UNIVERSITY OF CALIFORNIA SAN DIEGO

What a bouncing ball tells us about the brain, development, and autism

A dissertation submitted in partial satisfaction of the requirement for the degree

Doctor of Philosophy

in

Experimental Psychology

by

Andrew Marin

Committee in Charge:

Professor Leslie J Carver, Chair
Professor Seana Coulson
Professor Shafali Jeste
Professor Lindsey J. Powell
Professor Viola S Störmer

2024

The Dissertation of Andrew Marin is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2024

DEDICATION

This dissertation is dedicated to my parents, whose unwavering support has been foundational for my successes. I am deeply grateful for everything you have done for me. I also dedicate this work to my incredible mentors, whose guidance, wisdom, and inspiration have shaped me into the scholar I am today. Your passion for knowledge and dedication to teaching have been a constant source of motivation. And lastly, to Emily, the four-year-old kid who always tested my patience but forever changed my perception.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I'd like to express my gratitude to the research assistants who have been instrumental in supporting my dissertation research. Your dedication and hard work have been invaluable contributions to the success of these projects. I will always be indebted to your assistance and commitment. Thank you to Sophia Bokovikova, Kayce Padilla, Ipek Talu, Melanie Dratva, Zara Fearns, Mincong Wu, Gabriela Bernal, Elli Connell, Alexander Nam, Danna Wu, Haley Rippon, Amanda Salatino, Joshua Cervantes, Carmen Chen, Xiaoyang Liu, Rylie Pepper-gjerman, Elise Zhao, and all the rest I forgot to mention.

I would like to acknowledge all our participants, big and small, and the families who participated in my research. Your time and insights have been instrumental in advancing our understanding of how the brain anticipates.

I want to extend my heartfelt thanks to my committee members: Viola Störmer, Lindsey Powell, Seana Coulson, and Shafali Jeste. Your guidance, expertise, and invaluable feedback throughout this dissertation process has been exceptional, all of which has shaped this work and my growth as a scientist.

And lastly, I would like to thank Leslie Carver, my advisor and chair to this dissertation. I would like to express my appreciation for Leslie's invaluable guidance and support throughout my time here at UCSD. Her expertise, encouragement, and dedication have been instrumental in shaping this work and helping me navigate the challenges of my research. I am especially grateful for her trust in my abilities and for empowering me to be an independent scholar. Above all, I want to express my thanks for your kindness and belief in my potential.

Chapter 1, in full, is a reprint of the material as it appears in Expectations about dynamic visual objects facilitates early sensory processing of congruent sounds in *Cortex*. Marin, Andrew; Störmer, Viola S.; Carver, Leslie J. (2021). The dissertation author was the primary investigator and author of this paper.

Chapter 2, in part is currently being prepared for submission for publication of the material. Marin, Andrew; Pearson, Lucy; Wu, Mincong; Baker, Elizabeth; Carver, Leslie J. The dissertation author was the primary investigator and author of this material.

Chapter 3, in part is currently being prepared for submission for publication of the material. Marin, Andrew; Fearns, Zara; Dratva, Melanie; Powell, Lindsey J.; Störmer, Viola S.; Carver, Leslie J. The dissertation author was the primary investigator and author of this material.

VITA

2012 Bachelor of Science, Psychology, San Jose State University

2015 Master of Arts, General-Experimental Psychology, Cal. State University, Northridge

2024 Doctor of Philosophy, Experimental Psychology, University of California San Diego

PUBLICATIONS

Marin, A., Störmer. V.S., & Carver, L.J. (2021). Expectations about dynamic visual objects facilitates early sensory processing of congruent sounds. *Cortex, 144*, 198-211. DOI: 10.1016/j.cortex.2021.08.006

Tran, X. A., McDonald, N. M., Dickinson, A., Scheffler, A., Frohlich, J., Marin, A., Liu, C. K., Nosco, E., Sentürk, D., Dapretto, M., & Jeste, S. S. (2020). Functional connectivity during language processing in 3-month-old infants at familial risk for autism spectrum disorder. *European Journal of Neuroscience, 53*(5), 1621-1637. DOI: 10.1111/ejn.15005

