right lateralised for visuospatial memory (e.g. Groen et al., 2012; Whitehouse & Bishop, 2009). The numerically-lower lateralisation found in our study may reflect two crucial differences from previous fTCD research. First, previous fTCD studies that observed differential lateralisation in response to language and visuospatial tasks have used paradigms involving different stimuli in the two tasks (Groen et al., 2011, 2012; Rosch, Bishop, & Badcock, 2012b; Woodhead et al., 2018). This makes it difficult compare the results directly and conclude that differences are due to task, rather than stimuli. In contrast, our paradigm used identical displays of letters in both conditions, with only one minor variation in the auditorily-presented task instructions, making it more likely that differences in blood flow velocity are due to participants performing the task itself, rather than responses to different stimuli.

The second explanation for our lower lateralisation in individuals is that many previous analyses of lateralisation have been biased towards an increased detection of lateralisation. Previous fTCD research typically finds the peak in the left-minus-right blood flow velocity and concludes that if this difference is significantly different from zero, there is evidence for lateralisation. This method introduces statistical bias because it finds a maximum in the data and compares this maximum to zero, thus increasing the likelihood of type I error (probability of incorrectly rejecting the null hypothesis when it is true). In this study, we overcame this problem by computing the left-minus-right blood flow velocity difference over an entire pre-defined POI (as introduced by Woodhead et al., 2018), which brings the type I error back to the scientific standard of 5%. We believe that researchers interested in using fTCD to examine lateralisation of brain function must take care when making inferences based on peak LI analyses. This method is suitable to compare lateralisation between tasks or between groups, but should not be used to test whether individual or group lateralisation is different from chance, because the multiple comparisons inherent in selecting the peak from continuous data have not been accounted for. To check whether the LI analyses do make a difference in the lateralisation that we estimate, we ran a follow-up analysis of our data using the peak method that we believe to be flawed. Using this method, "significant" left lateralisation of language was still found in 50% of participants but right lateralisation of visuospatial memory increased from 20% to 50% of participants. This suggests that indeed the peak method increases bias towards lateralisation. However, this increase does not account for the entire difference that we observe compared to previous literature. This suggests that at least some of the difference between our results and previous literature may be due to differences in the stimuli used or population examined. It is possible that the lateralisation of children age 9-12 may not be as strong as previously suggested. However, our sample size was small (n = 20) and a larger study, with appropriate statistics, would be needed for a reliable estimate of the population's lateralisation.

Upon analysing the association between the lateralisation for the two tasks in individuals, we found a significant positive correlation between language and visuospatial memory lateralisation. In other words, despite the typical group-level left lateralisation for language and right lateralisation for visuospatial memory, individuals who were more left lateralised for language were also more leftwards lateralised for visuospatial memory. This correlation is consistent with previous reports (e.g. Flöel, Buyx, Breitenstein, Lohmann, & Knecht, 2005; Whitehouse & Bishop, 2009). It can be taken as evidence against a causal view of hemispheric specialisation in which localisation of one function to one hemisphere causes localisation of the other function to the other hemisphere (e.g. language is left lateralised *because* visuospatial memory is right lateralised; Whitehouse & Bishop, 2009). However, the data are also not well explained by the dominant alternate view, in which hemispheric lateralisation of each function is independent (Bryden, Hécaen, & DeAgostini, 1983), as this predicts no association in LI between tasks. Instead, our data suggest an *association* in which individuals who tend to rely more on their left hemisphere in one task, will also tend to rely more on this hemisphere in other task.

The second aim of this study was to design a paradigm that could be used to assess task-following abilities in non-verbal individuals. To this end, we computed the patterns of left-minus-right hemispheric activation in response to the two tasks. A consistent difference in the LI between tasks would indicate recruitment of different brain activity in response to the two instructions, and thus indicate preserved understanding. However, even though we found a robust statistical difference in the activation for the tasks at the group level, we could only observe a statistical difference in 55% of individual participants. One reason may be statistical power. We used 20 trials per condition, in line with previous fTCD research (Badcock, Nye, et al., 2012; Groen et al., 2011; Whitehouse & Bishop, 2009), as this amounted to 40 minutes of recording which was at the limit of what we felt that children could manage. However, 20 trials gave only .56 power to detect a medium effect size (d=.50), in the individual subject analysis, so we may have failed to detect differences in the remaining individual children due to insufficient numbers of trials. Another possibility is that some children were not sufficiently engaged in the task or did not perform the task. However, we did not observe any correlation between children's behavioural performance (frequency of self-reporting successful word generation or accuracy on the spatial task) and the size of their differential neural responses. Additionally, even though children reported the task to be engaging, it involved a long baseline period during which they were asked to clear their mind and "think of nothing". This might not be trivial, particularly for children, and it is possible that some participants engaged in language-related processes during the baseline period. This would have affected the lateralisation of the language task and potentially reduced the difference in the activation between the tasks. Because our ultimate goal is to use this paradigm in populations who have minimal speech and potentially unreliable behavioural responses, we chose not to ask participants to report out loud the words they generated on each trial. Instead, we only checked for compliance by asking participants whether they had generated many words or not. It is thus hard to assess whether participants were following the instructions properly for the language tasks. In the future, researchers may consider asking neurotypical children to report the words they generated to being to explore the contribution of task compliance to individual differences in lateralisation.

4.5 Conclusion

We measured brain activation in children using a portable and inexpensive neuroimaging device, fTCD. We analysed lateralisation of neural blood flow in response to a language task and a visuospatial memory task performed on identical visual stimuli. Two main findings emerged. First, by analysing the hemispheric activation pattern across two tasks we were able to robustly observe task-following from neural data in a group of children age 9-12. Statistical differences were further observed in 55% of individual children. This makes our approach potentially suitable for assessing task-following in groups, but the medium sensitivity at the individual level calls for further development of the paradigm and/or analyses before translation to clinical application. Second, our results indicate that the lateralisation of neurotypical children may not be as pronounced as previous research suggests. While previous fTCD research finds left lateralisation of language in about 70% of children, we were only able to see this pattern in 50% of our participants. Similarly, typical fTCD research finds right lateralisation of visuo-spatial memory in 70%, while we found this pattern in 20% of our participants. This was potentially due to the controlled stimulus presentation and/or less biased statistical assessment of lateralisation.

Overall, our methods constitute a promising step towards the neural measurement of task-following abilities in groups of children. These methods, however, need refining before they can be used as an assessment tool for individuals, possibly with more trials, an independent index of subject compliance, and refined paradigms to maximally differentiate between the two hemispheres.

4.6 References

Badcock, N. A., Holt, G., Holden, A., & Bishop, D. V. M. (2012). dopOSCCI: A functional transcranial Doppler ultrasonography summary suite for the assessment of cerebral lateralization of cognitive function. *Journal of Neuroscience Methods*, *204*(2), 383–388. https://doi.org/10.1016/j.jneumeth.2011.11.018

Badcock, N. A., Nye, A., & Bishop, D. V. M. (2012). Using functional transcranial Doppler ultrasonography to assess language lateralisation: Influence of task and difficulty level. *Laterality: Asymmetries of Body, Brain and Cognition*, *17*(6), 694–710. https://doi.org/10.1080/1357650X.2011.615128

Badcock, N. A., Spooner, R., Hofmann, J., Flitton, A., Elliott, S., Kurylowicz, L., … Keage, H. A. D. (2018). What Box: A task for assessing language lateralization in young children. *Laterality: Asymmetries of Body, Brain and Cognition*, *23*(4), 391–408. https://doi.org/10.1080/1357650X.2017.1363773

Bekinschtein, T. A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., & Naccache, L. (2009). Neural signature of the conscious processing of auditory regularities. *Proceedings of the National Academy of Sciences*, *106*(5), 1672. https://doi.org/10.1073/pnas.0809667106

Bekinschtein, T. A., Manes, F. F., Villarreal, M., Owen, A. M., & Della Maggiore, V. (2011). Functional Imaging Reveals Movement Preparatory Activity in the Vegetative State. *Frontiers in Human Neuroscience*, *5*. https://doi.org/10.3389/fnhum.2011.00005

Bishop, D. V. M., Watt, H., & Papadatou-Pastou, M. (2009). An efficient and reliable method for measuring cerebral lateralization during speech with functional transcranial Doppler ultrasound. *Neuropsychologia*, *47*(2), 587–590. https://doi.org/10.1016/j.neuropsychologia.2008.09.013

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4). Retrieved from https://brill.com/view/journals/sv/10/4/article-p433_15.xml

Bryden, M. P., Hécaen, H., & DeAgostini, M. (1983). Patterns of cerebral organization. *Brain and Language*, *20*(2), 249–262. https://doi.org/10.1016/0093-934x(83)90044-5

Cantiani, C., Choudhury, N. A., Yu, Y. H., Shafer, V. L., Schwartz, R. G., & Benasich, A. A. (2016). From Sensory Perception to Lexical-Semantic Processing: An ERP Study in

Non-Verbal Children with Autism. *PLoS ONE*, *11*(8). https://doi.org/10.1371/journal.pone.0161637

Deppe, M., Knecht, S., Henningsen, H., & Ringelstein, E. B. (1997). AVERAGE: a Windows® program for automated analysis of event related cerebral blood flow. *Journal of Neuroscience Methods*, *75*(2), 147–154. https://doi.org/https://doi.org/10.1016/S0165-0270(97)00067-8

Deppe, M., Knecht, S., Lohmann, H., & Ringelstein, E. B. (2004). A Method for the Automated Assessment of Temporal Characteristics of Functional Hemispheric Lateralization by Transcranial Doppler Sonography. *Journal of Neuroimaging*, *14*(3), 226–230. https://doi.org/10.1111/j.1552-6569.2004.tb00242.x

Flöel, A., Buyx, A., Breitenstein, C., Lohmann, H., & Knecht, S. (2005). Hemispheric lateralization of spatial attention in right- and left-hemispheric language dominance. *Behavioural Brain Research*, *158*(2), 269–275. https://doi.org/10.1016/j.bbr.2004.09.016

Groen, M. A., Whitehouse, A. J. O., Badcock, N. A., & Bishop, D. V. M. (2011). Where were those rabbits? A new paradigm to determine cerebral lateralisation of visuospatial memory function in children. *Neuropsychologia*, *49*(12), 3265–3271. https://doi.org/10.1016/j.neuropsychologia.2011.07.031

Groen, M. A., Whitehouse, A. J. O., Badcock, N. A., & Bishop, D. V. M. (2012). Does cerebral lateralization develop? A study using functional transcranial Doppler ultrasound assessing lateralization for language production and visuospatial memory. *Brain and Behavior, 2*(3), 256–269. https://doi.org/10.1002/brb3.56

Kilner, J. M. (2013). Bias in a common EEG and MEG statistical analysis and how to avoid it. *Clinical Neurophysiology*, *124*(10), 2062–2063. https://doi.org/10.1016/j.clinph.2013.03.024

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psychtoolbox-3. *Perception*, *36*(14), 1–16.

Knecht, S., Dräger, B., Deppe, M., Bobe, L., Lohmann, H., Flöel, A., … Henningsen, H. (2000). Handedness and hemispheric language dominance in healthy humans. *Brain, 123*(12), 2512–2518. https://doi.org/10.1093/brain/123.12.2512

Kohler, M., Keage, H. A. D., Spooner, R., Flitton, A., Hofmann, J., Churches, O. F., … Badcock, N. A. (2015). Variability in lateralised blood flow response to language is associated with language development in children aged 1–5 years. *Brain and Language*, *145–146*, 34–41. https://doi.org/10.1016/j.bandl.2015.04.004

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., & Baker, C. I. (2009). Circular analysis in systems neuroscience – the dangers of double dipping. *Nature Neuroscience*, *12*(5), 535–540. https://doi.org/10.1038/nn.2303

Lohmann, H., Ringelstein, E. B., & Knecht, S. (2006). Functional transcranial Doppler sonography. *Frontiers of Neurology and Neuroscience*, *21*, 251–260. https://doi.org/10.1159/000092437

Lust, J. M., Geuze, R. H., Groothuis, A. G. G., & Bouma, A. (2011). Functional cerebral lateralization and dual-task efficiency—Testing the function of human brain lateralization using fTCD. *Behavioural Brain Research*, *217*(2), 293–301. https://doi.org/10.1016/j.bbr.2010.10.029

Monti, M. M., Coleman, M. R., & Owen, A. M. (2009). Executive functions in the absence of behavior: functional imaging of the minimally conscious state. In S. Laureys, N. D. Schiff, & A. M. Owen (Eds.), *Progress in Brain Research* (Vol. 177, pp. 249–260). Retrieved from http://www.sciencedirect.com/science/article/pii/S0079612309177178

Owen, A. M., Coleman, M. R., Boly, M., Davis, M. H., Laureys, S., & Pickard, J. D. (2006). Detecting Awareness in the Vegetative State. *Science*, *313*(5792), 1402. https://doi.org/10.1126/science.1130197

Rodriguez Moreno, D., Schiff, N. D., Giacino, J., Kalmar, K., & Hirsch, J. (2010). A network approach to assessing cognition in disorders of consciousness. *Neurology*, *75*(21), 1871. https://doi.org/10.1212/WNL.0b013e3181feb259

Rosch, R. E., Bishop, D. V. M., & Badcock, N. A. (2012a). Lateralised visual attention is unrelated to language lateralisation, and not influenced by task difficulty – A functional transcranial Doppler study. *Neuropsychologia*, *50*(5), 810–815. https://doi.org/10.1016/j.neuropsychologia.2012.01.015

Rosch, R. E., Bishop, D. V. M., & Badcock, N. A. (2012b). Lateralised visual attention is unrelated to language lateralisation, and not influenced by task difficulty – A functional transcranial Doppler study. *Neuropsychologia*, *50*(5), 810–815. https://doi.org/10.1016/j.neuropsychologia.2012.01.015

Whitehouse, A. J. O., & Bishop, D. V. M. (2009). Hemispheric division of function is the result of independent probabilistic biases. *Neuropsychologia*, *47*(8), 1938–1943. https://doi.org/10.1016/j.neuropsychologia.2009.03.005

Woodhead, Z. V. J., Rutherford, H. A., & Bishop, D. V. M. (2018). Measurement of language laterality using functional transcranial Doppler ultrasound: a comparison of different tasks. *Wellcome Open Research*, *3*. https://doi.org/10.12688/wellcomeopenres.14720.2

Chapter 5

Neural assessment of lexico-semantic processing in minimally-verbal autism: a pilot case-series

Selene Petit[*,1], Nicholas A. Badcock[1], Elizabeth Pellicano[2], and Alexandra Woolgar[1,3]

[1] Department of Cognitive Science, Macquarie University, Australia

[2] Department of Educational Studies, Macquarie University, Australia

[3] Medical Research Council (UK), Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, UK

* Corresponding author at: Australian Hearing Hub, Level 3, 16 University Avenue, Macquarie University, NSW 2109, Australia.

Abstract

Emerging evidence suggests that standardised tests of cognition may underestimate the abilities of autistic individuals with minimal spoken language. Instead, objective and reliable neural assessments may help. In this pilot study, we employed an auditory-electroencephalography (EEG) paradigm that we have recently developed to test for neural evidence of individual-level lexico-semantic processing. We recorded EEG from three minimally-verbal autistic children while they listened to identical spoken words, presented in the context of congruent and incongruent sentences, and compared this to data from 20 age-matched, typically-developing controls. For each individual, we used univariate analyses and multivariate decoding to evaluate differences in their neural responses to the identical stimuli in the two contexts, which we took as evidence of lexico-semantic processing. Nineteen out of 20 typically developing children, and two of the three autistic children, showed a differential response in the two conditions. This study provides a proof-of-concept that EEG, together with a child-friendly auditory paradigm, can be used with the minimally-verbal autistic population. It also suggests that at least some minimally-verbal autistic children engage in lexico-semantic processing sufficient to drive differential neural responses that are detectable in a single session of EEG. This may open an important window into the hidden mental life of children with minimal spoken language.

5.1 Intro

The language abilities of autistic people[1] are extremely diverse, ranging from above-average linguistic skills to complete mutism (Paul, Chawarska, Cicchetti, & Volkmar, 2008; Rapin & Dunn, 2003; Tager-Flusberg & Kasari, 2013). Around 30% of autistic people are considered 'minimally-verbal', meaning they failed to acquire functional speech by school age (Kasari, Brady, Lord, & Tager-Flusberg, 2013; Mody & Belliveau, 2013; Tager-Flusberg & Kasari, 2013). Because of their absence of speech and its frequent co-occurrence with behavioural challenges, it can be particularly challenging to reliably estimate cognitive abilities in this population. In particular, standardised tests have been found to be unreliable due to the social constraints associated with the testing environment, in addition to the requirements for verbal or behavioural answers. As a result, the majority of autism research has focused on autistic individuals with average or above average language and intelligence (Tager-Flusberg & Kasari 2013).

Yet, despite the challenges associated with assessing minimally-verbal individuals, it is crucial to establish their level of cognitive abilities. In particular, accurate assessment of receptive language abilities may greatly impact these individuals' wellbeing and care. Here, we are interested in neural tests of language comprehension that would inform on the potentially hidden linguistic abilities of these children. Such objective assessments would be crucial, following the recent realisation that we may currently underestimate the level of language comprehension of minimally-verbal individuals. Two recent sources of evidence suggest that at least a subset of minimally-verbal autistic individuals have preserved language

[1] We use 'identify-first' language ('autistic person') rather than person-first language ('person with autism'), because it is the preferred term of autistic activists (e.g. Sinclair, 2013) and many autistic people and their families (Kenny et al., 2016) and is less associated with stigma (Gernsbacher, 2017).

processing. The first one comes from numerous anecdotal reports of minimally-verbal individuals who use alternative modes of communication, such as picture-exchange or pointing to letters on a board, and who display intact language comprehension (Higashida, 2016; Kedar, 2012, http://idoinautismland.com, https://iaminmyhead.com, https://nottootrapped.wordpress.com). The second source of evidence comes from recent electrophysiological evidence showing intact neural responses to linguistic stimuli (Cantiani et al., 2016; Coderre et al., 2019; DiStefano, Senturk, & Spurling Jeste, 2019).

Electroencephalography (EEG) is a passive method that allows real-time recording of the brain's electrical activity through sensors placed on the scalp (Luck, 2014). This technique is non-invasive, silent, and relatively inexpensive, making it suitable to investigate neural processing in minimally-verbal autistic children (Tager-Flusberg & Kasari, 2013; Tager-Flusberg et al., 2017). Event-related potentials (ERP) are the averaged neural responses to particular stimuli and provide insight into neural mechanisms involved with processing stimuli (Luck, 2014). The N400 ERP component is a negative-going potential, typically recorded while participants are presented with words, and shows a larger negativity when the words violate the semantic context in which they occur – this difference being referred to as the N400 effect (Kutas & Federmeier, 2011). For example, the spoken word "neck" evokes a larger N400 when presented within the incongruent sentence "There were candles on the birthday *neck"* compared to the congruent sentence "She wore a necklace around her *neck"*. Similarly, the N400 effect also occurs for spoken words following matched and mismatched pictures (e.g. the spoken word "bird" evokes a larger N400 when it is preceded by a picture of a desk, compared to a picture of a bird). As such, the presence of N400 effects in response to lexico-semantic manipulations is indicative of lexico-semantic processing. Following this, several studies have used the N400 as a marker of vocabulary knowledge and lexico-semantic processing in adults (Kutas, 1993; Nigam, Hoffman, & Simons, 1992; Perrin & García-Larrea, 2003), children (Atchley et al., 2006; Henderson, Baseler, Clarke, Watson, & Snowling, 2011; Rämä, Sirri, & Serres, 2013), and autistic people (Fishman, Yam, Bellugi, Lincoln, & Mills, 2011; McCleery et al., 2010; Ribeiro, Valasek, Minati, & Boggio, 2013).