Dickinson, A., Daniel, M., Marin, A., Goanker, B., Dapretto, M., McDonald, N. M., & Jeste, S. (2020). Multivariate neural connectivity patterns in early infancy predict later autism symptoms. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 6*(1), 59-69. DOI: 10.1016/j.bpsc.2020.06.003

Marin, A., Hutman, T., Ponting, C., McDonald, N. M., Carver, L.J., Baker, E., Daniel, M., Dickinson, A., Dapretto, M., Johnson, S. P., & Jeste, S. S. (2020). Electrophysiological signatures of visual statistical learning in three-month old infants at familial and low risk for autism spectrum disorder. *Developmental Psychobiology, 62*(6), 858-870. DOI: 10.1002/dev.21971

FIELDS OF STUDY

Major Field: Psychology

Studies in Developmental Psychology

ABSTRACT OF THE DISSERTATION

What a bouncing ball tells us about the brain, development, and autism

by

Andrew Marin

Doctor of Philosophy in Experimental Psychology

University of California San Diego, 2024

Professor Leslie Carver, Chair

In everyday perception, dynamic objects move and collide within physical environments, producing expected sounds. In this dissertation, I suggest that perceptual phenomena, like a bouncing ball, may offer mechanistic insights into: 1) how the brain anticipates sound via the integration of dynamic visual cues, 2) clinical populations who show differences in the ability to anticipate, and 3) the developmental emergence of skills used to anticipate sound. In a series of experiments, I presented neurotypical and autistic adults, and neurotypical infants a dynamic visual object that collides with a physical barrier, eliciting a sound at the point of expected collision (AV-synchronous), or unexpectedly before collision (AV-asynchronous). In chapter one, I recorded event-related potentials (ERPs) from neurotypical adults who were exposed to sounds that either synchronized with visual collision or occurred asynchronously before collision. I also included conditions where the object was occluded during synchronous collision, or when sound was presented without dynamic visual cues. I found that synchronous and occluded collision sounds elicited an attenuated auditory response relative to asynchronous or audio-only sounds. These results suggest that dynamic visual stimuli can help generate expectations about the timing of sound, which then facilitates the processing of auditory information that matches these expectations.

In chapter two, I replicated the same methods as in chapter one, but in a sample of autistic adults. Here, I observed greater amplitudes toward asynchrony in autism relative to neurotypicals, while no group differences toward fully visible or occluded synchrony emerged. These results suggest that neural responses to prediction errors are affected in autism, and not the integration of top-down expectations. In chapter three, I modified these methods for use in neurotypical infants to show that 4-to-5-month-olds look longer to bounce sounds that violate temporal expectations of when a bounce sound should occur.

These studies highlight the presence of neural mechanisms sensitive to predictable sound, which appear to be different in clinical populations like autism. Moreover, infants are sensitive to collision sounds, demonstrating that these perceptual skills are available early in life. Collectively, these methods could be further leveraged to understand the emergence of neurodevelopmental conditions like autism.

Introduction

Imagine a cracked egg dropping onto a sizzling pan, a deck of cards being shuffled, or hands clapping. Each experience has one thing in common: they involve dynamic visual objects signaling expectations toward sound. Processing expected auditory input generated by a moving visual object is a difficult task for the brain to implement. It requires the assessment of the object's spatial location moving within the physical constraints of the environment, and to then use this information to infer the relation between the sound and the object. Such processes are multimodal, requiring the brain to depend on integrative neural pathways between visual and auditory modalities. Critically, these pathways may be impacted in clinical populations, like autistic individuals, who experience widespread structural neural alterations (Belmonte, 2004; Bourgeron, 2015; Geschwind & Levitt, 2007; Parikshak et al., 2015; Port et al., 2014).

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition characterized by challenges in social communication and the presence of repetitive behaviors or interests (American Psychiatric Association, 2022). In addition, several sensory differences that include the processing of faces and emotional stimuli (Eussen et al., 2015; Harms et al., 2010; Pellicano et al., 2007; Uljarevic & Hamilton, 2013), multisensory integration (Stevenson et al., 2014; Wallace & Stevenson, 2014), and the presence of hypo- and hyper- sensory sensitivities (Baranek et al., 2013; Robertson & Baron-Cohen, 2017) have been identified in autistic individuals. Up to 90% of autistic people exhibit altered sensory processing sensitivities (Tavassoli et al., 2014; Tomchek & Dunn, 2007) that affect each sensory modality: touch (Marco et al., 2012; Puts et al., 2014), taste (Tavassoli & Baron-Cohen, 2012), smell (Galle et al., 2013; Rozenkrantz et al., 2015), audition (Bonnel et al., 2003), and vision (Simmons et al., 2009).

Clearly sensory sensitivities play a role in autism, influencing neural processing and shaping perceptual experiences that contribute to unique symptom profiles.

Numerous hypotheses have been proposed to explain the etiology of the heterogenous symptom profile that underlies autism. Recent theories posit that several symptoms of autism (e.g., insistence on sameness, deficits in social interaction) may stem from underlying differences in predictive coding (Cannon et al., 2021; Lawson et al., 2014; Pellicano & Burr, 2012; Sinha et al., 2014; Van Boxtel & Lu, 2013; Van de Cruys et al., 2014). A key concept of predictive coding is that the brain generates predictions about the present state of the world based on previous and current sensory input. Predictions (or expectations) serve as "priors", which are internal representations of the probabilistic structure of one's environment (Clark, 2013; Friston, 2005; Lawson et al., 2014) and are continually contrasted with current sensory input to contextualize and inform our perception (Dempster, 1968; Knill & Richards, 1996).

In predictive coding, predictions are not just generated, but are actively compared to the potential error of those predictions. If the prediction error informs our perception, prior expectations are readjusted to minimize future errors. Given that our sensory world behaves with some degree of uncertainty, and that the neural systems involved in processing sensory information are noisy themselves, prediction errors can be wrong or uninformative. Thus, the influence of top-down expectancies relative to bottom-up processing is mediated by the *precision*, or confidence bestowed upon the prediction error (Friston, 2010). If the resulting error signal is perceived as informative, it is then passed up the hierarchy to inform higher-level, top-down expectations to effectively resolve prediction errors in the future. High sensory precision would result in the individual placing more weight in prediction errors generated at each level of the processing hierarchy, while low sensory precision would attenuate the influence of bottom-up signals at each level, and bias perception toward prior beliefs. Therefore, a healthy balance between using informative errors to update one's expectations, or to employ an appropriate amount of precision when encountering error is necessary for effective mental models.

In this sense, predictive coding theory suggests that a primary function of the brain is to regulate prediction errors that can be encountered at any level of the processing stream, which is computed within contextualized mental models. This predictive mechanism is critical for learning, allowing us to anticipate perception and readjust when our expectation is violated. As such, differences in the integration and optimization of prediction errors has been proposed as a mechanism that may be responsible for predictive dysfunction in autism. Specifically, differences related to 1) generating predictions and/or 2) detecting violations to those predictions (i.e., prediction errors) have unique implications for learning within both social and non-social domains, and alterations to predictive mechanisms are what likely causes the key symptoms seen in autism (Cannon et al., 2021; Lawson et al., 2014; Pellicano & Burr, 2012; Sinha et al., 2014; Van Boxtel & Lu, 2013; Van de Cruys et al., 2014).

With this dissertation, I will look to quantify the mechanisms involved in predicting sound via the integration of dynamic visual cues. Grounded in predictive coding theory, I will use a novel electrophysiological (EEG) method to measure the brain's response to expected sounds that are either congruent or incongruent with an object's motion (chapter one). I will use these same methods to compare the neural response toward the generation of successful predictions versus ones that signify error in autistic adults (chapter two). I will then propose a study using similar methods to measure infant looking time toward collision events (chapter three). I will then close with a theoretical summary of how the mechanisms related to auditory predictions may help explain the emergence of autism early in development.