To my knowledge, only three studies have used the N400 to measure lexico-semantic processing in minimally-verbal autistic people. Cantiani et al. (2016) used a picture-word paradigm, in which they presented children with a picture, followed by its corresponding auditory label or a mismatched label. They measured the EEG response to the labels in ten neurotypical children and ten minimally-verbal autistic children. They found N400 effects in the neurotypical group but not in the autistic group, suggesting impaired lexico-semantic processing at the group level. However, when they examined individuals' waveforms, they found significant N400 effects in five out of ten autistic children, and eight out of ten neurotypical children, suggesting preserved lexico-semantic processing in at least a subset of the autistic children. In this study, however, the authors focused their analyses around a pre-specified region of interest centred around the vertex, thus they may have failed to detect neural effects occurring at other, unexpected locations.

More recently, DiStefano et al. (2019) measured the EEG response of neurotypical, verbal autistic, and minimally-verbal autistic children while they performed a picture-word paradigm similar to that of Cantiani et al. (2016). Contrary to Cantiani et al., DiStefano et al. found a significant N400 effect in all three groups, indicating preserved lexico-semantic processing even in the minimally-verbal group. This study, however, did not report individual results, making it difficult to evaluate the inter-individual heterogeneity of neural signals. Finally, Coderre et al. (2019) examined the use of three implicit measures of vocabulary knowledge in minimally-verbal autistic adults. They used a picture-word paradigm with words chosen to be either likely known or likely unknown by participants (as determined by their frequency in the Corpus of Contemporary American English), and they used three implicit measures of vocabulary knowledge: eye-movements, pupil dilations, and ERP responses. They tested five adults with severe autistic symptoms, including two who were minimally-verbal (i.e. with no functional spoken language). They had various detection rates of vocabulary knowledge with the different methods but neither of the minimally-verbal adults showed evidence for preserved vocabulary knowledge.

Overall, the mixed findings from the sparse literature on minimally-verbal autistic individuals (total of ten child and two adult individual datasets reported) call for additional research using objective and sensitive measures of lexico-semantic processing. Moreover, a limitation of the work so far has been the use of picture-word paradigms. The presence of N400 effects in response to this paradigm is evidence for lexico-semantic comprehension of the words. However, it does not allow us to conclude on discourse-level comprehension. In particular, it is possible that N400 effects recorded during picture-word paradigms may reflect learnt associations between pictures and labels. In addition, the use of both auditory and visual stimuli makes it difficult to interpret null effects in autistic individuals. An absence of N400 effects may reflect altered lexico-semantic processing but it may also result from altered visual attention or processing, or altered auditory-visual integration. Visual-attention deficits have previously been reported in autism, thus the use of visual stimuli may not be ideal to test this population (Behrmann, Thomas, & Humphreys, 2006; Ortiz-Mantilla, Cantiani, Shafer, & Benasich, 2019).

In a previous study, we developed a paradigm contrasting response to identical spoken words presented in congruent and incongruent spoken sentences, and validated the reliability of neural signals evoked by these semantic manipulations in neurotypical children (Petit, Badcock, Grootswagers, Rich, et al., 2019; Chapter 2 of this thesis). Three findings emerged from this study. First, despite a clear N400 effect at the group level, there was large heterogeneity of neural signals in individual children, with only 56% of children showing a statistically significant N400 effect. Second, a portable and easy-to-setup EEG system, Emotiv EPOC+, recorded similar ERPs to the research-grade Neuroscan SynAmps2. Research-grade EEG systems typically involve long setup times with experimenters scratching participants' scalp and connecting the electrodes with gel. EPOC+, however, is quick to setup—about 10 minutes—and does not require extensive touching of the scalp or to use conductive gel: it uses cotton-soaked saline to bridge the electrode to the scalp. This makes EPOC+ a promising alternative to test autistic children who might not tolerate traditional EEG systems. Third, the use of multivariate pattern analyses (MVPA) increased our individual detection rate to 88%. MVPA consists of training a machine-learning algorithm to distinguish the neural signals recorded in conditions of interest (e.g. here, congruent versus incongruent). It uses the signal recorded from all the electrodes, providing two major advantages over univariate ERP analyses. First, by extracting discriminative information from any electrode, it allows us to capture signals coming from unexpected locations, which would be missed if we confined univariate analyses to a few chosen electrodes. This is important considering the inconsistencies in the topography of the N400 effect typically reported in groups (Barber, Doñamayor, Kutas, & Münte, 2010; Thornhill & Van Petten, 2012) and the large inter-individual variability in neural signals previously reported in individual children (Petit, Badcock, Grootswagers, & Woolgar, 2019; Chapter 3 of this thesis), which may be even greater in autistic children (Happé, Ronald, & Plomin, 2006). Second, because MVPA considers the signal recorded from all the electrodes at once, it is not necessary to correct for multiple comparisons across electrodes or regions of interest, making it a more statistically powerful method. Finally, MVPA searches for reliable differences in patterns of activity without requiring the difference to be in any particular direction across electrodes or people (Woolgar, Golland, & Bode, 2014). Even though this approach is less specific (i.e. it does not allow us to observe the direction or timing of an effect), it is more sensitive, which, for our purpose, is an advantage. Indeed, given that conditions differ only in their lexico-semantic context, any reliable difference in activation evoked by identical stimuli indicates differential integration of the stimulus with the preceding semantic content of the sentence. Previous research using MVPA allowed successful detection of lexico-semantic processing in individual adults (Geuze, Gerven, Farquhar, & Desain, 2013; Tanaka, Watanabe, Maki, Sakriani, & Nakamura, 2019), and children (Petit, Badcock, Grootswagers, & Woolgar, 2019), making it a promising avenue for individual-level detection in autistic children.

In the current proof-of-concept study, we sought to apply our sentence congruency paradigm, previously validated in neurotypical children (Petit et al., 2019), to assess lexico-semantic processing in three minimally-verbal autistic children. We recorded from either a research-grade EEG system (n = 2) or a consumer-grade EEG system (Emotiv's EPOC+; n = 1), depending on the tolerance of each child. We recorded the EEG responses to spoken words at the end of congruent and incongruent sentences, and analysed the signal using univariate analyses and MVPA.

5.2 Methods

5.2.1 Participants

Five autistic children were recruited through advertisements on social media, via word-of-mouth, and through specialist schools in New South Wales, Australia. All participants had a formal diagnosis of autism and had minimal spoken language. Two of the participants did not tolerate the EEG system (see below), so our final sample consisted of three participants (aged 11:5, 15:4, and 5:10, in the order presented hereafter, all male). Their spoken language was first assessed through parent report (less than 50 spoken words used meaningfully) and confirmed during the Autism Diagnostic Observation Schedule – 2nd edition (ADOS-2; Lord et al., 2012), as evidenced by 20 or fewer different meaningful words spoken during the assessment (Table 1). All participating families received $20/hour for their participation. This study was approved by the Macquarie University Human Research Ethics Committee (Reference number: 3358).

To characterise the cognitive profile of our participants we administered several standardised assessments. First, we asked caregivers to complete the Vineland-3 Comprehensive Parent/Caregiver form (Vineland Adaptive Behavior Scales; Sparrow, Saulnier, Cicchetti, & Doll, 2016), which characterises the adaptive behaviours of children, and the Lifetime version of the Social Communication Questionnaire (SCQ; Rutter, Bailey, & Lord, 2003), which asks parents to report on their children's communication skills and social functioning. Second, we administered the ADOS on either the first or second visit. Additionally, we attempted to administer the Peabody Picture Vocabulary Test – 4th edition (PPVT-4; Dunn & Dunn, 2007), which measures receptive vocabulary. We were not able to establish a basal set for the PPVT for any of our participants.

The first autistic participant (A1) was a boy aged 11 years and 5 months. He was diagnosed with autism and moderate intellectual disability when he was 2-and-a-half-years old. A1 attends year 5 in a special unit at a public school. His family speaks Korean and English at home and he attends a school taught in English. His mother reported that he could use single words and occasionally sentences, especially for requesting purposes. For the last year, A1 has used rapid prompting method (RPM) with his school teacher. RPM is an educational method which facilitates communication based on pointing to letters on a letterboard (American Speech-Language-Hearing Association, 2018). His teacher and his mom reported that he is highly motivated to learn, especially through RPM, and that consistent repetition and daily practice allows him to learn. He obtained a standardised score of 59 in the communication domain of the Vineland-3, more than two standard deviations below the population mean (100), indicating impaired communication. In particular, he was reported to use words and simple sentences to request but to have difficulties producing more complex sentences. He obtained a composite (overall) Vineland score of 56 and an ADOS score of 17, scoring above the cut-off of 15, commensurate with his diagnosis of autism. Finally, he scored 20 on the SCQ, well above the autism cut-off of 15, indicating impaired social communication. During the ADOS assessment, he spoke 20 different words, all single words except one sentence, which were used to request and comment on the activities (e.g. “balloon”, “finished”, “can I have the bubbles please”).

The second autistic participant was a boy aged 15 years and 4 months (A2). A2 was diagnosed with autism when he was three. Around the age of two, he started talking but his parents noticed he used set phrases that were not flexible. He underwent Applied Behaviour Analysis therapy (Lovaas, 1987; Virués-Ortega, 2010) until 9 years of age but his language did not seem to develop. He started using RPM with his mom around 10 years of age and could use it for effective communication (i.e. independent communication, outside of simply repeating answers to questions from his mother) from around 12 years of age. He has now been home-schooled for two years. His Vineland communication score was 38, indicating impaired communication. His Vineland report indicates that he was able to produce full sentences but had difficulties with complex sentences and questions, such as asking “where” and “when” questions and making two-parts sentences. During the ADOS assessment and the EEG sessions, he expressed himself in full sentences but these sentences, which he repeated several times, appeared to be either nonsensical or delayed echolalia of sentences he heard previously during the day, and were not appropriate to the context of the testing (e.g. “we’ve got dips in the car”, “I got you with coconut”, “there’s nothing to tell”, “take the train”). Therefore, although this participant produced more than 50 words, he still met the criteria for minimally-verbal as he did not use language in a meaningful way. His speech was also marked by strange intonation (excessive loudness and odd intonation). A2 scored 27 on this modified ADOS, indicative of severe autism. However, we must interpret A2’s ADOS scores with extreme caution for two reasons. First, we chose to administer the module 1 of the ADOS, based on his functional speech, which was extremely limited. However, due to his age, most of the materials from module 1 may have been poorly suited (i.e. too childish). We found it difficult to get A2 interested in any of the ADOS materials, potentially impairing his score. Second, we attempted to modify some of the activities to be more suited for his age (e.g. by swapping the toys used during the free play for some of the items used in module 3). This compromised the standardisation of the assessment as we did not use the exact same materials from the module 1 kit.

The third autistic participant was a boy aged 5 years and 10 months (A3). A3 received an autism diagnosis when he was five, after his parents noticed severe language delay. A3’s family speaks English and Russian at home. His mom reports that he seems to understand some language and has recently started saying a few words, such as colours and numbers. He started attending day care this year. His Vineland communication score was 38 and he spoke 5 words during the ADOS assessment, indicating spoken language impairment. He obtained a composite Vineland score of 54 and an SCQ score of 18, indicating difficulties in adaptive behaviour and social communication. Finally, his autism diagnosis was confirmed by an ADOS score of 25.

The brain responses of these children were compared to those of 20 neurotypical children aged 6 to 15 years (M = 11:2, SD = 2:9, 12 male, 8 female). Sixteen of these neurotypical children were tested as part of our previous methods-validation study (Petit et al., 2019) and an additional four were tested to supplement the group for the current study, following the same experimental protocol. All neurotypical children spoke English as their first language and had no reported history of brain injury, autism, or language impairment.

Table 1. Summary of communication abilities for autistic participants

Participant	A1	A2	A3
Language spoken at home	English and Korean	English	English and Russian
Child speaks	Y	Y	S
Total words spoken during ADOS	20	> 50	5
Says simple sentences (parent report)	Y	Y	N
Says the name of at least 10 objects (Vineland item)	Y	Y	N
Says at least 50 words (Vineland item)	Y	Y	N
Communication domain Vineland standard score	59	38	38
Follows 1-step command	Y	Y	Y
Alternative communication	RPM for 1 year	RPM for 6 years	N

Notes: Y = yes; N = no; S = sometimes. ADOS = Autism Diagnosis Observational Schedule-2; RPM = Rapid Prompting Method. Vineland standard scores are scaled so that the mean score for each participant's age group is 100, with a standard deviation (SD) of 15. As such, a score below 70 is two SD below the mean.

Table 2. Standardised Assessment of autistic participants

Participant		**A1**	**A2**	**A3**
Age		11:7	15:4	5:10
Vineland standard scores (with 95% confidence intervals)	Communication	59 (54 – 64)	38 (34 – 42)	38 (34 – 42)
	Daily living skills	65 (60 – 70)	41 (36 - 46)	60 (56 – 64)
	Socialisation	28 (24 – 32)	36 (32 – 40)	50 (46 – 54)
	Composite	56 (53 – 59)	43 (40 – 46)	54 (52 – 56)
SCQ		20	27	18
ADOS	Social affect	15	20	17
	Restricted and repetitive behaviours	2	7	8
	Total	17	27	25

Note: SCQ = Social Communication Questionnaire; ADOS = Autism Diagnostic Observational Schedule-2. Vineland standard scores are scaled so that the mean score for each participant's age group is 100, with a standard deviation (SD) of 15. As such, a score below 70 is two SD below the mean. SCQ scores range from 0 to 39, with higher scores indicating more social communication impairments. SCQ scores are interpreted compared to a cut-off of 15, with scores above the cut-off suggesting a diagnosis of autism. ADOS scores range from 0 to 28, with higher scores indicating larger difficulties. For children with few to no words (e.g. A3), a score of 16 and above suggest a diagnosis of autism, while for children with some words (e.g. A1 and A2), a score of 12 and above suggest a diagnosis of autism.

5.2.2 Stimuli

As described in the validation study by Petit et al. (2019), stimuli comprised 112 spoken sentences, half of them congruent and half of them incongruent. The congruent sentences were based on the list from Block and Baldwin (2010), consisting of 400 high-close-probability sentences, meaning that at least 67% of their adult participants completed a given sentence stem (e.g. "she went to the bakery for a loaf of") with the same target word (e.g. "bread"). Of these sentences, we selected 56 that were appropriate for children, for which the target word was of high-frequency and likely acquired by age 5 (Kuperman et al., 2012). These sentences formed the congruent condition. To create a balanced incongruent condition, we recombined the sentence stems and target words, such that the incongruent target words would be unexpected but grammatically correct. We also ensured that the incorrect target did not rhyme or start with the same sound as the correct target. The final set of stimuli thus consisted of 56 congruent and 56 incongruent sentences.

All stimuli were recorded by a female native British English speaker in a sound-proof room. The sentence stems and target words were recorded separately to avoid possible co-articulation, and later recombined with a 100 ms gap between the sentence stems and the targets.

5.2.3 EEG equipment

For neurotypical participants, we recorded the EEG signals using two EEG systems simultaneously (Petit et al., 2019). The research-grade Neuroscan SynAmps2 (Scan version 4.3) was first setup on the head through an elastic cap (EasyCap, Herrsching, Germany) that held 34 Ag-AgCl electrodes placed according to the international 10-20 system, including M1 (online reference), AFz (ground electrode), and M2. We also recorded eye movements from four electrodes placed on the outer canthi of each eye, and above and below the right eye.

Neuroscan data was sampled at 1000 Hz with an online bandpass filter from 1 to 100 Hz. We ensured all impedances were below 10 kΩ, before setting up the second EEG system. The EPOC+ system was then placed over the Neuroscan setup, with electrodes connecting the scalp through custom-made slits in the EasyCap. Electrodes were bridged to the scalp using saline-soaked cotton pieces (i.e. cotton-rolls). The EPOC+ is a portable headset with flexible plastic arms holding 16 gold-plated sensors according to the international 10-20 system (Figure 1), including M1 (online reference; i.e. common mode sense), and M2 (feed-forward reference used to reduce external electrical interference; i.e. driven-right leg). Two electrodes (O1 and O2) were used as part of a custom-built event-marking system, which consisted of a transmitter box plugged in to the presentation computer and a receiver box attached to the headset (Badcock et al., 2015; Thie, 2013). Simultaneous to the presentation of the target words, a tone was fed to the transmitter box, which converted it into an infrared light pulse. This was sent to the receiver box and converted into an electrical signal which was inserted in O1 and O2. The signals from the other 12 electrodes were high-pass filtered online with a 0.16 Hz cut-off, pre-amplified and low-pass filtered at an 83 Hz cut-off. The analogue signals were then digitised at 2048 Hz. The digitised signal was filtered using a 5th-order sinc filter to notch out 50Hz and 60 Hz, low-pass filtered and down-sampled to 256 Hz (specifications taken from the EPOC+ system web forum). The effective bandwidth was 0.16–43 Hz. We ensured all impedances were 'green' in the TestBench software. In Petit et al. (2019), we found that EPOC+ recorded ERPs of similar shape and amplitude to Neuroscan, making it a viable alternative to Neuroscan. However, we found that EPOC+'s data was not particularly well suited for multivariate pattern analyses, which were the most sensitive analyses on the Neuroscan data.

For the autistic children, we met each child and their family in advance of their first EEG session in order to establish contact and to initiate desensitisation to the EEG materials. For this purpose, we provided the parents with an EEG cap (EasyCap, Herrsching, Germany) and instructed them on how to set it up on the head and strap it under the chin with the Velcro strap. We also provided the families with a printed, bespoke social story (Gray, 1994), picturing the steps involved in the EEG setup, so they could review them with their child in advance of the EEG session. On the EEG test day, depending on the level of tolerance of each child, we used Neuroscan or EPOC+. Our preferred choice was Neuroscan because of its higher sensitivity. To reduce the setup time, we only recorded from 15 electrodes, excluding AFz (see Figure 1). However, if a child could not tolerate the EasyCap or the electrodes attached to the mastoids (M1 and M2), we recorded from EPOC+ (all the EPOC+ electrodes were setup, see Figure 1).

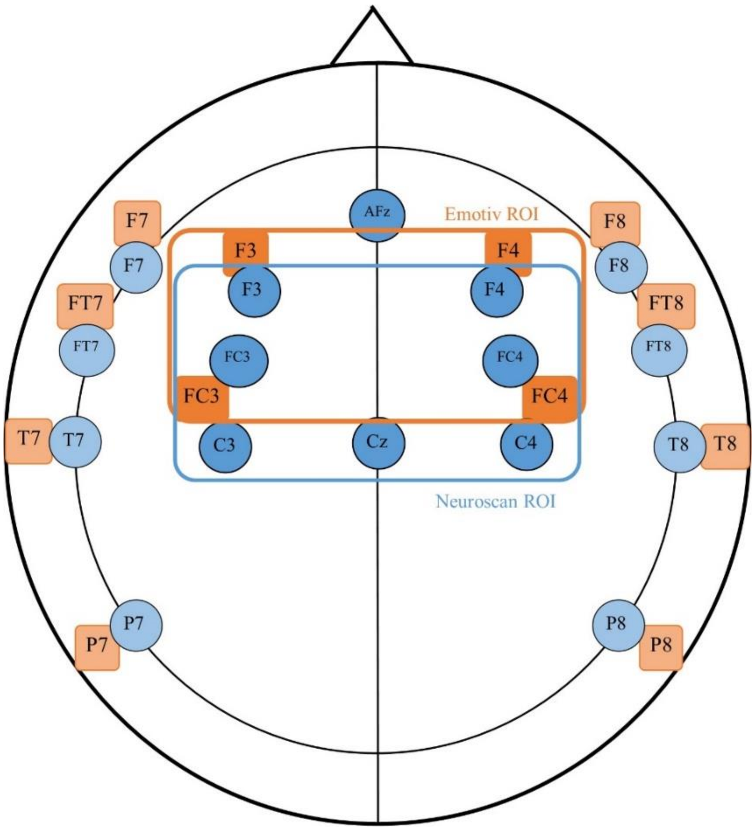


Figure 1. Electrode position on the scalp for the Neuroscan (blue circles) and EPOC+ (orange rectangles) systems for testing autistic participants. The regions of interest for each system are shown.

5.2.4 EEG experimental procedure

We used visual aids to help familiarise the autistic participants with the lab and the experimental procedures. Once we established that the participant was comfortable, we setup the chosen EEG system (for autistic participants) or both EEG systems (for neurotypical participants). Then, participants were seated in front of a monitor and began the experiment. We presented all stimuli using Psychophysics Toolbox 3 (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) in MATLAB (R2017b). In an initial practice block, we explained that evil aliens were interfering with messages that we were sending to our astronaut friends. Participants had to listen to each message and count the number of "incorrect" messages (i.e. incongruent sentences) in each block. We then presented five sentences at a comfortable listening volume. Two sentences were incongruent and three were congruent, and we asked participants to answer how many were incorrect. No autistic participant was able to answer at this stage, which meant that we could not check that they understood the instructions. In contrast, all neurotypical participants answered correctly. Before starting the experimental blocks, we repeated the instructions, emphasising that they must decide in their head, for each sentence, if it was correct or incorrect.

After the practice, participants took part in two 20-minute sessions. In each session, all 112 sentences (56 congruent, 56 incongruent) were presented. Each session was divided into 16 blocks of 4 to 10 trials, with short breaks between each block. To reinforce the story and cue children when to attend, we presented an image of a satellite centrally on the screen to signal the onset of each trial, and kept this display on for the whole trial. Two seconds after the satellite appeared, we presented the sentence through the speakers. The satellite remained onscreen for a further 1.5 s after the presentation of the target word. There was then a 2 s inter-trial-interval before the next trial, during which a fixation cross was shown on a blank screen (Figure 2). The order of the sentences was the same for all the children and was pseudorandomised to prevent order effects (see Petit et al., 2019). All neurotypical participants, and one autistic participant (A1), completed the two sessions on the same day, while the remaining two autistic participants completed the sessions over two non-consecutive days (13 days apart from A2 and 9 for A3).

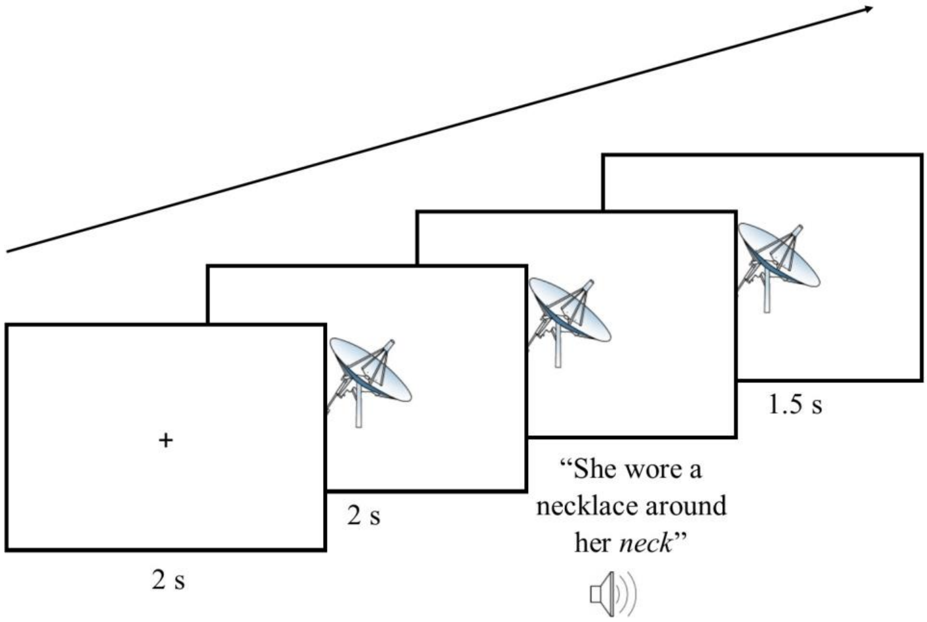


Figure 2. Trial sequence during the EEG experiment. A picture of a satellite appeared on screen for 2 seconds before the sentence was played, and remained on screen for 1.5 seconds after.

5.2.5 EEG data processing

5.2.5.1 Preprocessing

We processed all EEG signals in EEGLAB (v13.4.4b, Delorme and Makeig, 2004) in MATLAB (R2017b). The signals from Neuroscan and EPOC+ were analysed the same way, with the exception of the sampling rate (500 Hz for Neuroscan, 256 Hz for EPOC+). We first applied a band-pass filter between 0.1 and 40Hz, then cut the data into epochs from -100 ms to 1000 ms around target word onset. We ran an Independent Component Analysis (ICA) on all the epochs. The EPOC+ ICA did not reveal eye movements artefacts, as seen in Badcock et al.

(2015) and Petit et al. (2019), possibly because we did not record from face electrodes which provide the strongest eye-movement signals. Components corresponding to eye-blinks and eye-movements were removed from the Neuroscan data. We applied a baseline correction to the epochs by subtracting the averaged signal from the 100 ms preceding target onset. We then used these minimally-cleaned data for multivariate pattern analyses (MVPA). For univariate analyses, we further removed epochs with extreme values (±150μV) from each electrode separately. For Neuroscan, an average of 20 epochs (SD = 17, min = 0, and max = 55) were rejected for the neurotypical participants and 44 epochs for the autistic participants (72 for A1 and 16 for A2). For the EPOC+, an average of 16 epochs (SD = 18, min = 1, and max = 76) were rejected for the neurotypical participants and 112 epochs (50% of the data) were rejected for the autistic participant.

5.2.5.2 Multivariate pattern analyses

We first analysed the Neuroscan data using a temporally- and spatially-unconstrained MVPA approach that we developed previously to maximise individual sensitivity to lexico-semantic processing in typical children (Petit, Badcock, Grootswagers, & Woolgar, 2019). For this, we used a linear support vector machine (SVM) classifier to test for consistent patterns of neural responses that discriminated between our two conditions (congruent and incongruent). In this analysis, we concatenated the data from all the EEG electrodes and time points, for each trial. We then separated the data into a training set and a testing set. The training set consisted of all of the trials but four, corresponding to the four presentations of a single target word (which appeared twice in the congruent condition and twice in the incongruent condition). The testing set consisted in these four left-out trials. We then trained the classifier to establish a decision boundary that best discriminated the conditions based on the training data, then we tested the accuracy in classifying the testing set. On this scheme, successful classification can only result from different responses to the two conditions which generalise over different words (i.e. from training to test set) and discriminate between presentations of the identical auditory token (the left-out word) in the two conditions. We then repeated this procedure 56 times, leaving out a different target each time, and averaged the accuracies of all the repetitions to get a decoding accuracy for each participant. To determine whether the accuracy was significantly above chance (50%), we ran a permutation test. This consisted of randomly shuffling the condition labels of the data ('congruent' or 'incongruent'), then repeating the decoding as described above. We repeated this permutation 1000 times to yield a null distribution of accuracies. Our actual accuracy was considered significantly above chance if it was bigger than 95% of the permutations' accuracies ($\alpha = .05$). For neurotypical participants, all 32 electrodes were used by the classifier (i.e. not M1 and M2). In contrast, we only set up 15 electrodes for the autistic children. To check whether the additional number of electrodes for neurotypical participants impacted the decoding results, we ran the same decoding with only the 15 electrodes matching the autistic group. We compared the results using a paired-sample t-test. We also compared the decoding accuracy of Neuroscan (both with all electrodes and with 15 electrodes) and EPOC+ using paired-sample t-tests.

5.2.5.3 Univariate analyses

We then examined single-subject univariate effects, i.e. the difference between the congruent and incongruent ERPs over time. We defined a region of interest (ROI) based on previous studies that found the N400 effect to be maximal over fronto-central areas (Cruse et al., 2014; Petit, Badcock, Grootswagers, Rich, et al., 2019). Our ROI for Neuroscan was FC3, FC4, C3, C4, Cz, F3, and F4, and for EPOC+ was F3, F4 FC3, and FC4 (see Figure 1). Averaging more electrodes together (7 electrodes for Neuroscan versus 4 electrodes for EPOC+) should provide an advantage as it increases the signal-to-noise ratio. However, as we were not directly comparing Neuroscan and EPOC+, we chose to maximise the number of electrodes in each system's ROI that surrounded Cz, resulting in a different number of electrodes for the two systems. We first averaged all the epochs that were not rejected during pre-processing (i.e. clean epochs) within the ROI, then we analysed the ERP difference (congruent-minus-incongruent ERP) using independent-sample t-tests at each time point. We corrected for multiple comparisons using a temporal-cluster correction (Guthrie & Buchwald, 1991). This method estimates the minimum number of consecutive time points that need to be significant at an uncorrected threshold of $p < 0.05$ to be considered a significant cluster after correction for multiple comparisons (Guthrie & Buchwald, 1991; detailed in Petit, Badcock, Grootswagers, Rich, et al., 2019).

5.3 Results

5.3.1 Behavioural results

All participants were instructed to count how many sentences were incorrect during each block. Neurotypical participants were highly accurate (mean percent correct: M = 96.29%, SD = 3.36%, range = [88.4%, 100%]), indicating that they were making correct semantic judgements for each sentence. None of the autistic children were able to give overt responses.

5.3.2 Multivariate pattern analyses

We first evaluated the accuracy of an SVM classifier to distinguish between neural signals for the semantically congruent and incongruent conditions for Neuroscan and EPOC+ data. Using Neuroscan, we could differentiate the conditions in 18/20 neurotypical participants (90%), indicating a good level of sensitivity (Figure 3). Decoding was also significantly above chance for one of the two autistic participants scanned with this device (Figure 3, right panel). This suggests intact lexico-semantic processing of the sentences for this participant.

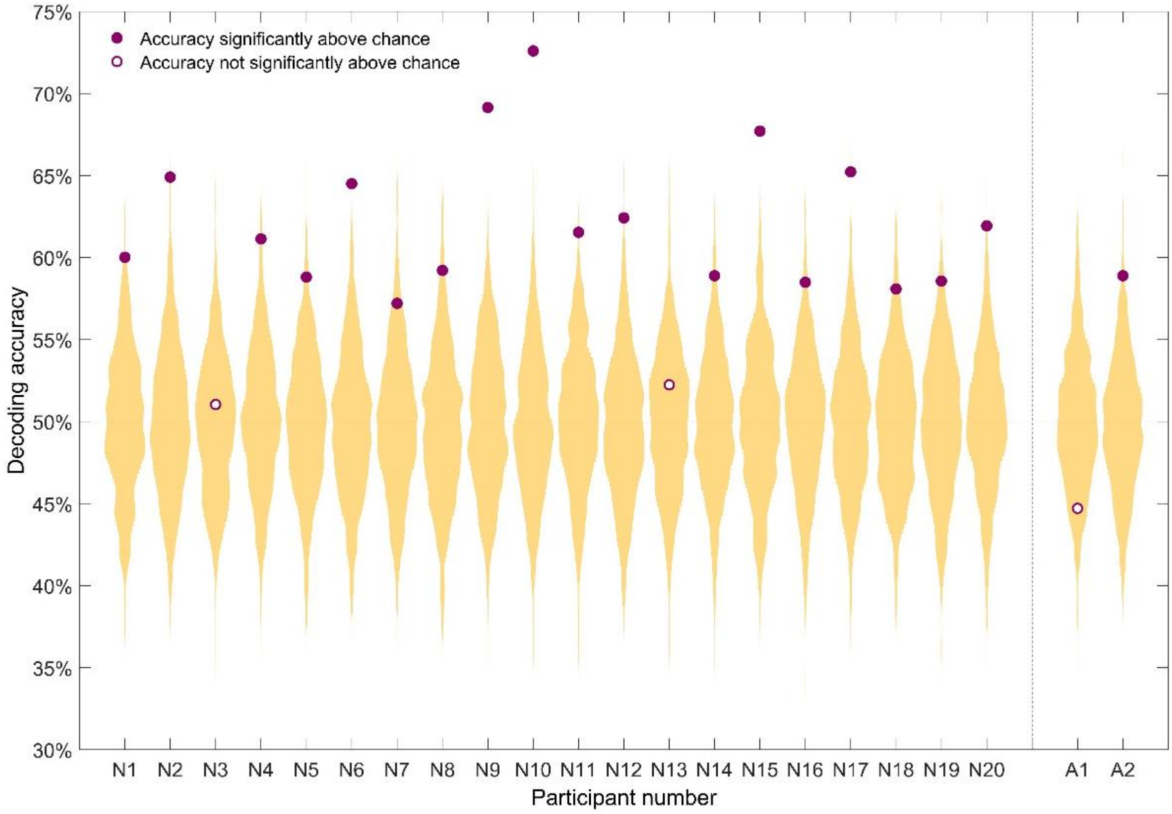


Figure 3. Individual decoding accuracy for classification of identical target words in congruent and incongruent contexts using Neuroscan. N indicates neurotypical participants and A indicates autistic participants. Purple circles indicate decoding accuracy for each participant. The yellow distribution shows the null distribution obtained by the permutation test for that participant. 90% of neurotypical participants and one autistic participant had a decoding accuracy significantly above chance (full purple circles). Neurotypical participants are ordered by age; A1's age would fit between N9 and N10, and A2's age would fit between N16 and N17.

Because the neurotypical data comprised more electrodes than the autistic data (33 versus 15, respectively), we checked whether decoding accuracy was impacted by number of electrodes. For this, we repeated the analysis on the neurotypical participants' data, using only the 15 electrodes used with the autistic participants. There was no evidence that reducing the number of electrodes influenced decoding accuracy ($t_{(19)} = 1.60$, $p = 0.13$; see Supplementary Figure 1).

Next, we considered the data from EPOC+ (Figure 4). In the neurotypical participants, decoding accuracy was significantly lower for EPOC+ compared to both Neuroscan with all electrodes ($t_{(19)} = 3.85$, $p = .0011$) and Neuroscan with only 15 electrodes ($t_{(19)} = 3.29$, $p = .0039$; see Supplementary Figure 1). Significant effects were detected in only 35% of typical controls. This lower decoding accuracy for EPOC+ has previously been reported and may reflect numerically lower sensitivity in this system (Petit, Badcock, Grootswagers, Rich, et al., 2019). Nonetheless, the autistic participant tested with EPOC+ (A3) was significantly above chance and, in fact, showed the highest classification accuracy of the group at 70% correct. This suggests that this participant engaged in lexico-semantic processing of the linguistic stimuli sufficient to drive distinct neural responses detectable with this technology.

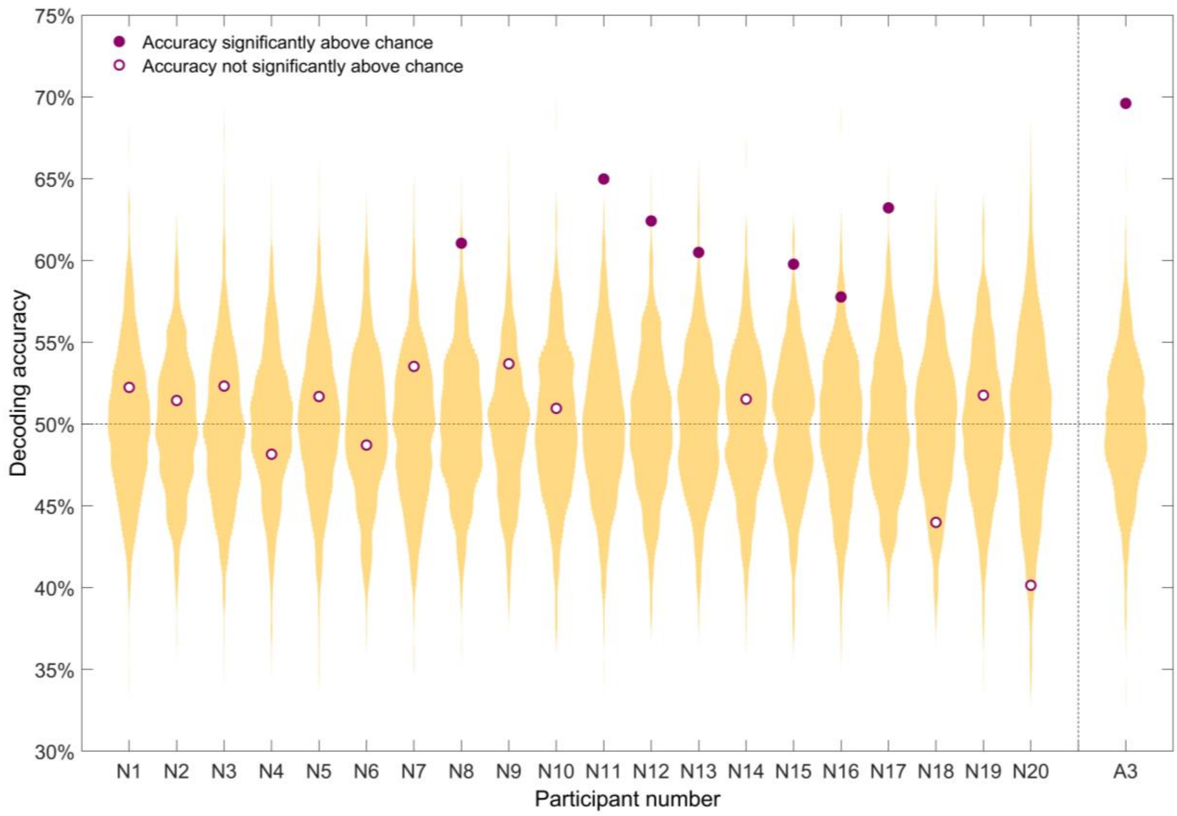


Figure 4. Individual decoding accuracy for classification of identical target words in congruent and incongruent contexts, using EPOC+. N indicates neurotypical participants and A indicates autistic participants. Purple circles indicate decoding accuracy for each participant. The yellow distribution shows the null distribution obtained by the permutation test for that participant.

35% of neurotypical participants and one autistic participant had a decoding accuracy significantly above chance (full purple circles). Chance (50%) is indicated by the horizontal dashed line. Neurotypical participants are ordered by age: A3's age would fit before N1.

5.3.3 Univariate analyses

Next, we further characterised the responses to the two conditions by examining univariate differences in response within a region of interest centred around Cz (see Figure 1). In Figure 5, we show the ERPs for the controls using Neuroscan and autistic participants 1 and 2. In the neurotypical group, 16 participants showed at least one significant cluster of significant time points, indicating a significant univariate effect. In the autistic group, the one participant (A2) who had shown a significant multivariate effect also showed a significant univariate effect, with two immediately adjacent significant clusters (from 148 to 480 ms). In Figure 6, we show the ERPs for the controls using EPOC+ and autistic participant 3. The sensitivity of EPOC+ was numerically lower, with 14 neurotypical participants showing a significant effect. Participant A3 showed a reversal of the typical N400 effect, with a more negative ERP in the congruent than the incongruent condition.