References

American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.). https://doi.org/10.1176/appi.books.9780890425787

Baranek, G. T., Watson, L. R., Boyd, B. A., Poe, M. D., David, F. J., & McGuire, L. (2013). Hyporesponsiveness to social and nonsocial sensory stimuli in children with autism, children with developmental delays, and typically developing children. *Development and Psychopathology*, *25*(2), 307–320. https://doi.org/10.1017/S0954579412001071

Belmonte, M. K. (2004). Autism and abnormal development of brain connectivity. *Journal of Neuroscience*, *24*(42), 9228–9231. https://doi.org/10.1523/JNEUROSCI.3340-04.2004

Bonnel, A., Mottron, L., Peretz, I., Trudel, M., Gallun, E., & Bonnel, A. M. (2003). Enhanced pitch sensitivity in individuals with autism: A signal detection analysis. *Journal of Cognitive Neuroscience*, *15*(2), 226–235. https://doi.org/10.1162/089892903321208169

Bourgeron, T. (2015). From the genetic architecture to synaptic plasticity in autism spectrum disorder. *Nature Reviews Neuroscience*, *16*, 551-563. https://doi.org/10.1038/nrn3992

Cannon, J., O'Brien, A. M., Bungert, L., & Sinha, P. (2021). Prediction in autism spectrum disorder: A systematic review of empirical evidence. *Autism Research*, *14*(4), 604–630. https://doi.org/10.1002/aur.2482

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Dempster, A. P. (1968). A generalization of bayesian inference. *Journal of the Royal Statistical Society: Series B (Methodological)*, *30*(2), 205–232. https://doi.org/10.1111/j.2517-6161.1968.tb00722.x

Eussen, M. L. J. M., Louwerse, A., Herba, C. M., Van Gool, A. R., Verheij, F., Verhulst, F. C., & Greaves-Lord, K. (2015). Childhood facial recognition predicts adolescent symptom severity in autism spectrum disorder. *Autism Research*, *8*(3), 261–271. https://doi.org/10.1002/aur.1443

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Galle, S. A., Courchesne, V., Mottron, L., & Frasnelli, J. (2013). Olfaction in the autism spectrum. *Perception*, *42*(3), 341–355. https://doi.org/10.1068/p7337

Geschwind, D. H., & Levitt, P. (2007). Autism spectrum disorders: Developmental disconnection syndromes. *Development*, *17*(1), 103–111. https://doi.org/10.1016/j.conb.2007.01.009

Harms, M. B., Martin, A., & Wallace, G. L. (2010). Facial emotion recognition in autism spectrum disorders: A review of behavioral and neuroimaging studies. *Neuropsychology Review*, *20*(3), 290–322. https://doi.org/10.1007/s11065-010-9138-6

Knill, D. C., & Richards, W. (Eds.). (1996). *Perception as Bayesian Inference*. Cambridge University Press.

Lawson, R. P., Rees, G., & Friston, K. J. (2014). An aberrant precision account of autism. *Frontiers in Human Neuroscience*, *8*, 302. https://doi.org/10.3389/fnhum.2014.00302

Marco, E. J., Khatibi, K., Hill, S. S., Siegel, B., Arroyo, M. S., Dowling, A. F., Neuhaus, J. M., Sherr, E. H., Hinkley, L. N. B., & Nagarajan, S. S. (2012). Children with autism show reduced somatosensory response: An MEG study. *Autism Research*, *5*(5), 340–351. https://doi.org/10.1002/aur.1247

Parikshak, N. N., Gandal, M. J., & Geschwind, D. H. (2015). Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. *Nature Reviews Genetics*, *16*(8), 441–458. https://doi.org/10.1038/nrg3934

Pellicano, E., & Burr, D. (2012). When the world becomes 'too real': A Bayesian explanation of autistic perception. *Trends in Cognitive Sciences*, *16*(10), 504–510. https://doi.org/10.1016/j.tics.2012.08.009

Pellicano, E., Jeffery, L., Burr, D., & Rhodes, G. (2007). Abnormal adaptive face-coding mechanisms in children with autism spectrum disorder. *Current Biology*, *17*(17), 1508–1512. https://doi.org/10.1016/j.cub.2007.07.065