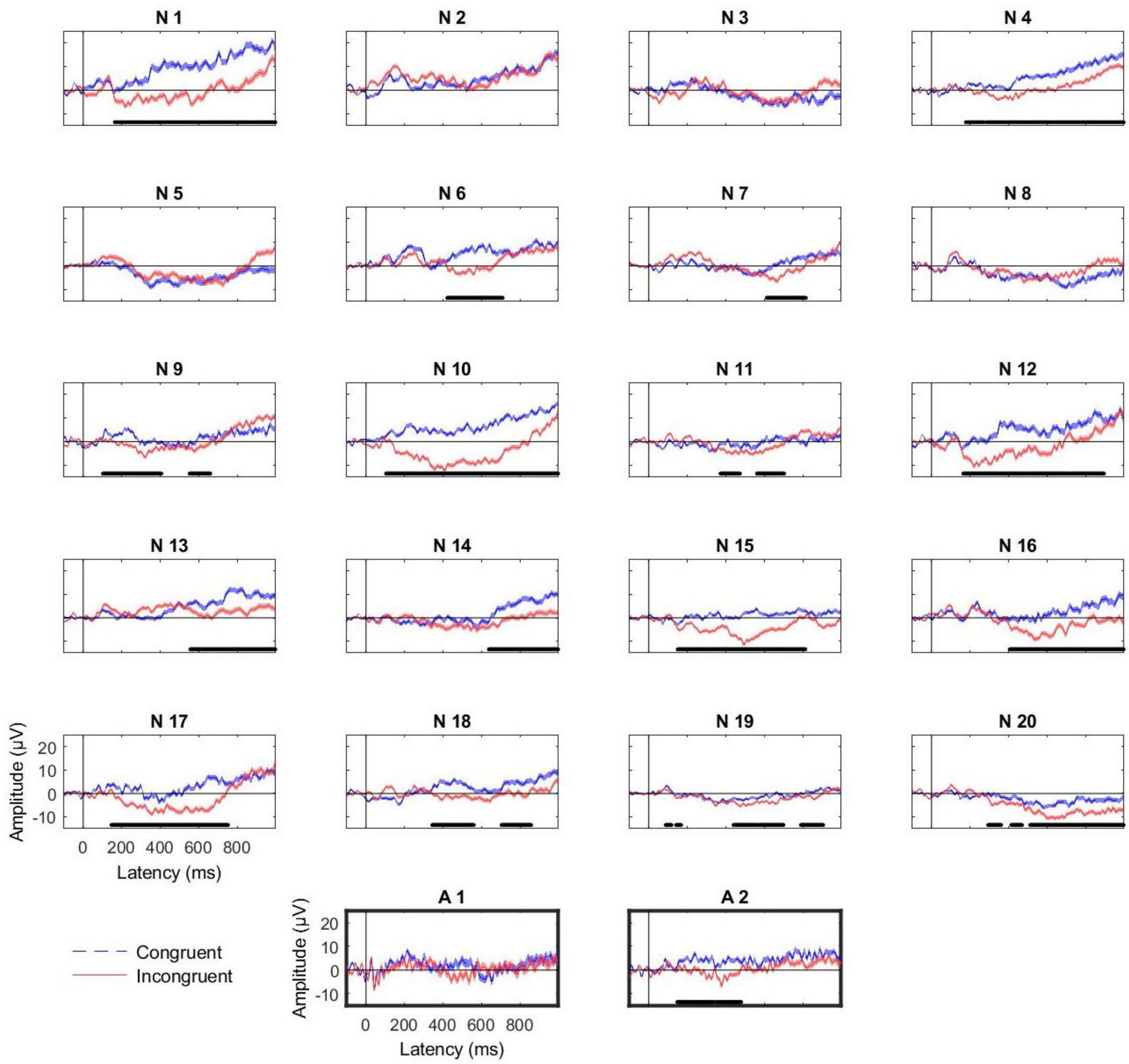


Figure 5. Individual participant event-related potentials to target words following congruent (dashed blue) and incongruent (solid red) sentences in our region of interest, measured with Neuroscan. Shading indicates standard error of the mean. Clusters of statistically significant N400 effects are indicated with a solid, horizontal, black line. N indicates neurotypical participants and A indicates autistic participants. 80% (16/20) of controls and one autistic participant showed a statistical univariate N400 effect. Participants are ordered by age, as in Figure 3.

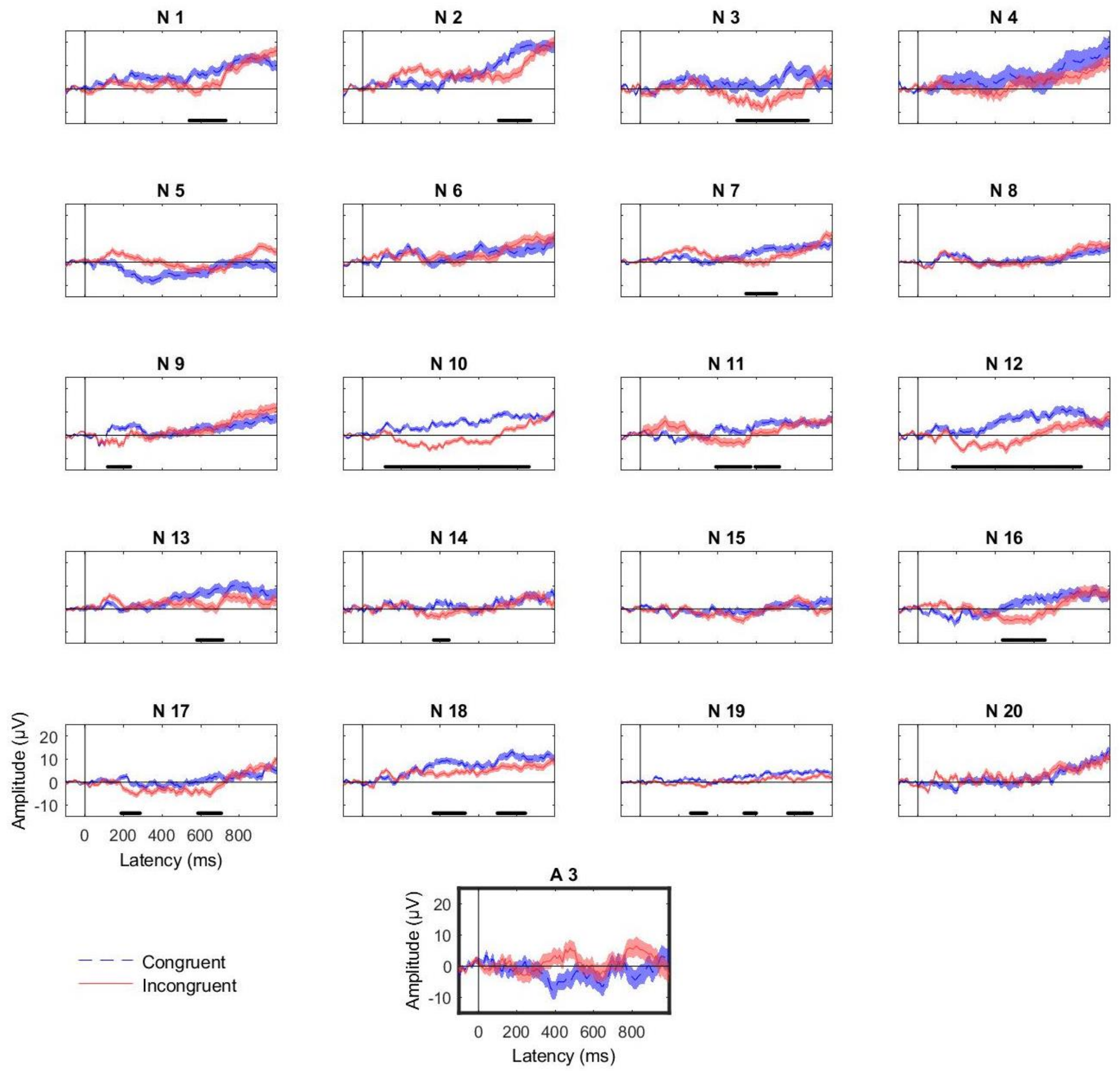


Figure 6. Individual participant event-related potentials to target words following congruent (dashed blue) and incongruent (solid red) sentences in our region of interest, measured with EPOC+. Shading indicates standard error of the mean. Clusters of statistically significant N400 effects are indicated with a solid, horizontal, black line. N indicates neurotypical participants and A indicates autistic participants. 70% (14/20) of controls showed a statistical univariate N400 effect, while the autistic participant did not. Participants are ordered by age, as in Figure 3.

5.4 Discussion

Despite the urgent need to develop reliable tests of cognition for minimally-verbal autistic individuals, there is a paucity of research on this population. The current proof-of-concept study set out to use our recently-developed auditory N400 paradigm to assess lexico-semantic processing in minimally-verbal autistic children. We presented three autistic children and 20 neurotypical children with sentences ending with congruent or incongruent words. We asked participants to covertly decide on each sentence's semantic validity and to count the number of incorrect sentences, if possible. Having previously validated the Emotiv EPOC+ as a viable but less sensitive alternative to the research-grade Neuroscan system (Petit et al., 2019), we tailored the materials to the tolerance of each child. Two of the autistic children tolerated Neuroscan, while one did not and was therefore tested with EPOC+. We analysed the data using both multivariate and univariate analyses for each individual. Using Neuroscan, we decoded the semantic condition in 90% of the neurotypical children with MVPA, indicating that they engaged in different neural processes in the two semantic conditions. In addition, 80% of neurotypical children showed a statistically significant univariate N400 effect. Of the two autistic participants tested with Neuroscan, one showed significant decoding and a significant univariate N400 effect (A2), indicating preserved receptive lexico-semantic processing. The other autistic participant did not show neural evidence of lexico-semantic processing (A1). Using EPOC+, we decoded the semantic condition in 35% of neurotypical participants and found a significant N400 effect in 70% of neurotypical participants. The autistic participant tested with EPOC+ (A3) showed significant decoding but showed a reversed N400-like effect (congruent condition more negative than incongruent condition).

The neural data from A2 suggest that he engaged in differential processing of identical spoken words depending on the lexico-semantic context in which they were presented. His neural processing was consistently different when the target word was congruent with, and/or predicted by, the context compared to when it was not (multivariate analysis), and the temporal profile of this difference matched that of typically developing controls (univariate analysis). We did not present the topographic maps of neural activation over time as the electrodes layout used for the autistic participants did not cover the whole head (i.e. less coverage of the centro-parietal than the frontal regions). Thus, it is currently not possible to describe the spatial profile of neural activation. Nevertheless, these data imply that he was able to integrate the identical auditory token with its lexico-semantic context. We take this as evidence of receptive language ability that is not necessarily apparent from his behaviour or his profile of relatively proficient but meaningless use of spoken language.

At the time of testing, A2 had been using RPM for five years. We deliberately included children with alternative communication in our sample even though some definitions of “minimally-verbal” would not include them (e.g. Tager-Flusberg & Kasari, 2013). They were included because they did not demonstrate functional speech, in line with the classic definition of minimally-verbal. Furthermore, in light of the ongoing controversy regarding RPM and the lack of scientific evidence to support its validity (Schlosser et al., 2019), our test provides independent evidence of lexico-semantic processing in this individual. We also note that A2 was our oldest autistic participant (15 years) but we did not find evidence for an impact of age on decoding effects in neurotypical children (as indicated by non-significant Pearson’s correlation between decoding accuracy and age: $r = .10$, $p = .66$; see Supplementary Figure 2).

The results from participant A3 were intriguing. A3 was tested with EPOC+, which has a lower sensitivity than Neuroscan for decoding (based on the neurotypical participants’ data). Yet, A3 showed statistically reliable decoding (in fact, higher decoding accuracy than any neurotypical participant), suggesting that he consistently engaged in different neural processes for the congruent versus incongruent sentences. However, A3’s univariate data showed a *reversed* N400-like pattern, with more negative potential for congruent than incongruent sentences. A similar N400 reversal was reported by DiStefano et al. (2019) for two subgroups of autistic children in their picture-word paradigm. One subgroup consisted of four verbal autistic children who showed a robust ERP difference between the two semantic conditions, but with a more negative potential for the congruent than the incongruent condition. The other consisted of two minimally-verbal children who showed an initially reversed N400-like effect, followed by a typical response at longer latencies. It is unclear why this reversed pattern occurs, and whether or not to take it as evidence for lexico-semantic processing.

A similar N400 reversal has previously been observed in visual-masking paradigms, where a prime word is masked (i.e. it does not reach conscious access), then a target word is presented. In these paradigms, a larger N400 for related compared to unrelated targets has been observed and coined the "negative priming effect" (Bermeitinger, Frings, & Wentura, 2008; Wentura & Frings, 2005). This effect has been explained in terms of center-surround inhibition (Carr & Dagenbach, 1990), where a masked prime only weakly activates its node. To maintain this relative activation, surrounding semantic nodes are inhibited, leading to a more difficult retrieval of semantic words related to the masked prime, which is observed as an increased N400. In the case of spoken sentences, it is possible that a reversed N400 reflects higher resource allocation to the semantic content of the sentence frame, to the detriment of sentence-level semantic integration. In turn, this could inhibit the related semantic target word. This effect may thus reflect weaker or impaired, but not absent, semantic processing. A much larger sample size is needed to test this possibility and to extract possible individual factors that contribute to this cognitive phenomenon. Previous research has also found reversed N400-like effects in children but only for words outside their vocabulary range (Byrne, Dywan, & Connolly, 1995). As this participant was the youngest and his family are bilingual, it is possible that the word stimuli we used were not known by this participant. Given these considerations, we interpret the results from A3 with great caution. We recommend further testing of this participant, perhaps using more sensitive EEG systems (see below) if he can tolerate them in the future. It would also be interesting to examine if/how this pattern changes as he matures.

We did not detect neural evidence of lexico-semantic processing in the remaining autistic participant (A1). It is possible therefore, that this child was yet to acquire lexico-semantic processing at the time of testing. However, lack of evidence for a difference must not be taken as evidence for a lack of lexico-semantic processing. Instead, care must be taken in interpreting this null effect. In neurotypical children, whose lexico-semantic processing is not in question, we were unable to reach 100% detection rate, even when using the least constrained approach of multivariate decoding of the data from all the electrodes and time points. It is possible that children not showing an effect failed to do the task or that the size of the effect was too small for us to pick up. In the case of neurotypical children, their behavioural accuracy for semantic judgements was near perfect, suggesting that they performed the task accurately. For the autistic children, however, we were unable to verify whether they were judging the semantics of the sentences. Even though we reminded them to decide in their head whether each sentence was correct or incorrect, it is possible that they did not understand, lacked motivation, or could not follow the instructions (Tager-Flusberg et al., 2017). Thus, at this stage, our test is limited to inference *for* preserved lexico-semantic processing and does not allow us to draw any conclusions regarding an absence of processing.

Contrary to previous research that used picture-words paradigms, we used purely auditory stimuli with congruent and incongruent spoken sentences. In addition to not requiring visual attention, the advantage of our approach is that a positive result is evidence of sentence-level comprehension: a higher form of lexico-semantic integration compared with picture-word matching. However, we do note that a positive result may also be driven by learnt associations between key words in the sentences. For example, the words “birthday” and “cake” are heard together more often than “birthday” and “door”, therefore “door” may be more surprising than “cake” in the context “There were candles on the birthday door” simply because these two words are usually not presented together. In addition, a drawback of our paradigm is that sentence-level processing requires processing of phonology, syntax, and semantic, in addition to working memory to encode the meaning of the full sentences. It is possible that some minimally-verbal autistic children lack such high-level linguistic processing, while retaining single-word knowledge. In this case, our paradigm may fail to detect this ability and for this a picture-word paradigm would be more appropriate. Future tests may incorporate these two approaches to complement each other and evaluate the level of lexico-semantic processing.

This study served as a proof of concept that our methods can be used with minimally-verbal autistic children. However, in addition to the three participants tested successfully, we attempted to test two more children without success. The main issue was tolerance to either EEG system, despite habituation and desensitisation to the materials. It is likely that the hyper-sensory atypicalities experienced by autistic individuals (Posar & Visconti, 2018; Tomchek & Dunn, 2007) are problematic when setting up electrodes on their heads, even with the more tolerable EPOC+. For these reasons, it might be beneficial to turn to other EEG devices that are currently being independently validated. For example, a recent validation of the Cognionics 64-electrodes EEG system for classification of single-trials is encouraging (Callan, Durantin, & Terzibas, 2015), although the authors did not compare it to a traditional EEG system, making it difficult to assess recording quality. This particular system is active, allowing for better impedances without the need for scratching the scalp, and comprises a full 64-electrodes array. Future research should aim to provide the most tolerable testing environment to increase comfort and engagement for special populations.

Taken together, the significant evidence for lexico-semantic processing in our oldest and most verbal autistic child A2, in line with previous findings of neural evidence of lexico-semantic processing in minimally-verbal autistic children (Cantiani et al., 2016; DiStefano et al., 2019), points to cases of preserved receptive language abilities despite limited speech. Further work is needed to establish the prevalence of such abilities and the extent to which they apply to completely non-speaking children, which may have transformative implications for the care and education of autistic children.

5.5 Conclusion

We present the neural responses of three autistic children with minimal functional spoken language during an entirely auditory N400 paradigm. We found neural evidence of lexico-semantic processing in one child, as seen by significant multivariate decoding of semantic conditions and significant univariate N400 effects. This study represents a proof-of-concept that a covert auditory paradigm can be applied to infer lexico-semantic processing of spoken words in minimally-verbal autistic children. Future research using these methods on a larger sample will allow us to investigate the prevalence of intact receptive language processing in this population, currently neglected in the research literature.

5.6 References

American Speech-Language-Hearing Association. (2018). *Rapid Prompting Method*.

Atchley, R. A., Rice, M. L., Betz, S. K., Kwasny, K. M., Sereno, J. A., & Jongman, A. (2006). A comparison of semantic and syntactic event related potentials generated by children and adults. *Brain and Language*, *99*(3), 236–246. https://doi.org/10.1016/j.bandl.2005.08.005

Badcock, N. A., Preece, K. A., Wit, B., Glenn, K., Fieder, N., Thie, J., & McArthur, G. (2015). Validation of the Emotiv EPOC EEG system for research quality auditory event-related potentials in children. *PeerJ*, *3*, e907. https://doi.org/10.7717/peerj.907

Barber, H. A., Doñamayor, N., Kutas, M., & Münte, T. (2010). Parafoveal N400 effect during sentence reading. *Neuroscience Letters*, *479*(2), 152–156. https://doi.org/10.1016/j.neulet.2010.05.053

Behrmann, M., Thomas, C., & Humphreys, K. (2006). Seeing it differently: visual processing in autism. *Trends in Cognitive Sciences*, *10*(6), 258–264. https://doi.org/10.1016/j.tics.2006.05.001

Bermeitinger, C., Frings, C., & Wentura, D. (2008). Reversing the N400: event-related potentials of a negative semantic priming effect. *NeuroReport*, *19*(15), 1479. https://doi.org/10.1097/WNR.0b013e32830f4b0b

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.

Byrne, J. M., Dywan, C. A., & Connolly, J. F. (1995). An innovative method to assess the receptive vocabulary of children with cerebral palsy using event-related brain potentials. *Journal of Clinical and Experimental Neuropsychology*, *17*(1), 9–19. https://doi.org/10.1080/13803399508406576

Callan, D. E., Durantin, G., & Terzibas, C. (2015). Classification of single-trial auditory events using dry-wireless EEG during real and motion simulated flight. *Frontiers in Systems Neuroscience*, *9*, 11.

Cantiani, C., Choudhury, N. A., Yu, Y. H., Shafer, V. L., Schwartz, R. G., & Benasich, A. A. (2016). From Sensory Perception to Lexical-Semantic Processing: An ERP Study in Non-Verbal Children with Autism. *PLOS ONE*, *11*(8), e0161637. https://doi.org/10.1371/journal.pone.0161637

Carr, T. H., & Dagenbach, D. (1990). Semantic Priming and Repetition Priming From Masked Words: Evidence for a Center-Surround Attentional Mechanism in Perceptual Recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(2), 341–350. https://doi.org/10.1037/0278-7393.16.2.341

Coderre, E. L., Chernenok, M., O'Grady, J., Bosley, L., Gordon, B., & Ledoux, K. (2019). Implicit Measures of Receptive Vocabulary Knowledge in Individuals With Level 3 Autism. *Cognitive and Behavioral Neurology*, *32*(2), 95. https://doi.org/10.1097/WNN.0000000000000194

Cruse, D., Beukema, S., Chennu, S., Malins, J. G., Owen, A. M., & McRae, K. (2014). The reliability of the N400 in single subjects: Implications for patients with disorders of consciousness. *NeuroImage : Clinical*, *4*, 788–799. https://doi.org/10.1016/j.nicl.2014.05.001

DiStefano, C., Senturk, D., & Spurling Jeste, S. (2019). ERP Evidence of Semantic Processing in Children with ASD. *Developmental Cognitive Neuroscience*, 100640. https://doi.org/10.1016/j.dcn.2019.100640

Dunn, L. M., & Dunn, D. M. (2007). PPVT-4: Peabody picture vocabulary test. *Pearson Assessments*.

Fishman, I., Yam, A., Bellugi, U., Lincoln, A., & Mills, D. (2011). Contrasting patterns of language-associated brain activity in autism and Williams syndrome. *Social Cognitive and Affective Neuroscience*, *6*(5), 630–638. https://doi.org/10.1093/scan/nsq075

Gernsbacher, M. A. (2017). Editorial Perspective: The use of person-first language in scholarly writing may accentuate stigma. *Journal of Child Psychology and Psychiatry*, *58*(7), 859–861. https://doi.org/10.1111/jcpp.12706

Geuze, J., Gerven, M. A. J. van, Farquhar, J., & Desain, P. (2013). Detecting Semantic Priming at the Single-Trial Level. *PLOS ONE*, *8*(4), e60377. https://doi.org/10.1371/journal.pone.0060377

Gray, C. (1994). Social stories. *Arlington, TX: Future Horizons*.

Guthrie, D., & Buchwald, J. S. (1991). Significance testing of difference potentials. *Psychophysiology*, *28*(2), 240–244.

Happé, F., Ronald, A., & Plomin, R. (2006). Time to give up on a single explanation for autism. *Nature Neuroscience*, *9*(10), 1218–1220. https://doi.org/10.1038/nn1770

Henderson, L. M., Baseler, H. A., Clarke, P. J., Watson, S., & Snowling, M. J. (2011). The N400 effect in children: Relationships with comprehension, vocabulary and decoding. *Brain and Language*, *117*(2), 88–99. https://doi.org/10.1016/j.bandl.2010.12.003

Higashida, N. (2016). *The Reason I Jump: The Inner Voice of a Thirteen-Year-Old Boy with Autism* (Reprint edition; K. A. Yoshida & D. Mitchell, trans.). New York: Random House Trade Paperbacks.

Kasari, C., Brady, N., Lord, C., & Tager-Flusberg, H. (2013). Assessing the minimally verbal school-aged child with autism spectrum disorder. *Autism Research: Official Journal of the International Society for Autism Research*, *6*(6), 479–493. https://doi.org/10.1002/aur.1334

Kedar, I. (2012). *Ido in Autismland: Climbing Out of Autism's Silent Prison*. Sharon Kedar.

Kenny, L., Hattersley, C., Molins, B., Buckley, C., Povey, C., & Pellicano, E. (2016). Which terms should be used to describe autism? Perspectives from the UK autism community. *Autism: The International Journal of Research and Practice*, *20*(4), 442–462. https://doi.org/10.1177/1362361315588200

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psychtoolbox-3. *Perception*, *36*(14), 1–16.

Kutas, M. (1993). In the company of other words: Electrophysiological evidence for single-word and sentence context effects. *Language and Cognitive Processes*, *8*(4), 533–572. https://doi.org/10.1080/01690969308407587

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Lord, C., Rutter, M., Dilavore, P., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism Diagnostic Observation Schedule (2nd Ed.). Torrance, CA: Western Psychological Services*.

Lovaas, O. I. (1987). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology*, *55*(1), 3–9. https://doi.org/10.1037//0022-006x.55.1.3

Luck, S. J. (2014). *An introduction to the event-related potential technique*.

McCleery, J. P., Ceponiene, R., Burner, K. M., Townsend, J., Kinnear, M., & Schreibman, L. (2010). Neural correlates of verbal and nonverbal semantic integration in children with autism spectrum disorders. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *51*(3), 277–286. https://doi.org/10.1111/j.1469-7610.2009.02157.x

Mody, M., & Belliveau, J. W. (2013). Speech and Language Impairments in Autism: Insights from Behavior and Neuroimaging. *North American Journal of Medicine & Science*, *5*(3), 157–161.

Nigam, A., Hoffman, J. E., & Simons, R. F. (1992). N400 to Semantically Anomalous Pictures and Words. *J. Cognitive Neuroscience*, *4*(1), 15–22. https://doi.org/10.1162/jocn.1992.4.1.15

Ortiz-Mantilla, S., Cantiani, C., Shafer, V. L., & Benasich, A. A. (2019). Minimally-verbal children with autism show deficits in theta and gamma oscillations during processing of semantically-related visual information. *Scientific Reports*, *9*(1), 5072. https://doi.org/10.1038/s41598-019-41511-8

Paul, R., Chawarska, K., Cicchetti, D., & Volkmar, F. (2008). Language Outcomes of Toddlers With Autism Spectrum Disorders: A Two Year Follow-Up. *Autism Research : Official Journal of the International Society for Autism Research*, *1*(2), 97–107. https://doi.org/10.1002/aur.12

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.

Perrin, F., & García-Larrea, L. (2003). Modulation of the N400 potential during auditory phonological/semantic interaction. *Cognitive Brain Research*, *17*(1), 36–47. https://doi.org/10.1016/S0926-6410(03)00078-8

Petit, S., Badcock, N. A., Grootswagers, T., Rich, A. N., Brock, J., Nickels, L., … Woolgar, A. (2019). Towards an individualised neural assessment of receptive language in children. *bioRxiv*, 566752. https://doi.org/10.1101/566752

Petit, S., Badcock, N. A., Grootswagers, T., & Woolgar, A. (2019). Even neurotypical children are heterogeneous: using multivariate decoding to improve individual subject analysis of lexico-semantic EEG data. *bioRxiv*, 797175. https://doi.org/10.1101/797175

Posar, A., & Visconti, P. (2018). Sensory abnormalities in children with autism spectrum disorder. *Jornal De Pediatria*, *94*(4), 342–350. https://doi.org/10.1016/j.jped.2017.08.008

Rämä, P., Sirri, L., & Serres, J. (2013). Development of lexical–semantic language system: N400 priming effect for spoken words in 18- and 24-month old children. *Brain and Language*, *125*(1), 1–10. https://doi.org/10.1016/j.bandl.2013.01.009

Rapin, I., & Dunn, M. (2003). Update on the language disorders of individuals on the autistic spectrum. *Brain & Development*, *25*(3), 166–172.

Ribeiro, T. C., Valasek, C. A., Minati, L., & Boggio, P. S. (2013). Altered semantic integration in autism beyond language: a cross-modal event-related potentials study. [Miscellaneous Article]. *Neuroreport*, *24*(8), 414–418. https://doi.org/10.1097/WNR.0b013e328361315e

Rutter, M., Bailey, A., & Lord, C. (2003). Social communication questionnaire (SCQ). *Torrance, CA: WPS*.

Schlosser, R. W., Hemsley, B., Shane, H., Todd, J., Lang, R., Lilienfeld, S. O., … Odom, S. (2019). Rapid Prompting Method and Autism Spectrum Disorder: Systematic Review Exposes Lack of Evidence. *Review Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s40489-019-00175-w

Sinclair, J. (2013). Why I dislike “person first” language. *Autonomy, the Critical Journal of Interdisciplinary Autism Studies*, *1*(2).

Sparrow, S. S., Saulnier, C. A., Cicchetti, D. V., & Doll, E. A. (2016). *Vineland-3 : Vineland adaptive behavior scales*. /z-wcorg/.

Tager-Flusberg, H., & Kasari, C. (2013). Minimally Verbal School-Aged Children with Autism Spectrum Disorder: The Neglected End of the Spectrum. *Autism Research*, *6*(6), 468–478. https://doi.org/10.1002/aur.1329

Tager-Flusberg, H., Plesa Skwerer, D., Joseph, R. M., Brukilacchio, B., Decker, J., Eggleston, B., … Yoder, A. (2017). Conducting research with minimally verbal participants with autism spectrum disorder. *Autism: The International Journal of Research and Practice*, *21*(7), 852–861. https://doi.org/10.1177/1362361316654605

Tanaka, H., Watanabe, H., Maki, H., Sakriani, S., & Nakamura, S. (2019). Electroencephalogram-Based Single-Trial Detection of Language Expectation Violations in Listening to Speech. *Frontiers in Computational Neuroscience*, *13*. https://doi.org/10.3389/fncom.2019.00015

Thie, J. (2013). *A wireless marker system to enable evoked potential recordings using a wireless EEG system (EPOC) and a portable computer* (No. e32v1). Retrieved from PeerJ PrePrints website: https://peerj.com/preprints/32v1

Thornhill, D. E., & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components. *International Journal of Psychophysiology*, *83*(3), 382–392. https://doi.org/10.1016/j.ijpsycho.2011.12.007

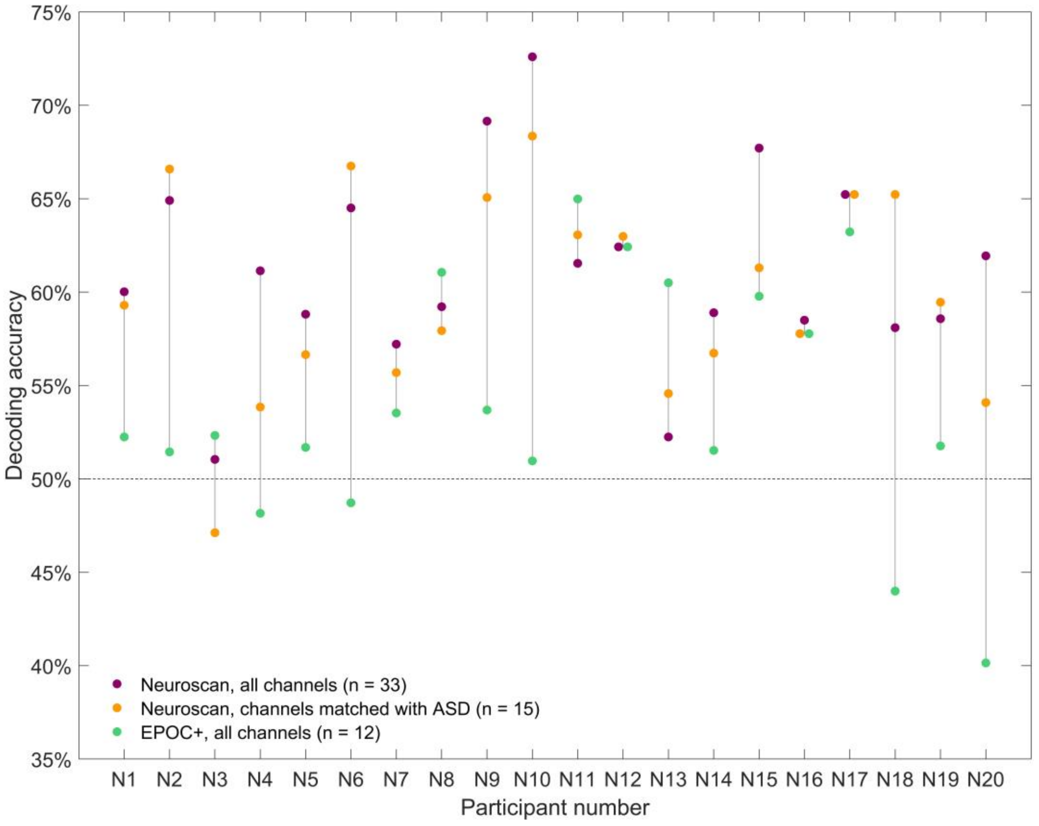
Tomchek, S. D., & Dunn, W. (2007). Sensory processing in children with and without autism: a comparative study using the short sensory profile. *The American Journal of Occupational Therapy: Official Publication of the American Occupational Therapy Association*, *61*(2), 190–200. https://doi.org/10.5014/ajot.61.2.190

Virués-Ortega, J. (2010). Applied behavior analytic intervention for autism in early childhood: meta-analysis, meta-regression and dose-response meta-analysis of multiple outcomes. *Clinical Psychology Review*, *30*(4), 387–399. https://doi.org/10.1016/j.cpr.2010.01.008

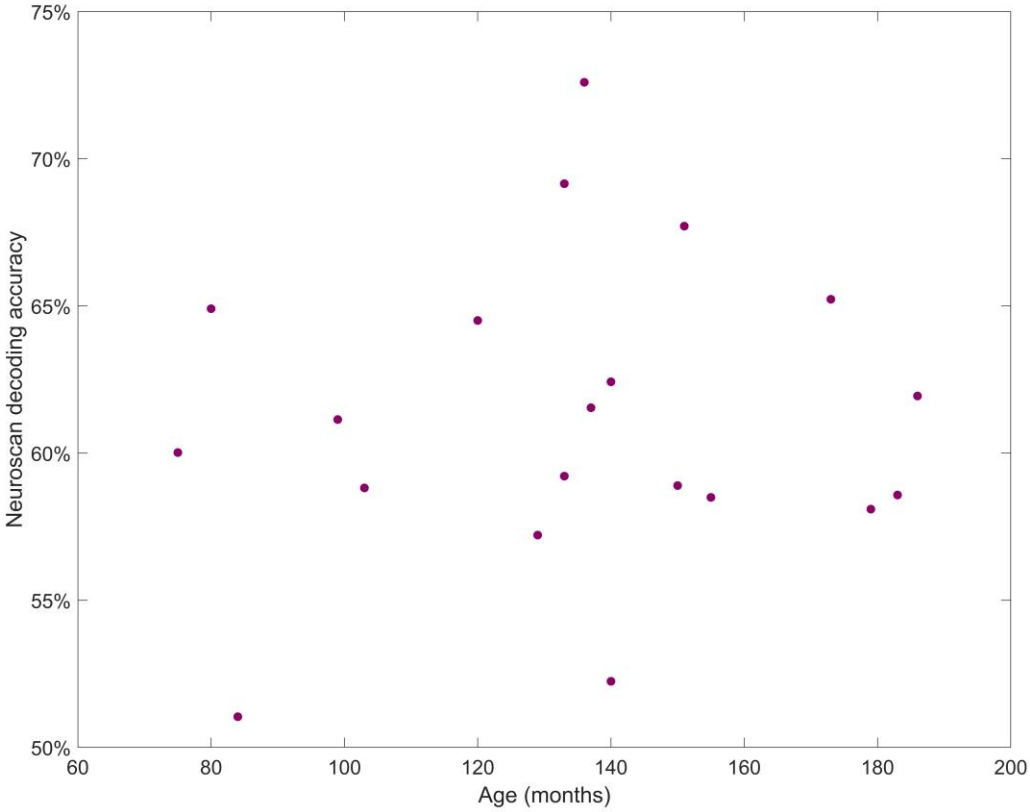
Wentura, D., & Frings, C. (2005). Repeated Masked Category Primes Interfere With Related Exemplars: New Evidence for Negative Semantic Priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 108–120. https://doi.org/10.1037/0278-7393.31.1.108

Woolgar, A., Golland, P., & Bode, S. (2014). Coping with confounds in multivoxel pattern analysis: What should we do about reaction time differences? A comment on Todd, Nystrom & Cohen 2013. *NeuroImage*, *98*, 506–512. https://doi.org/10.1016/j.neuroimage.2014.04.059

5.7 Supplementary materials



Supplementary Figure 1: Individual decoding accuracy for classification of identical target words in congruent and incongruent contexts, using all Neuroscan electrodes, or only the electrodes matching the ASD group, or all EPOC+ electrodes in neurotypical participants. Purple circles indicate Neuroscan decoding accuracy using all electrodes, orange circles indicate Neuroscan decoding accuracy using only the electrodes used in the ASD group, and green electrodes indicate EPOC+ decoding accuracy. There was no significant differences in accuracy for the Neuroscan when decoding on all, or only 15, electrodes. However, EPOC+ decoding accuracy was significantly lower than both Neuroscan decoding conditions.



Supplementary Figure 2: Neuroscan decoding accuracy across age for neurotypical participants. There was no significant correlation between age and decoding accuracy.

Chapter 6

Discussion

The aim of this thesis was to develop and evaluate a neural test of language comprehension suitable for use with minimally-verbal autistic children. The major challenge was to develop an approach that was sensitive at the individual subject level. In the first three experimental chapters I presented an evaluation of four new paradigms using electroencephalography (EEG) and functional transcranial Doppler ultrasound (fTCD) in neurotypical children. In the final experimental chapter I reported a small case-series demonstrating that the most sensitive paradigm can be used with minimally-verbal autistic children. In this final chapter I first discuss the development of paradigms to measure neural signals in individual children, including my findings of the heterogeneity in the neurotypical population. I reflect on the challenges associated with individual data, and the methods and analysis improvements that can be implemented to increase reliability. Next, I discuss the results obtained in the small sample of minimally-verbal autistic children. I then consider the limitations of the current research and offer future directions for this work, before concluding on the overall significance of this thesis.

6.1 Developing sensitive neural tests in neurotypical children

For minimally-verbal children, cognitive abilities may not be apparent through their behaviour and cannot be readily tested with standard assessments (Tager-Flusberg, 2000; Tager-Flusberg & Kasari, 2013). For this reason, paradigms investigating neural correlates of lexico-semantic processing may give insight into hidden language abilities in these children, in the absence of overt communication. I thus set out to develop electrophysiological paradigms that measure lexico-semantic processing. Two main findings were lacking from the literature on electrophysiological measures of language comprehension. First, the majority of typical electrophysiological research only reported group results, making it difficult to capture the heterogeneity in individual participants' data. Second, the few electrophysiological studies of receptive language processing in minimally-verbal autistic children (most of which were published since the start of my PhD) either did not include a comparison group or did not carry out detailed statistical analyses of the neurotypical children's data (Cantiani et al., 2016; Coderre et al., 2019; DiStefano, Senturk, & Spurling Jeste, 2019). Thus, the first goal of my thesis was to design experimental paradigms that index neural language processing in children and establish their sensitivity in the neurotypical population.

6.1.1 Designing sensitive EEG paradigms – the N400

In Chapters 2 and 3 I explored EEG paradigms using the N400 event-related potential (ERP) as a neural marker of lexico-semantic processing. The N400 has been widely studied in adults, children, and special populations (see the extensive review by Kutas & Federmeier, 2011). However, to my knowledge, only a few studies systematically report the individual detection rate of N400 effects, typically finding a medium detection rate around 50%, and all of this individual-subject work has been carried out with adults only (Beukema et al., 2016; Cruse et al., 2014; Rohaut et al., 2015). To address this gap in the literature I designed two auditory N400 paradigms contrasting words in congruent and incongruent contexts, and embedded them in child-friendly experiments (Chapter 2). Experiment 1 used normatively-associated versus unrelated word pairs, and Experiment 2 used congruent versus incongruent sentences. In groups of children aged 6-12 I found large univariate N400 effects (N = 15 and N = 16 respectively) with both paradigms when recording with a research-grade EEG system. However, at the individual level, there was large inter-individual variability in the strength, location, and timing of the N400 effect. Experiments 1 and 2 elicited significant univariate N400 effects in 47% and 56% of participants, respectively. In an attempt to increase this individual reliability, I built on Experiment 2 and created a third paradigm that contrasted congruent and incongruent auditory sentences together with short, cartoon-like animations that matched the semantic content of each sentence (Chapter 3). This paradigm was designed to increase participants' engagement to the task, helping them focus on the semantic content of each sentence. However, similar to Chapter 2, the individual detection rate was moderate at best, with 45% of participants showing a reliable univariate N400 effect. Overall, using a covert design, my results approximate those found previously in adults (Beukema et al., 2016; Cruse et al., 2014; Rohaut et al., 2015) with around half of participants (children or adults) showing statistically significant univariate effects. It is unclear why only some individuals show a statistical N400 effect but I discuss several potential hypotheses below.