Port, R. G., Gandal, M. J., Roberts, T. P. L., Siegel, S. J., & Carlson, G. C. (2014). Convergence of circuit dysfunction in ASD: A common bridge between diverse genetic and environmental risk factors and common clinical electrophysiology. *Frontiers in Cellular Neuroscience*, *8*. https://doi.org/10.3389/fncel.2014.00414

Puts, N. A. J., Wodka, E. L., Tommerdahl, M., Mostofsky, S. H., & Edden, R. A. E. (2014). Impaired tactile processing in children with autism spectrum disorder. *Journal of Neurophysiology*, *111*(9), 1803–1811. https://doi.org/10.1152/jn.00890.2013

Robertson, C. E., & Baron-Cohen, S. (2017). Sensory perception in autism. *Nature Reviews Neuroscience*, *18*(11), 671–684. https://doi.org/10.1038/nrn.2017.112

Rozenkrantz, L., Zachor, D., Heller, I., Plotkin, A., Weissbrod, A., Snitz, K., Secundo, L., & Sobel, N. (2015). A mechanistic link between olfaction and autism spectrum disorder. *Current Biology*, *25*(14), 1904–1910. https://doi.org/10.1016/j.cub.2015.05.048

Simmons, D. R., Robertson, A. E., McKay, L. S., Toal, E., McAleer, P., & Pollick, F. E. (2009). Vision in autism spectrum disorders. *Vision Research*, *49*(22), 2705–2739. https://doi.org/10.1016/j.visres.2009.08.005

Sinha, P., Kjelgaard, M. M., Gandhi, T. K., Tsourides, K., Cardinaux, A. L., Pantazis, D., Diamond, S. P., & Held, R. M. (2014). Autism as a disorder of prediction. *Proceedings of the National Academy of Sciences*, *111*(42), 15220. https://doi.org/10.1073/pnas.1416797111

Stevenson, R. A., Siemann, J. K., Schneider, B. C., Eberly, H. E., Woynaroski, T. G., Camarata, S. M., & Wallace, M. T. (2014). Multisensory temporal integration in autism spectrum disorders. *Journal of Neuroscience*, *34*(3), 691–697. https://doi.org/10.1523/JNEUROSCI.3615-13.2014

Tavassoli, T., & Baron-Cohen, S. (2012). Taste identification in adults with autism spectrum conditions. *Journal of Autism and Developmental Disorders*, *42*(7), 1419–1424. https://doi.org/10.1007/s10803-011-1377-8

Tavassoli, T., Miller, L. J., Schoen, S. A., Nielsen, D. M., & Baron-Cohen, S. (2014). Sensory over-responsivity in adults with autism spectrum conditions. *Autism*, *18*(4), 428–432. https://doi.org/10.1177/1362361313477246

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

Uljarevic, M., & Hamilton, A. (2013). Recognition of emotions in autism: A formal meta-analysis. *Journal of Autism and Developmental Disorders*, *43*(7), 1517–1526. https://doi.org/10.1007/s10803-012-1695-5

Van Boxtel, J., & Lu, H. (2013). A predictive coding perspective on autism spectrum disorders. *Frontiers in Psychology*, *4*, 19. https://doi.org/10.3389/fpsyg.2013.00019

Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: Predictive coding in autism. *Psychological Review*, *121*(4), 649–675. https://doi.org/10.1037/a0037665

Wallace, M. T., & Stevenson, R. A. (2014). The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. *Neuropsychologia*, *64*, 105–123. https://doi.org/10.1016/j.neuropsychologia.2014.08.005

Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Research Report

Expectations about dynamic visual objects facilitates early sensory processing of congruent sounds

Andrew Marin [a,*], *Viola S. Störmer* [b] *and Leslie J. Carver* [a]

[a] *University of California, San Diego (UCSD), Psychology Department, La Jolla, CA, USA*
[b] *Dartmouth College, Department of Psychological and Brain Sciences, Hanover, NH, USA*

ARTICLE INFO

Article history:
Received 16 February 2021
Reviewed 19 March 2021
Revised 17 May 2021
Accepted 5 August 2021
Action editor D. Samuel Schwarzkopf
Published online 21 September 2021

Keywords:
Audio-visual
Expectation
ERP
Temporal synchrony

ABSTRACT

The perception of a moving object can lead to the expectation of its sound, yet little is known about how visual expectations influence auditory processing. We examined how visual perception of an object moving continuously across the visual field influences early auditory processing of a sound that occurred congruently or incongruently with the object's motion. In Experiment 1, electroencephalogram (EEG) activity was recorded from adults who passively viewed a ball that appeared either on the left or right boundary of a display and continuously traversed along the horizontal midline to make contact and elicit a bounce sound off the opposite boundary. Our main analysis focused on the auditory-evoked event-related potential. For audio-visual (AV) trials, a sound accompanied the visual input when the ball contacted the opposite boundary (AV-synchronous), or the sound occurred before contact (AV-asynchronous). We also included audio-only and visual-only trials. AV-synchronous sounds elicited an earlier and attenuated auditory response relative to AV-asynchronous or audio-only events. In Experiment 2, we examined the roles of expectancy and multisensory integration in influencing this response. In addition to the audio-only, AV-synchronous, and AV-asynchronous conditions, participants were shown a ball that became occluded prior to reaching the boundary of the display, but elicited an expected sound at the point of occluded collision. The auditory response during the AV-occluded condition resembled that of the AV-synchronous condition, suggesting that expectations induced by a moving object can influence early auditory processing. Broadly, the results suggest that dynamic visual stimuli can help generate expectations about the timing of sounds, which then facilitates the processing of auditory information that matches these expectations.

[*] *Corresponding author*. University of California, San Diego (UCSD), Psychology Department, 9500 Gilman Drive, La Jolla, CA, 92093.
E-mail addresses: amarin@ucsd.edu (A. Marin), viola.s.stoermer@dartmouth.edu (V.S. Störmer), ljcarver@ucsd.edu (L.J. Carver).

https://doi.org/10.1016/j.cortex.2021.08.006

1. Introduction

In everyday life, dynamic visual objects often predict accompanying sounds. For example, observing two hands moving closer together precedes the onset of a clap, or a marble contacting another results in a sound precisely at the point of collision. These scenarios showcase how motion (i.e., directionality, speed, etc.) and physical cues (i.e., artificially defined object boundaries, collision, etc.) of dynamic visual objects in natural sensory environments elicit expected sounds at precise moments in time and space. Perceiving such events uniquely highlights how visual anticipation can directly interact with auditory processing—yet little is known about how auditory processing is influenced by preceding visual information about moving objects. We know that dynamic visual objects can elicit sounds in natural sensory environments, but does this visually driven anticipation facilitate early auditory processing?

Dynamically moving visual objects often generate sounds that can be predicted from the temporal expectancy laid forth by the object itself. Accurately inferring the source of sounds generated by a moving object involves matching temporally synchronous visual information with the sound. Such inferences may reflect a mechanism designed to exploit the temporal and spatial information of a moving object to make predictions about expected sounds in the environment. Questions regarding stimulus prediction and its brain bases have received considerable amounts of attention, particularly within the auditory domain (for review, Lange, 2013). Event-related potentials (ERPs), which reflect the averaged electroencephalogram (EEG) response time-locked to a particular event, have been utilized to examine how early auditory responses are shaped by predictable sensory information. One particular ERP response that is modulated by predictable sounds is the auditory evoked potential (i.e., the N1–P2 complex), which is an early sensory response elicited after a sound. Many studies have reported that the auditory response is attenuated when hearing temporally predictable sounds (Clementz et al., 2002; D'Andrea-Penna et al., 2020; Ford et al., 2007; Ford & Hillyard, 1981; Kononowicz & van Rijn, 2014; Lange, 2009; Menceloglu et al., 2020; Schafer et al., 1981). Auditory response suppression toward expected sounds has usually been interpreted within a general predictive coding framework (Friston, 2005; Lange, 2013), where the reduction of the auditory response is thought to arise due to top-down expectancies matching bottom-up sensory input. The synchronous match between bottom-up and top-down signals is thought to reduce the error signal of the predicted sound, which results in an overall reduction of the evoked ERP (Baldeweg, 2007; Lange, 2013). Yet, many of the studies mentioned here cued the expectation of the sound within the temporal domain by providing the perceiver foreknowledge about when a sound would occur. However, in natural sensory environments, sounds are more likely to be preceded by visual stimuli that are often moving across time and space.