6.1.1.1 Individual factors contributing to N400 heterogeneity

First, I asked whether inter-individual variability could be attributed to individual factors such as age or cognitive abilities. In Chapter 2 I assessed participants' verbal and non-verbal abilities using a vocabulary test (Peabody Picture Vocabulary Test; PPVT; Dunn & Dunn, 2007) and a matrix reasoning test (from the Kaufmann Brief Intelligence Test; Kaufman & Kaufman, 2004), respectively. Although our group was small for assessing correlations, I found no hint of relationships between these measures and the strength of N400 effects. This is consistent with several studies that did not find a relationship between vocabulary knowledge and the N400 in adults (Boudewyn, Long, & Swaab, 2012) or children (Henderson, Baseler, Clarke, Watson, & Snowling, 2011). However, other studies have demonstrated that related factors such as cognitive control (as measured by degree of inhibition of irrelevant cognitive processes), verbal working memory, and overall language proficiency (as measured by listening and reading comprehension of stories) correlate with the N400 (Boudewyn et al., 2012; Boudewyn, Long, & Swaab, 2013; Henderson et al., 2011; Landi & Perfetti, 2007; Perfetti, Wlotko, & Hart, 2005; Tanner, Goldshtein, & Weissman, 2018). In their study on cognitive control, Boudewyn et al. (2012) presented 50 adults with spoken congruent sentences in which the final word was associated or unassociated with a preceding word in the sentence (e.g. "In her haste she forgot to buy the *apples* and *oranges*/*bread*"). They found N400 effects for unrelated compared to related words. Furthermore, they found that better cognitive control, as measured by less interference on a Stroop task, predicted smaller N400 effects. This may indicate that individuals with better cognitive control focus less on the associative relations between words and more on the general meaning of the sentence. In the example above, participants with higher cognitive control may be less sensitive to the local lexical associations between *apples* and *bread* compared to their peers with low cognitive control, thus showing a smaller N400 effect. Elsewhere, some authors have suggested that verbal working memory is an important predictive factor of the N400 due to the taxing cognitive demands of integrating complex lexico-semantic meaning (Boudewyn, Long, & Swaab, 2013; Tanner, Goldshtein, & Weissman, 2018). Other authors disagree, instead suggesting that language proficiency and comprehension are the important predictors of the N400 (Henderson et al., 2011; Landi & Perfetti, 2007; Perfetti et al., 2005). In their review on individual differences in the neural markers of language, Tanner et al. (2018) conclude that it remains unclear which factors best predict inter-individual variability. My thesis was not designed to be a detailed study of individual differences in N400 and does not conclude this debate. Nevertheless, my findings do suggest that there is inter-individual variability in N400 amongst typically developing children and highlight the importance of larger studies addressing this issue. In future work it will be important to confirm whether this variability is due to individual-subject factors such as cognitive and language proficiency, as opposed to factors such as EEG data quality or participants' task engagement. One avenue would be to invite children for multiple testing sessions and examine the test-retest reliability of our paradigms (see Kiang, Patriciu, Roy, Christensen, & Zipursky, 2013).

6.1.1.2 The N400 across development

In addition to cognitive factors, there is evidence that the N400 changes with age. In an early developmental study, Holcomb, Coffey, and Neville (1992) analysed the ERPs of 130 participants aged 5 to 26 years in response to auditory and written congruent or incongruent sentences. While all the age groups showed an N400 effect, the younger participants (7 to 12 years) showed an N400 to both incongruent and congruent sentences. In contrast, the older participants (13 to 26 years) showed an N400 only for the incongruent condition. Holcomb et al. also found that the latency and amplitude of the N400 effect decreased from 5 to 15 years, before stabilizing, consistent with the view that neuronal connections keeps maturing until the mid-teen years (Huttenlocher, 1979; Huttenlocher, de Courten, Garey, & Van der Loos, 1982). However, one limitation of this study is that the authors did not control for the difficulty of stimuli across their age range. It is possible that the larger ERPs recorded in younger participants reflects increased difficulty or reduced familiarity with the linguistic stimuli used. In contrast to Holcomb et al., Byrne et al. (1999) controlled for age regarding stimulus difficulty by choosing age-appropriate words from the PPVT. They presented 56 children aged 5 to 12 with pictures and auditory labels that matched or mismatched the pictures. The youngest group's (5 to 6 years) N400 effect was largely distributed across the scalp, potentially reflecting a broad, less efficient lexico-semantic processing. In contrast, as age increased, the distribution of the N400 became less frontal and more central, more akin to adult-like N400.

More recently, four studies compared the N400 responses of adults and children (Atchley et al., 2006; Friederici & Hahne, 2001; Hahne, Eckstein, & Friederici, 2004; Juottonen, Revonsuo, & Lang, 1996). The first major finding from these studies is a delayed N400 in younger children compared to older children and adults (Hahne et al., 2004; Friederici & Hahne, 2001) or a delayed N400 in children compared to adults (Atchley et al., 2006; Juottonen et al., 1996). Three of these studies also found a larger N400 effect in children than in adults (Atchley et al., 2006; Friederici & Hahne, 2001; Juottonen et al., 1996). Overall, the few available studies of the N400 across development point to larger amplitude, longer latency, and more distributed scalp topography for children compared to adults. Despite these interesting findings, it is unclear why these developmental changes exist. It is possible that the larger amplitude reported in children corresponds to an increased reliance on contextual information (Holcomb et al., 1992). My studies were not designed as a test of developmental changes in N400 but the literature suggests that age differences might explain some of the variability between individuals in my data (age range: 6 to 12). However, there was no obvious association between age and presence of N400 effects in my (small) studies. Because it was not possible to extract clear peaks in individual ERPs, I did not conduct latency analyses. More data would be needed to assess developmental effects in my paradigms, such as a longitudinal assessment of the N400 in a cohort of children.

6.1.1.3 Task-related factors and neural processes contributing to heterogeneity

Several other factors may influence the presence and size of N400 effects. First, it is possible that I simply lacked sensitivity to detect small effects that were present in all individuals. This would be due to an insufficient number of trials contributing to participants’

ERP. Due to the type of statistical analyses that I used (temporal cluster correction for multiple comparisons), it was not possible to conduct power analyses that would indicate the number of trials needed to achieve sufficient power. All three studies from Chapters 2 and 3 used at least 94 trials per condition, which would give a power of .96 to detect an effect of size .5 in a within-subject, one-tailed t-test. This is similar or higher than what is typically used in children's N400 research (Atchley et al., 2006; Benau, Morris, & Couperus, 2011; Hahne et al., 2004; Henderson et al., 2011). Adding more trials may be beneficial in cases where effects are small but it is also necessary to balance the requirement for plentiful data with keeping the experimental session as short as possible, to maintain participants' focus.

It is also possible that only a subset of our participants engaged and/or performed the required task. Keeping in mind that I wanted to use the paradigm with a clinical population, I opted for a semi-covert design, asking participants to rate each sentence in their mind, then occasionally prompting to check for compliance. In Chapter 2 Experiment 1, I asked participants to rate the overall ability of friendly aliens to produce pairs of words that were associated. In Chapter 2 Experiment 2, I asked participants to count the number of incongruent sentences. Finally, in Chapter 3, I occasionally asked participants whether the last sentence they heard was congruent or incongruent. This allowed us to implement only minimal changes when testing autistic children (Chapter 5). However, Cruse et al. (2014) reported that an overt design where participants are prompted to give an answer on every trial was more successful than a covert or passive design at eliciting significant N400 effects in individual adults (75% detection rate for their overt task versus 58% detection rate for a covert task similar to ours). We are thus facing a trade-off between the task-demands we can impose on autistic participants and higher sensitivity for tasks that require more engagement. Interestingly, Experiment 2 had the best detection rate using multivariate decoding analyses, which may have reflected the counting task being the most engaging (see further discussion of this result below).

Another explanation for the observed inter-individual variability in the N400 is the possibility that different people may rely on different neural substrates for processing similar stimuli. Such theories have been put forward in studies of first and second language learning, which report individual differences in neural responses to linguistic stimuli. For example, Osterhout (1997) presented adult participants with correct or syntactically anomalous sentences (e.g. "The little girl read the sad story *cried*") and found that these anomalies elicited N400 effects in some participants but P600 effects in others. Osterhout interpreted these inter-individual differences as differences in the individual sensitivity to the semantic versus syntactic implications of the anomaly. Similar results have been found using anaphors (Nieuwland & Van Berkum, 2008), animacy violations (Nakano, Saron, & Swaab, 2010), and in second language learners (Osterhout, McLaughlin, Kim, Greenwald, & Inoue, 2004; Tanner, Mclaughlin, Herschensohn, & Osterhout, 2013). Thus, different individuals may process the same information in different ways.

It is also theoretically possible that similar cognitive processes manifest as different EEG signatures between individuals. For example, ERPs associated with lexico-semantic integration may have occurred at different timings and/or location depending on individual subject neural factors including processing speed and brain morphology. In addition, inter-individual differences in head shape and size may result in ERPs from similar neural generators manifesting in different sensors over individuals. Particularly considering that EEG has poor spatial resolution, due to the electrical signals being distorted and blurred from travelling through the skull and scalp before reaching the electrodes (Makeig, Bell, Jung, & Sejnowski, 1996; Srinivasan, Nunez, Tucker, Silberstein, & Cadusch, 1996), group average topology may be overly restrictive when guiding analysis at the individual subject level. Thus, in order to increase our sensitivity at the individual level and account for individual differences in topography, I also analysed data using multivariate pattern analyses (MVPA).

6.1.2 Increasing sensitivity using MVPA

6.1.2.1 Decoding lexico-semantic processing

MVPA uses machine-learning algorithms to extract patterns of neural activity that differ between two experimental conditions. I used two versions of MVPA in this thesis. In Chapter 2 I employed time-resolved MVPA by training a classifier to use any differences between semantic conditions from all EEG electrodes at each time point in turn. For Experiment 1, contrasting related and unrelated word pairs, the individual detection rate was similar for MVPA than univariate analyses (47% for univariate versus 53% for MVPA). However, for Experiment 2, contrasting congruent and incongruent sentences, the individual detection rate was numerically higher for MVPA than univariate analyses, with only two participants not showing significant decoding (56% for univariate versus 88% for MVPA). In Chapter 3 I went further in increasing the sensitivity of MVPA by allowing a classifier to train on the data from all the electrodes *and* all the time points. Again, this approach numerically improved our detection rate of neural responses to semantic anomalies (45% for univariate versus 65% for unconstrained MVPA; time-resolved MVPA yielded only a 30% detection rate).

This unconstrained MVPA approach has recently been used to investigate neural responses to semantic anomalies in adults. In their EEG study, Tanaka et al. (2019) could successfully decode the semantic conditions of sentences (correct or anomalous) in a group of adults. More interestingly, Geuze, Gerven, Farquhar, and Desain (2013) could decode the semantic category of word pairs (semantically related versus unrelated) at the individual level in all their adult participants. Unfortunately they do not report the individual univariate results, making it impossible to assess whether the signals they decoded matched onto the N400 effect. They obtained a higher detection rate than all of my studies, potentially because their participants were adults whose attention and engagement was higher and movement minimal, which in turn yielded better quality data.

A shortcoming of this unconstrained MVPA approach is that it is largely unspecific and does not allow us to examine the timing or location of the discriminative brain signals. I thus faced another trade-off between the highest sensitivity of MVPA and a reduced specificity regarding the neural signals contributing to decoding. For my purpose of assessing neural responses to lexico-semantic anomalies, it is possible to use MVPA to make statistical inference about the experimental conditions being processed differently and examine the time course and topography of the corresponding neural signals using univariate analyses. On its own, a positive MVPA result suggests that the context of the sentence must have been sufficiently processed to influence the way in which the identical auditory token is responded to. Even though I cannot conclude that complete sentence comprehension occurred (see Limitations section), I would not expect decoding to be successful if no understanding was present.

6.1.2.2 Maximum sensitivity obtained in Experiment 2

As stated above, I obtained the maximal sensitivity in the paradigm contrasting congruent and incongruent auditory sentences when analysing the Neuroscan data with MVPA. Here I discuss several possibilities that may explain the superior sensitivity of this paradigm associated with this EEG system and analysis method.

First, for this paradigm, Neuroscan data analysed with MVPA had a clear numerical advantage over EPOC+' (88% sensitivity for Neuroscan versus 25% for EPOC+). This may be caused by several factors. We first considered that the higher number of electrodes of the Neuroscan (33 versus 12 for EPOC+) and/or its superior electrode placement for N400 effects contributed to the superior classification. However, follow-up analyses revealed that reducing the Neuroscan's data to only the 12 electrodes adjacent to the EPOC+ electrodes did not reduce decoding accuracy. It is also possible that EPOC+' higher impedance led to noisier signal, which may have led to poorer decoding accuracy. In particular, we used minimal data processing before classification. In order to maintain a balanced design for decoding (Grootswagers, Wardle, & Carlson, 2017) we did not reject any epochs, even if they exceeded a voltage threshold indicating that they were noisy. It is thus possible that poorer classification for EPOC+ simply reflects noisier data.

A second interesting point is the superior sensitivity of the auditory-sentences paradigm over both the word-pair and the auditory-and-visual-sentences paradigms. This finding may be explained by the auditory stimuli used: the auditory sentences contained a silent gap before the final word, possibly enhancing the target-word expectation and/or providing a more appropriate baseline for EEG analysis. The task was also different: participants had to count the number of incorrect sentences, as opposed to rating the overall relatedness of words in experimental blocks, or covertly judging each stimulus. This may have required greater engagement, thus enhancing the neural response. Because this paradigm was more sensitive only when using MVPA, it is also possible that incrementing working memory, or other cognitive processes associated with counting the incorrect sentences, contributed to the different responses for the congruent and incongruent conditions. In this case, the classifier may use information that does not relate to semantic judgement *per se*. However, significant classification based on incrementing a number in one condition and not the other would still have to be based on recognising the difference between the two conditions which eventually indicates processing of the sentence-level meaning. In that case, the differences used by the classifier may not be reflected in the traditional N400 effect, thus explaining the superiority of MVPA over univariate analyses, but would still be suitable for my purpose of indexing receptive language processing.

6.1.3 Using methods suited to autistic children

When designing my paradigms, I tried to make them applicable for use with minimally-verbal autistic children. Two major difficulties in testing this populations are typically reported. The first one is difficulty dealing with task demands, which could be attributed to one or more of the following: lack of motivation, failure to understand the task instructions, inability to establish rapport with the examiner, anxiety, or frustration (Tager-Flusberg, 2000). I embedded the paradigms in engaging stories: in the EEG paradigms (Chapters 2 and 3) children were asked to help capture evil aliens that were interfering with messages sent to astronauts or to rate friendly aliens who were trying to learn English. In the fTCD paradigm (Chapter 4) children were helping pirates to collect treasures and return home after being caught in a storm. These stories were designed to keep neurotypical children engaged and interested in the task but I acknowledge that it is possible that autistic children may find these additional instructions confusing or socially demanding. To address this, I designed my tasks such that focusing on the story was not mandatory to complete the task and, in fact, the only relevant task was to attend to the auditory stimuli. When testing autistic children, I reiterated that the important thing was to focus on listening and understanding the sentences.

The second difficulty reported when testing autistic children arises from sensory differences, which can manifest as hyposensitivity, hypersensitivity, or sensory seeking behaviours and which can occur within all sensory modalities (Posar & Visconti, 2018; Tomchek & Dunn, 2007). In addition, autistic children may have difficulties remaining still during testing, making methods such as functional magnetic resonance imaging or magnetoencephalography unusable in some cases. To mitigate these sensory and motor issues, I chose two methods that are relatively easy to set up and/or less sensitive to movement. In particular, in Chapter 2 I evaluated the signals recorded by a consumer-grade EEG system, Emotiv's EPOC+, as an easy-to-setup alternative to research-grade systems. Furthermore, in Chapter 4 I explored another avenue using a method relatively insensitive to movement, fTCD.

6.1.3.1 Adapted EEG systems

Traditional EEG research involves long setups in electrically-shielded rooms with expensive materials and preparation of participants' skin before applying the electrodes. All these factors usually preclude the use of EEG outside the lab. With increasing demands to make EEG technology more accessible, recent years have seen the development of many portable, inexpensive, and easy-to-setup EEG systems. Several studies validating and comparing these systems to traditional EEG setups have been conducted (Badcock et al., 2013, 2015; Callan, Durantin, & Terzibas, 2015; de Lissa, Sörensen, Badcock, Thie, & McArthur, 2015; Debener, Minow, Emkes, Gandras, & De Vos, 2012; Duvinage et al., 2013; Forney et al., 2013; Grummett et al., 2015; Lin, Wang, & Jung, 2014; Ries, Touryan, Vettel, McDowell, & Hairston, 2014; Rogers, Johnstone, Aminov, Donnelly, & Wilson, 2016).

From these studies, it appears that most of the emerging portable EEG devices may record signals of similar or slightly worse quality than research-grade systems. Portable systems are usually noisier, likely due to participants movements and/or poorer contact with the scalp. Regarding the EPOC system (the predecessor of EPOC+), several studies have found good quality of recordings for ERP such as auditory and mismatch negativity (Badcock et al., 2013, 2015), the face-sensitive N170 component (de Lissa et al., 2015), and for possible brain-computer interface applications using flashing visual stimuli (Debener et al., 2012; Lin et al., 2014; Zander et al., 2011). To my knowledge, no existing studies have investigated the capacity of the EPOC or EPOC+ to record the N400 component. In Chapter 2, I validated the EPOC+ against our research-grade Neuroscan system. In contrast to most comparisons studies to date, but in line with Badcock et al. (2013, 2015) and de Lissa et al. (2015), I recorded EEG from the two systems simultaneously, allowing the direct comparison of signal quality.

I found that EPOC+ recorded ERPs of similar shape and amplitude to Neuroscan at our key electrodes of interest (FC3 and FC4), with medium correlations between the waveforms (mean Spearman's ρ across conditions and experiments = .59). However, I also found that EPOC+ was generally noisier than Neuroscan, as indicated by a higher number of noisy trials that were rejected during data pre-processing. It is important to note that across my studies, and in line with previous literature, I found the N400 effect to be largest at central locations. Yet, EPOC+ does not have electrodes at these locations and is therefore not be ideal for this particular ERP. Furthermore, I found significantly poorer performance of EPOC+ with MVPA, compared to Neuroscan. In Experiment 1 of Chapter 2, I could successfully decode semantic condition from the Neuroscan's data both at the group level and in 53% of individuals. In contrast, I could only decode from EPOC+'s data in 13% of individuals and not at the group level. For Experiment 2, I could decode from Neuroscan and EPOC+'s data at the group level and in 88% of individuals using Neuroscan or 25% of individuals using EPOC+. This suggests that, in its current form, EPOC+ is not sensitive for MVPA analyses of semantic processing in individuals. This is in line with some previous studies that found lower decoding accuracy for EPOC compared to research-grade systems (Duvinage et al., 2013; Ries et al., 2014), although other studies have successfully decoded data using EPOC, albeit using more robust and early ERP components (e.g. steady-state visual response, Lin et al., 2014; P300, Debener et al., 2012). Interestingly, Debener et al. modified the EPOC by removing electrodes from the headset and fixing them to an elastic cap. This allowed a better fit to the scalp and also the freedom to re-arrange the electrodes at any locations (including central locations), which may be useful for future N400 work. A more recent EEG device from Emotiv, EPOC Flex, was developed with movable electrodes, allowing the recording over central locations, and may be an interesting option in the future.

Overall, I did not find significant differences between the EPOC+ and research-grade EEG in univariate analyses but it did perform worse than Neuroscan for MVPA. Thus, to balance the apparent sensitivity and accessibility of the two technologies when testing autistic children, I first set out to record from Neuroscan and if the child could not tolerate it, I tried to record from EPOC+.

6.1.3.2 Functional Transcranial Doppler ultrasound (fTCD)

Following my aim of designing paradigms that can be applied to test children with sensory and behavioural difficulties, I explored another avenue using a portable technology, fTCD. In Chapter 4, I presented a novel paradigm using controlled stimuli to elicit task-following in neurotypical children. I recorded the blood-flow velocity to the two hemispheres in response to two task instructions (generating words versus remembering the location of letters on the screen). I found that the language task was more left-lateralised, while the visuospatial task was bilateral. By comparing the left-minus-right blood-flow velocity during the two tasks, I could observe task following in the group of participants and in half of the individual participants. An important advantage of observing task-following, as opposed to EEG responses to sentences, is that it allows for strong inference of language comprehension. For instance, a different EEG response to semantically congruent and incongruent sentences suggests lexico-semantic processing of language but it may also reflect different processing, such as recognizing statistical regularities in speech (see Limitations section below). Intact task-following however, requires understanding the task instructions, thus providing stronger inference for language comprehension.

FTCD allows the continuous measurement of the haemodynamic response, thus giving indirect measurement of neural activity. It presents three clear advantages over other haemodynamic measurement methods, such as functional magnetic resonance imaging (fMRI). First, it is relatively insensitive to movement, more so than most neuroimaging methods (Lohmann, Ringelstein, & Knecht, 2006). Second, it is portable and easy to set up, allowing its potential use outside of the laboratory. This has also allowed the recording of neural responses in populations for whom lying still in fMRI scanners may not be possible, such as children (Bishop, Watt, & Papadatou-Pastou, 2009; Haag et al., 2010). Third, it is much less expensive than fMRI, providing an inexpensive alternative to fMRI scanning in some cases (Lohmann et al., 2006).

In my case, I used fTCD as a way to measure covert task-following in children, taking advantage of its insensitivity to movement. Compared to previous work, my approach had two elements that I introduced to improve scientific rigor. First, my paradigm had the clear advantage of using the same stimuli in both tasks, alleviating any possible confound of stimuli-related activity and task-related activity. Second, I analysed the lateralization by averaging velocity data within a pre-defined period of interest, instead of analysing the peak in the data. This allowed an unbiased analysis of lateralisation. However, using my unbiased design and analysis, I found lateralisation effects to be somewhat weaker than anticipated based on previous literature (Bishop et al., 2009; Groen, Whitehouse, Badcock, & Bishop, 2011, 2012; Haag et al., 2010) and could only find statistical evidence of task-following in around half of the neurotypical children. Thus, despite designing child-friendly paradigms that aimed to keep children engaged to the tasks, I failed to improve the detection rate compared to my most successful EEG paradigm.

6.2 Neural signals of minimally-verbal autistic children

As I have discussed in this thesis, for children with minimal spoken language, it is not trivial to assess true cognitive abilities. Anecdotal evidence from non-speaking people who use augmentative and alternative communication such as rapid prompting method sometimes receive passionate criticism due to the lack of scientific evidence to support their claims (Schlosser et al., 2019). One of the motivations for my work was to supplement the subjectivity of these reports with objective assessments of language that do not rely on an individual's motor and speech behaviour. In neurotypical children, the most reliable paradigm was the one that used EEG to measure neural responses to semantic anomalies in spoken sentences (Chapter 2, Experiment 2). Thus, the next logical step of this work was to assess whether this test could be applied to minimally-verbal children.

To this end, I presented three minimally-verbal children, aged 11:5, 15:4, and 5:10, with my sentence paradigm. Despite limited spoken language, I found evidence of lexico-semantic processing from the neural signals of the oldest child, who was also the most verbal. This was confirmed by both significant multivariate decoding and by a statistically significant N400 effect in a pre-defined region of interest, providing strong evidence for intact lexico-semantic processing in this child. I also found significant multivariate decoding in the youngest child. However, this child's ERPs showed a reverse N400-like pattern, with congruent sentences eliciting a more negative deflection than incongruent sentences. This pattern is difficult to interpret and may reflect altered but present lexico-semantic processing. I did not find equivalent evidence in the remaining autistic participant, reflecting either absent or altered lexico-semantic processing, or a lack of sensitivity of my methods; unfortunately, no conclusions can be drawn from these null results.

Some heterogeneity in this very small autistic sample is not surprising, given the heterogeneity in our data for neurotypical children and the well-known heterogeneity for cognitive profiles in autism (Jones & Klin, 2009). Moreover, the few previous studies including minimally-verbal children in EEG research have similarly found contrasting results. Using a picture-word paradigm, Cantiani et al. (2016) found ERP evidence for lexico-semantic processing in half of their minimally-verbal participants and in 90% of their neurotypical participants. In addition, DiStefano et al. (2019) found large inter-individual differences in neural signals to matched and mismatched picture-words, with some minimally-verbal children showing a typical N400 effect and others showing reversed N400-like effect. Finally a recent study on autistic children with severe features (characterized in the study as 'Level-3 autism') found no ERP or eye-tracking evidence for lexico-semantic comprehension in the two children with minimal spoken language (Coderre et al., 2019).

Overall, from my results and previous literature, it appears that inter-individual variability is large both in the neurotypical and the autistic population. This emphasises the need for assessment at the individual subject level. My paradigm adds to the scarce literature by providing the first report of neural signals in a purely auditory paradigm. This design allowed me to alleviate concerns about autistic children having difficulties with visual attention or integrating information across modalities, by requiring them to only focus on the auditory sentences. My data provide an initial proof-of-concept that my paradigm can be used to indicate lexico-semantic processing in some autistic children. Establishing the presence of such neural signals in at least some minimally-verbal children could in turn influence their care, treatment, and the methods used to communicate with them.

6.3 Limitations and future directions

The first, important, limitation on interpretation of this work concerns what level of linguistic processing is reflected by the N400 component – an issue that is currently unresolved in the literature. The N400 effect is commonly accepted to reflect the continuous processing of lexico-semantic properties of stimuli. Some authors have argued that the N400 arises after the incoming word has been recognised and reflects the integration of the incoming stimuli with the semantic context encoded and stored in working memory (Brown & Hagoort, 1993; Hagoort, Baggio, & Wlllems, 2009). This "post-lexical" view is consistent with the fact that the N400 is affected by high-level linguistic manipulations, such as in response to different pragmatic situations (Hunt, Politzer-Ahles, Gibson, Minai, & Fiorentino, 2013). However, this view is not supported by evidence finding N400 in response to pseudowords (Friedrich, Eulitz, & Lahiri, 2006). Pseudowords do not have a mental representation in the lexicon and therefore should not have any meaning. As such, the N400 arising for these stimuli cannot reflect their integration with a previous semantic context. Based on this evidence, some authors have instead argued that the N400 reflects a *pre-lexical* process – i.e. prior to word recognition itself (Deacon, Dynowska, Ritter, & Grose-Fifer, 2004). However, this view does not explain discourse-level N400 effects or evidence for N400 being modulated by the semantic context (Szewczyk & Schriefers, 2018). In my work, I carefully designed my experiments so that the same target words and contexts would appear in both semantic conditions (congruent and incongruent). As such, I postulate that any reliable difference in the neural activity between conditions must reflect differential processing of the target word given its semantic context, which in turn must reflect differential integration of the word and context when they do and do not match.

The second question that directly applies to my results pertains to the level of *semantic* processing that we are accessing. A difference in neural signals between the congruent and incongruent condition suggests that different neural processes were engaged. However, it remains possible that this difference reflects statistical learning of word co-occurrences. For example, in the sentences: “There were candles on the birthday *cake*” and “There were candles on the birthday *door*”, it is possible that, without understanding the meaning of the full sentence, a larger N400 for the incongruent condition reflects the lower co-occurrence of “birthday” and “door” in the same sentence. Many studies have shown that statistical learning of the occurrence of spoken words and visual scenes is a key learning process for infants (see review by Romberg & Saffran, 2010). In addition, previous EEG studies found N400 responses to statistical learning in adults (Abla, Katahira, & Okanoya, 2008; Cunillera, Toro, Sebastián-Gallés, & Rodríguez-Fornells, 2006; Sanders, Newport, & Neville, 2002). However, in all of these studies, the statistical learning occurred during the course of the experiment. In my case, a possible influence of the statistical regularities of the environment would reflect lifelong exposure to language – an effect that was difficult to assess in my studies. Previous EEG studies on minimally-verbal autistic children have all used picture-word paradigms (Cantiani et al., 2016; Coderre et al., 2019; DiStefano et al., 2019), which may be more sensitive to the statistical learning phenomenon (infants learn language by associating spoken words with visual objects). My study contrasting congruent and incongruent sentences could be less sensitive to this phenomenon, as sentence-level comprehension is a higher-level process requiring continuous integration and encoding of the words in working memory but would not be completely devoid of it.

Third, even though sentence stimuli may be less sensitive to statistical learning of regularities, they pose a different problem. Spoken language comprehension is typically seen as a hierarchical process, ranging from speech perception to lexical recognition to semantic integration. This latter process itself relies on both the stored semantic memory of the incoming word and the encoding and retrieval of the previous context from working memory (Caplan &

Waters, 1999; Kutas & Federmeier, 2000). In particular, studies with adults show that difficulties with working memory lead to impaired language comprehension when the stimuli are complex sentences (King & Just, 1991; Miyake, Carpenter, & Just, 1994). As such, failure to detect neural responses to semantic anomalies during my sentences paradigm may reflect impaired higher-level semantic comprehension but could also relate to deficits at any lower level in the processing chain. In particular, previous research suggests that several processing steps may be atypical in minimally-verbal autism. Some authors have noted difficulties in working memory of verbal autistic adults and children (Minshew, Luna, & Sweeney, 1999; Russell, Jarrold, & Henry, 1996; Steele, Minshew, Luna, & Sweeney, 2007). However, it is unclear how these results translate to minimally-verbal children. In addition, it is widely accepted that autistic people have atypical basic auditory processing. A review by O'Connor (2012) reveals various levels of difficulties in the processing of prosody and speech in noise (namely basic perceptual atypicalities, such as pitch and loudness). Such difficulties have also been found for minimally-verbal individuals (Yau, McArthur, Badcock, & Brock, 2015; Cantiani et al., 2016). For these reasons, a null result in my current experiment is particularly difficult to interpret, as we are not able to distinguish at which stage the atypicalities arise. However, a positive neural result in a child reflects intact lexico-semantic processing of language, as well as intact processing along all the previous processing steps. Future research may use incremental paradigms, that allows one to first check whether basic sensory processing is affected, then how speech is perceived compared to non-speech, then build up to sentence-level comprehension. For example, an interesting avenue to explore incrementing processing of auditory information was introduced by Ghosh Hajra et al. (2018). In their study, they presented interlaced frequent and deviant tones, and congruent and incongruent word pairs in a five minutes stream. This allowed them to investigate basic auditory responses to tones, attention (through the oddball effect), and semantic processing. Even though they do not report the individual waveforms, they obtained above-chance decoding accuracy of the semantic condition in individuals, making such fast paradigms a promising option for assessing special populations.

Another limitation of the current work derives from the task demands associated with counting the incorrect sentences in my sentence paradigm. There are still inconsistencies regarding whether overt decisions about the stimuli are needed to elicit N400 effects. On the one hand, studies have shown that as long as the stimuli are attended, overt responses are not necessary to elicit N400 effects. For example, Bentin, Kutas, and Hillyard (1995) designed a dichotic listening task with concurrent presentation of different words to the two ears of participants. Participants were instructed to attend to one ear and ignore the other. Even though no overt decision was required from the participants, the authors observed N400 effects in response to semantic manipulations for the attended but not the unattended stimuli. In a similar study in the visual domain, McCarthy and Nobre (1993) found the same results, with attended written words eliciting N400 effects while unattended words did not. Finally, Connolly, Stewart, and Phillips (1990) compared the amplitude of N400 effects following sentences with incongruent and congruent endings, with participants engaging in different levels of processing. They found no differences in the N400 effects whether participants were asked to simply listen to sentences or make orthographic, phonological, or semantic decisions about the last word. From these studies, it thus seems that attention to the stimuli is critical to elicit N400 effects, while the actual task is less important. However, all these studies were conducted with adults, for whom staying attentive throughout an experimental session may be easier than for children. These studies also do not report on the individual N400 effects.

In their N400 study on adults, Cruse et al. (2014) found that asking participants to press a button indicating whether they just heard a semantically related or unrelated word pairs was the most successful design to elicit N400 effects in individuals (75% reliability). In contrast, asking participants to make this judgement *silently* decreased the reliability to 58%. Finally, when they instructed their participants to simply listen to the sentences, without prompting them to judge their semantic content, their detection rate dropped to 0%. However, in line with the previous findings mentioned above, N400 effects were present at the group level for all conditions. It seems that as long as participants are attending to the stimuli, N400 effects can be detected at the group level. However, it is unclear whether this is also the case in children. Furthermore, this does not apply to individual ERPs, where task-demands play a crucial role in the presence of N400 effects. Ideally, a paradigm tailored to minimally-verbal autistic children would be completely passive, as we cannot ensure that they will complete the required task, but this approach risks minimizing our neuroimaging sensitivity. In my case, I opted for a middle ground: to assume competence in our participants, give them the cover story but emphasise the need to silently judge the semantic content of each sentence (covert design), and repeatedly remind them to pay attention to all the sentences. The slightly different instructions given to neurotypical and autistic participants mean that it is possible that the neurotypical and autistic groups performed slightly different processes. However, this does not diminish any positive evidence that I find in response to the semantic context because differential processing of the identical auditory token in the absence of the additional task (i.e. counting) is still evidence for differential integration of that word with its context.

Finally, the implications of Chapter 5 are clearly limited by the small sample size of minimally-verbal autistic children. This chapter constitutes a proof-of-concept that my methods can be used with minimally verbal autistic children and that evidence of lexico-semantic processing was present in at least one child. However, a much larger sample size would be beneficial for several reasons. First, it would allow us to estimate how the minimally-verbal autistic group compares to the neurotypical group in terms of neural signals. It would be interesting, for example, to observe the amplitude and latency of the N400 effect at the group level (though with all the caveats raised here and elsewhere concerning group-level analysis). Second a larger group would allow us to estimate prevalence: how common intact language processing is in the minimally-verbal population. This could have important implications for how minimally-verbal children are cared for and treated. Finally, a larger sample size may allow us to investigate potential factors that contribute to the presence or absence of lexico-semantic processing between individuals. For example, it is possible that the presence of neural signals reflecting lexico-semantic integration will correlate with communication abilities, age, or autism severity, or have predictive power for future language development.

I also have to note that I attempted to test five autistic participants and two were unable to tolerate either EEG system, despite previous familiarisation/desensitisation. This indicates that even though these methods can be successfully applied to some autistic children, methodological improvements are still needed to make this technology accessible to the entire minimally-verbal population. With the emergence of new EEG devices, solutions balancing comfort and signal quality may provide promising avenues for extending the current work to more children.

In the future, it may be beneficial to combine approaches to better evaluate language abilities of autistic children. Kasari et al. (2013), and Tager-Flusberg and Kasari (2013) encourage a multi-disciplinary approach that takes into account parent reports, child observation, and objective measures to fully characterise the profile of these children. Another interesting avenue would be to assess the development of the N400 in individual autistic children, as they age and/or become verbal. Little is known about this transition from pre-verbal to verbal and the N400 may potentially be used as a biomarker to distinguish children who will acquire language from those who will not. Previous studies have already attempted to use the N400 as a biomarker in the context of dementia (Olichney et al., 2008), where reduced N400 effects in adults with mild cognitive impairment indicated likely progression to dementia. It may also be interesting to combine my EEG measures with eye-tracking. Eye-tracking technology has previously allowed to assess vocabulary knowledge in infants and autistic individuals (Brady, Anderson, Hahn, Obermeier, & Kapa, 2014; Gredebäck, Johnson, & von Hofsten, 2009). I decided not to use eye tracking in this work because of the visual attention deficits reported in autistic people (Behrmann, Thomas, & Humphreys, 2006; Ortega, López, & Aboitiz, 2008). Nonetheless, it may be interesting to present visual pictures and their auditory labels, versus mismatching labels. We could then combine the measurement of the neural responses with tracking participants' gaze, so as to have multiple avenues to assess language comprehension.

A final consideration regards to the ethical implications of my results. We must take great care in divulging individual-subject results to families. This is particularly important in case where we do not find neural evidence for lexico-semantic processing, as null results are not interpretable, so it is important to correctly inform families that no conclusions can be drawn from my test. Positive results must also be divulged with care as this information may have important repercussions, and at the statistical thresholds typically used, will reflect a false positive up to 5% of the time. As more objective evidence emerges, it is likely that we will see a shift in the care of minimally-verbal individuals. It will then be crucial for researchers to work in close collaborations with families and therapists to ensure that this new information benefits autistic individuals.

6.4 Conclusion

For autistic children with minimal spoken language, objective tests of language comprehension are urgently needed. Reliable assessment of their language abilities would be transformative for their care and communication opportunities, yet behavioural tests are notoriously unreliable in this population. Over five neuroimaging studies, this thesis documents the development of neural approaches to meet this challenge and provides a proof-of-concept that electrophysiology may be suitable to uncover intact lexico-semantic processing in some minimally-verbal autistic children. It provides insight into methodological prospects and limitations of neural tests of language processing, as well as theoretical insight into individual heterogeneity in the neurotypical and autistic population. Perhaps more importantly, it constitutes a step towards decoding hidden language abilities in those who cannot speak.

6.5 References

Abla, D., Katahira, K., & Okanoya, K. (2008). On-line Assessment of Statistical Learning by Event-related Potentials. *Journal of Cognitive Neuroscience*, *20*(6), 952–964. https://doi.org/10.1162/jocn.2008.20058

Atchley, R. A., Rice, M. L., Betz, S. K., Kwasny, K. M., Sereno, J. A., & Jongman, A. (2006). A comparison of semantic and syntactic event related potentials generated by children and adults. *Brain and Language*, *99*(3), 236–246. https://doi.org/10.1016/j.bandl.2005.08.005

Badcock, N. A., Mousikou, P., Mahajan, Y., de Lissa, P., Thie, J., & McArthur, G. (2013). Validation of the Emotiv EPOC® EEG gaming system for measuring research quality auditory ERPs. *PeerJ*, *1*, e38. https://doi.org/10.7717/peerj.38

Badcock, N. A., Preece, K. A., Wit, B., Glenn, K., Fieder, N., Thie, J., & McArthur, G. (2015). Validation of the Emotiv EPOC EEG system for research quality auditory event-related potentials in children. *PeerJ*, *3*, e907. https://doi.org/10.7717/peerj.907

Behrmann, M., Thomas, C., & Humphreys, K. (2006). Seeing it differently: Visual processing in autism. *Trends in Cognitive Sciences*, *10*(6), 258–264. https://doi.org/10.1016/j.tics.2006.05.001

Benau, E. M., Morris, J., & Couperus, J. W. (2011). Semantic Processing in Children and Adults: Incongruity and the N400. *Journal of Psycholinguistic Research*, *40*(3), 225–239. https://doi.org/10.1007/s10936-011-9167-1

Bentin, S., Kutas, M., & Hillyard, S. A. (1995). Semantic processing and memory for attended and unattended words in dichotic listening: Behavioral and electrophysiological evidence. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(1), 54–67. https://doi.org/10.1037/0096-1523.21.1.54

Beukema, S., Gonzalez-Lara, L. E., Finoia, P., Kamau, E., Allanson, J., Chennu, S., … Cruse, D. (2016). A hierarchy of event-related potential markers of auditory processing in disorders of consciousness. *NeuroImage : Clinical*, *12*, 359–371. https://doi.org/10.1016/j.nicl.2016.08.003

Bishop, D. V. M., Watt, H., & Papadatou-Pastou, M. (2009). An efficient and reliable method for measuring cerebral lateralization during speech with functional transcranial Doppler ultrasound. *Neuropsychologia*, *47*(2), 587–590. https://doi.org/10.1016/j.neuropsychologia.2008.09.013

Boudewyn, Megan A., Long, D. L., & Swaab, T. Y. (2012). Cognitive control influences the use of meaning relations during spoken sentence comprehension. *Neuropsychologia*, *50*(11), 2659–2668. https://doi.org/10.1016/j.neuropsychologia.2012.07.019

Boudewyn, Megan Ann, Long, D. L., & Swaab, T. Y. (2013). Effects of Working Memory Span on Processing of Lexical Associations and Congruence in Spoken Discourse. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00060

Brady, N. C., Anderson, C. J., Hahn, L. J., Obermeier, S. M., & Kapa, L. L. (2014). Eye tracking as a measure of receptive vocabulary in children with autism spectrum disorders. *Augmentative and Alternative Communication*, *30*(2), 147–159.

Brown, C., & Hagoort, P. (1993). The Processing Nature of the N400: Evidence from Masked Priming. *Journal of Cognitive Neuroscience*, *5*(1), 34–44. https://doi.org/10.1162/jocn.1993.5.1.34

Byrne, J. M., Connolly, J. F., MacLean, S. E., Dooley, J. M., Gordon, K. E., & Beattie, T. L. (1999). Brain activity and language assessment using event-related potentials: Development of a clinical protocol. *Developmental Medicine and Child Neurology*, *41*(11), 740–747. https://doi.org/10.1017/s0012162299001504

Callan, D. E., Durantin, G., & Terzibas, C. (2015). Classification of single-trial auditory events using dry-wireless EEG during real and motion simulated flight. *Frontiers in Systems Neuroscience*, *9*, 11.

Cantiani, C., Choudhury, N. A., Yu, Y. H., Shafer, V. L., Schwartz, R. G., & Benasich, A. A. (2016). From Sensory Perception to Lexical-Semantic Processing: An ERP Study in Non-Verbal Children with Autism. *PLOS ONE*, *11*(8), e0161637. https://doi.org/10.1371/journal.pone.0161637

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, *22*(1), 77–94. https://doi.org/10.1017/S0140525X99001788

Coderre, E. L., Chernenok, M., O'Grady, J., Bosley, L., Gordon, B., & Ledoux, K. (2019). Implicit Measures of Receptive Vocabulary Knowledge in Individuals With Level 3 Autism. *Cognitive and Behavioral Neurology*, *32*(2), 95. https://doi.org/10.1097/WNN.0000000000000194

Connolly, J. F., Stewart, S. H., & Phillips, N. A. (1990). The effects of processing requirements on neurophysiological responses to spoken sentences. *Brain and Language*, *39*(2), 302–318. https://doi.org/10.1016/0093-934X(90)90016-A

Cruse, D., Beukema, S., Chennu, S., Malins, J. G., Owen, A. M., & McRae, K. (2014). The reliability of the N400 in single subjects: Implications for patients with disorders of consciousness. *NeuroImage : Clinical*, *4*, 788–799. https://doi.org/10.1016/j.nicl.2014.05.001

Cunillera, T., Toro, J. M., Sebastián-Gallés, N., & Rodríguez-Fornells, A. (2006). The effects of stress and statistical cues on continuous speech segmentation: An event-related brain potential study. *Brain Research*, *1123*(1), 168–178. https://doi.org/10.1016/j.brainres.2006.09.046

de Lissa, P., Sörensen, S., Badcock, N., Thie, J., & McArthur, G. (2015). Measuring the face-sensitive N170 with a gaming EEG system: A validation study. *Journal of Neuroscience Methods*, *253*, 47–54. https://doi.org/10.1016/j.jneumeth.2015.05.025

Deacon, D., Dynowska, A., Ritter, W., & Grose-Fifer, J. (2004). Repetition and semantic priming of nonwords: Implications for theories of N400 and word recognition. *Psychophysiology*, *41*(1), 60–74. https://doi.org/10.1111/1469-8986.00120

Debener, S., Minow, F., Emkes, R., Gandras, K., & De Vos, M. (2012). How about taking a low-cost, small, and wireless EEG for a walk? *Psychophysiology*, *49*(11), 1617–1621.

DiStefano, C., Senturk, D., & Spurling Jeste, S. (2019). ERP Evidence of Semantic Processing in Children with ASD. *Developmental Cognitive Neuroscience*, 100640. https://doi.org/10.1016/j.dcn.2019.100640

Dunn, L. M., & Dunn, D. M. (2007). PPVT-4: Peabody picture vocabulary test. *Pearson Assessments*.

Duvinage, M., Castermans, T., Petieau, M., Hoellinger, T., Cheron, G., & Dutoit, T. (2013). Performance of the Emotiv Epoc headset for P300-based applications. *BioMedical Engineering OnLine*, *12*, 56. https://doi.org/10.1186/1475-925X-12-56

Forney, E., Anderson, C., Davies, P., Gavin, W., Taylor, B., & Roll, M. (2013). A comparison of EEG systems for use in P300 spellers by users with motor impairments in real-world environments. *Proceedings of the Fifth International Brain-Computer Interface Meeting: Defining the Future*.

Friederici, A., & Hahne, A. (2001). *Development Patterns of Brain Activity Reflecting Semantic and Syntactic Processes*. https://doi.org/10.1075/lald.24.11fri

Friedrich, C. K., Eulitz, C., & Lahiri, A. (2006). Not every pseudoword disrupts word recognition: An ERP study. *Behavioral and Brain Functions*, *2*, 36. https://doi.org/10.1186/1744-9081-2-36

Geuze, J., Gerven, M. A. J. van, Farquhar, J., & Desain, P. (2013). Detecting Semantic Priming at the Single-Trial Level. *PLOS ONE*, *8*(4), e60377. https://doi.org/10.1371/journal.pone.0060377

Ghosh Hajra, S., Liu, C. C., Song, X., Fickling, S. D., Cheung, T. P. L., & D'Arcy, R. C. N. (2018). Multimodal characterization of the semantic N400 response within a rapid evaluation brain vital sign framework. *Journal of Translational Medicine*, *16*. https://doi.org/10.1186/s12967-018-1527-2

Gredebäck, G., Johnson, S., & von Hofsten, C. (2009). Eye tracking in infancy research. *Developmental Neuropsychology*, *35*(1), 1–19.

Groen, M. A., Whitehouse, A. J. O., Badcock, N. A., & Bishop, D. V. M. (2011). Where were those rabbits? A new paradigm to determine cerebral lateralisation of visuospatial memory function in children. *Neuropsychologia*, *49*(12), 3265–3271. https://doi.org/10.1016/j.neuropsychologia.2011.07.031

Groen, M. A., Whitehouse, A. J. O., Badcock, N. A., & Bishop, D. V. M. (2012). Does cerebral lateralization develop? A study using functional transcranial Doppler ultrasound assessing lateralization for language production and visuospatial memory. *Brain and Behavior*, *2*(3), 256–269. https://doi.org/10.1002/brb3.56

Grootswagers, T., Wardle, S. G., & Carlson, T. A. (2017). Decoding dynamic brain patterns from evoked responses: A tutorial on multivariate pattern analysis applied to time-series neuroimaging data. *Journal of Cognitive Neuroscience*, *29*(4), 677–697. https://doi.org/10.1162/jocn_a_01068

Grummett, T., Leibbrandt, R., Lewis, T., DeLosAngeles, D., Powers, D., Willoughby, J., … Fitzgibbon, S. (2015). Measurement of neural signals from inexpensive, wireless and dry EEG systems. *Physiological Measurement*, *36*(7), 1469.

Haag, A., Moeller, N., Knake, S., Hermsen, A., Oertel, W. H., Rosenow, F., & Hamer, H. M. (2010). Language lateralization in children using functional transcranial Doppler sonography. *Developmental Medicine & Child Neurology*, *52*(4), 331–336. https://doi.org/10.1111/j.1469-8749.2009.03362.x

Hagoort, P., Baggio, G., & Wlllems, R. M. (2009). Semantic unification. In *The cognitive neurosciences, 4th ed* (pp. 819–835). Cambridge, MA, US: Massachusetts Institute of Technology.

Hahne, A., Eckstein, K., & Friederici, A. D. (2004). Brain signatures of syntactic and semantic processes during children's language development. *Journal of Cognitive Neuroscience*, *16*(7), 1302–1318. https://doi.org/10.1162/0898929041920504

Henderson, L. M., Baseler, H. A., Clarke, P. J., Watson, S., & Snowling, M. J. (2011). The N400 effect in children: Relationships with comprehension, vocabulary and decoding. *Brain and Language*, *117*(2), 88–99. https://doi.org/10.1016/j.bandl.2010.12.003

Holcomb, P. J., Coffey, S. A., & Neville, H. J. (1992). Visual and auditory sentence processing: A developmental analysis using event-related brain potentials. *Developmental Neuropsychology*, *8*(2–3), 203–241. https://doi.org/10.1080/87565649209540525

Hunt, L., Politzer-Ahles, S., Gibson, L., Minai, U., & Fiorentino, R. (2013). Pragmatic inferences modulate N400 during sentence comprehension: Evidence from picture-sentence verification. *Neuroscience Letters*, *534*, 246–251. https://doi.org/10.1016/j.neulet.2012.11.044

Huttenlocher, P. (1979). Synaptic density in human frontal cortex—Developmental changes and effects of aging. *Brain Research*, *163*(2), 195–205. https://doi.org/10.1016/0006-8993(79)90349-4

Huttenlocher, P. R., de Courten, C., Garey, L. J., & Van der Loos, H. (1982). Synaptogenesis in human visual cortex—Evidence for synapse elimination during normal development. *Neuroscience Letters*, *33*(3), 247–252. https://doi.org/10.1016/0304-3940(82)90379-2

Jones, W., & Klin, A. (2009). Heterogeneity and homogeneity across the autism spectrum: The role of development. *Journal of the American Academy of Child and Adolescent Psychiatry*, *48*(5), 471–473.

Juottonen, K., Revonsuo, A., & Lang, H. (1996). Dissimilar age influences on two ERP waveforms (LPC and N400) reflecting semantic context effect. *Cognitive Brain Research*, *4*(2), 99–107. https://doi.org/10.1016/0926-6410(96)00022-5

Kasari, C., Brady, N., Lord, C., & Tager-Flusberg, H. (2013). Assessing the minimally verbal school-aged child with autism spectrum disorder. *Autism Research: Official Journal of the International Society for Autism Research*, *6*(6), 479–493. https://doi.org/10.1002/aur.1334

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman brief intelligence test KBIT 2 ; manual*.

Kiang, M., Patriciu, I., Roy, C., Christensen, B. K., & Zipursky, R. B. (2013). Test-retest reliability and stability of N400 effects in a word-pair semantic priming paradigm. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *124*(4), 667–674. https://doi.org/10.1016/j.clinph.2012.09.029

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, *30*(5), 580–602. https://doi.org/10.1016/0749-596X(91)90027-H

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*(12), 463–470. https://doi.org/10.1016/S1364-6613(00)01560-6

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Landi, N., & Perfetti, C. A. (2007). An electrophysiological investigation of semantic and phonological processing in skilled and less-skilled comprehenders. *Brain and Language*, *102*(1), 30–45. https://doi.org/10.1016/j.bandl.2006.11.001

Lin, Y.-P., Wang, Y., & Jung, T.-P. (2014). Assessing the feasibility of online SSVEP decoding in human walking using a consumer EEG headset. *Journal of NeuroEngineering and Rehabilitation*, *11*(1), 119. https://doi.org/10.1186/1743-0003-11-119

Lohmann, H., Ringelstein, E. B., & Knecht, S. (2006). Functional transcranial Doppler sonography. *Frontiers of Neurology and Neuroscience*, *21*, 251–260. https://doi.org/10.1159/000092437

Makeig, S., Bell, A. J., Jung, T.-P., & Sejnowski, T. J. (1996). Independent Component Analysis of Electroencephalographic Data. In D. S. Touretzky, M. C. Mozer, & M. E. Hasselmo (Eds.), *Advances in Neural Information Processing Systems 8* (pp. 145–151). Retrieved from http://papers.nips.cc/paper/1091-independent-component-analysis-of-electroencephalographic-data.pdf

McCarthy, G., & Nobre, A. C. (1993). Modulation of semantic processing by spatial selective attention. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *88*(3), 210–219. https://doi.org/10.1016/0168-5597(93)90005-A

Minshew, N. J., Luna, B., & Sweeney, J. A. (1999). Oculomotor evidence for neocortical systems but not cerebellar dysfunction in autism. *Neurology*, *52*(5), 917. https://doi.org/10.1212/WNL.52.5.917

Miyake, A., Carpenter, P. A., & Just, M. A. (1994). A capacity approach to syntactic comprehension disorders: Making normal adults perform like aphasic patients. *Cognitive Neuropsychology*, *11*(6), 671–717. https://doi.org/10.1080/02643299408251989

Nakano, H., Saron, C., & Swaab, T. Y. (2010). Speech and span: Working memory capacity impacts the use of animacy but not of world knowledge during spoken sentence comprehension. *Journal of Cognitive Neuroscience*, *22*(12), 2886–2898. https://doi.org/10.1162/jocn.2009.21400

Nieuwland, M. S., & Van Berkum, J. J. A. (2008). The interplay between semantic and referential aspects of anaphoric noun phrase resolution: Evidence from ERPs. *Brain and Language*, *106*(2), 119–131. https://doi.org/10.1016/j.bandl.2008.05.001

O'Connor, K. (2012). Auditory processing in autism spectrum disorder: A review. *Neuroscience & Biobehavioral Reviews*, *36*(2), 836–854. https://doi.org/10.1016/j.neubiorev.2011.11.008

Olichney, J. M., Taylor, J. R., Gatherwright, J., Salmon, D. P., Bressler, A. J., Kutas, M., & Iragui-Madoz, V. J. (2008). Patients with MCI and N400 or P600 abnormalities are at very high risk for conversion to dementia. *Neurology*, *70*(19 Part 2), 1763. https://doi.org/10.1212/01.wnl.0000281689.28759.ab

Ortega, R., López, V., & Aboitiz, F. (2008). Voluntary modulations of attention in a semantic auditory-visual matching Task: An ERP study. *Biological Research*, *41*(4), 453–460. https://doi.org/10.4067/S0716-97602008000400010

Osterhout, L. (1997). On the brain response to syntactic anomalies: Manipulations of word position and word class reveal individual differences. *Brain and Language*, *59*(3), 494–522. https://doi.org/10.1006/brln.1997.1793

Osterhout, Lee, McLaughlin, J., Kim, A., Greenwald, R., & Inoue, K. (2004). Sentences in the brain: Event-related potentials as real-time reflections of sentence comprehension and language learning. *The On-Line Study of Sentence Comprehension: Eyetracking, ERP, and Beyond*, 271–308.

Perfetti, C. A., Wlotko, E. W., & Hart, L. A. (2005). Word Learning and Individual Differences in Word Learning Reflected in Event-Related Potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1281–1292. https://doi.org/10.1037/0278-7393.31.6.1281

Posar, A., & Visconti, P. (2018). Sensory abnormalities in children with autism spectrum disorder. *Jornal De Pediatria*, *94*(4), 342–350. https://doi.org/10.1016/j.jped.2017.08.008

Ries, A. J., Touryan, J., Vettel, J., McDowell, K., & Hairston, W. D. (2014). A Comparison of Electroencephalography Signals Acquired from Conventional and Mobile Systems. *Journal of Neuroscience and Neuroengineering*, *3*(1), 10–20. https://doi.org/10.1166/jnsne.2014.1092

Rogers, J. M., Johnstone, S. J., Aminov, A., Donnelly, J., & Wilson, P. H. (2016). Test-retest reliability of a single-channel, wireless EEG system. *International Journal of Psychophysiology*, *106*, 87–96.

Rohaut, B., Faugeras, F., Chausson, N., King, J.-R., Karoui, I. E., Cohen, L., & Naccache, L. (2015). Probing ERP correlates of verbal semantic processing in patients with impaired consciousness. *Neuropsychologia*, *66*, 279–292. https://doi.org/10.1016/j.neuropsychologia.2014.10.014

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(6), 906–914. https://doi.org/10.1002/wcs.78

Russell, J., Jarrold, C., & Henry, L. (1996). Working Memory in Children with Autism and with Moderate Learning Difficulties. *Journal of Child Psychology and Psychiatry*, *37*(6), 673–686. https://doi.org/10.1111/j.1469-7610.1996.tb01459.x

Sanders, L. D., Newport, E. L., & Neville, H. J. (2002). Segmenting nonsense: An event-related potential index of perceived onsets in continuous speech. *Nature Neuroscience*, *5*(7), 700–703. https://doi.org/10.1038/nn873

Schlosser, R. W., Hemsley, B., Shane, H., Todd, J., Lang, R., Lilienfeld, S. O., … Odom, S. (2019). Rapid Prompting Method and Autism Spectrum Disorder: Systematic Review Exposes Lack of Evidence. *Review Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s40489-019-00175-w

Srinivasan, R., Nunez, P. L., Tucker, D. M., Silberstein, R. B., & Cadusch, P. J. (1996). Spatial sampling and filtering of EEG with spline laplacians to estimate cortical potentials. *Brain Topography*, *8*(4), 355–366.

Steele, S. D., Minshew, N. J., Luna, B., & Sweeney, J. A. (2007). Spatial Working Memory Deficits in Autism. *Journal of Autism and Developmental Disorders*, *37*(4), 605–612. https://doi.org/10.1007/s10803-006-0202-2

Szewczyk, J. M., & Schriefers, H. (2018). The N400 as an index of lexical preactivation and its implications for prediction in language comprehension. *Language, Cognition and Neuroscience*, *33*(6), 665–686. https://doi.org/10.1080/23273798.2017.1401101

Tager-Flusberg, H. (2000). The challenge of studying language development in children with autism. *Methods for Studying Language Production*, 313–332.

Tager-Flusberg, H., & Kasari, C. (2013). Minimally Verbal School-Aged Children with Autism Spectrum Disorder: The Neglected End of the Spectrum. *Autism Research*, *6*(6), 468–478. https://doi.org/10.1002/aur.1329

Tanaka, H., Watanabe, H., Maki, H., Sakriani, S., & Nakamura, S. (2019). Electroencephalogram-Based Single-Trial Detection of Language Expectation

Violations in Listening to Speech. *Frontiers in Computational Neuroscience*, *13*. https://doi.org/10.3389/fncom.2019.00015

Tanner, D., Goldshtein, M., & Weissman, B. (2018). Individual differences in the real-time neural dynamics of language comprehension. *Psychology of Learning and Motivation*, *68*, 299–335.

Tanner, D., Mclaughlin, J., Herschensohn, J., & Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence*. *Bilingualism: Language and Cognition*, *16*(2), 367–382. https://doi.org/10.1017/S1366728912000302

Tomchek, S. D., & Dunn, W. (2007). Sensory processing in children with and without autism: A comparative study using the short sensory profile. *The American Journal of Occupational Therapy: Official Publication of the American Occupational Therapy Association*, *61*(2), 190–200. https://doi.org/10.5014/ajot.61.2.190

Yau, S. H., McArthur, G., Badcock, N. A., & Brock, J. (2015). Case study: Auditory brain responses in a minimally verbal child with autism and cerebral palsy. *Frontiers in Neuroscience*, *9*. https://doi.org/10.3389/fnins.2015.00208

Zander, T. O., Lehne, M., Ihme, K., Jatzev, S., Correia, J., Kothe, C., … Nijboer, F. (2011). A dry EEG-system for scientific research and brain–computer interfaces. *Frontiers in Neuroscience*, *5*, 53.

Appendix of this thesis has been removed as it may contain sensitive/confidential content