The perception of simultaneity of discrete audio-visual (AV) events in time and space plays a large role in determining if the two sensory events will be perceptually bounded as one, or perceived as two separate events (Körding et al., 2007; for review, see Wallace & Stevenson, 2014). Discrete sensory events that remain in close temporal proximity to one another are more likely to be integrated as one, whereas sensory events that are further away in time and space are more likely to be perceived as two distinct events (Spence, 2007; Stevenson, Zemtsov, & Wallace, 2012; Stevenson, Fister, et al., 2012; van Wassenhove et al., 2007). Simple AV stimuli like pure auditory tones and geometrical visual shapes have been associated with an enlarged auditory neural response compared to the sum of unimodal presentations (Fort et al., 2002; Giard & Peronnet, 1999; Molholm et al., 2002). Enhancement of early neural responses while perceiving multisensory simultaneity has been theorized as a general principle of multisensory processing (Meredith et al., 1987). Other demonstrations of AV integration in natural environments involve the perception of speech sounds, human actions, and dynamic visual objects. The auditory-evoked potential has been found to occur earlier in time and elicit a smaller amplitude response when speech sounds are paired with synchronous mouth movements compared to auditory-only presentations, which has been interpreted as auditory processing being suppressed when paired with visual information (van Wassenhove et al., 2005). In the case of speech perception, visual information originating from mouth movements precedes paired auditory outputs by tens to a few hundreds of milliseconds (van Wassenhove et al., 2005), likely leading to strong expectations about when and what sound will appear. Other research has shown that such expectations can also be triggered by non-speech stimuli, for example human actions (i.e., a hand clap) and dynamic objects (i.e., a hammer tapping a cup), which have been also shown to attenuate the auditory response (Stekelenburg & Vroomen, 2007). Critically, decreases in the auditory response were not seen with objects that did not provide anticipatory visual motion information (Stekelenburg & Vroomen, 2007), suggesting that the amplitude reduction underlying the early auditory response occurs when visual information provides clear expectations about the onset of a sound. Together, these studies suggest that a smaller amplitude of the auditory response may arise in situations that allow a relatively long build-up of visual expectations (e.g., through the movement of hands or lips), in that the visual information allows one to predict the upcoming acoustic signal, and subsequently reduces uncertainty and lowers computational demands of auditory brain regions (Besle et al., 2004; Vroomen & Stekelenburg, 2010).

Another study further supports the notion that changes in early auditory processing only arise when visual information reliably predicts sound onset. Vroomen & Stekelenburg found early attenuation of the auditory evoked potential when viewing simple visual stimuli that provided expectations of an anticipated sound compared to auditory-alone presentations (2010). In their task, for AV expectation trials, two disks appeared to the extreme left and right of a vertically aligned rectangle that was presented at the center of a display. Here, each visual disk moved toward the rectangle and eventually collided with it, compressing it and eliciting a synchronous pure tone at the point of collision. Participants were also exposed to two other conditions where 1) the dynamic visual stimuli collided with the rectangle but did not contain an expected sound (visual-only) and 2) audio-only trials that contained sound with no visual stimulus. These three trial types appeared in a single block, in random order. Importantly, in a new block of trials, participants were exposed to a new AV condition that did not provide visual expectations about when the sound would appear. In this condition, there were no visual disks, but the rectangle eventually compressed and made a sound upon doing so. Within this block, subjects were also presented with the same audio- and visual-only trials described previously, and each were presented in a random order. In a follow-up experiment, Vroomen & Stekelenburg presented the same two AV trials, as well as the audio- and visual-only conditions explained above. In addition, they 1) provided two new sensory conditions where the sound either happened before or after the collision event (AV-asynchronous; early and late) and 2) manipulated whether these sensory conditions appeared in a fixed or random order (2010). Here, the early amplitude response during fixed block ordering was reduced while perceiving synchronous AV expectations compared to audio-alone input, an effect not seen during mixed block ordering. This led the authors to suggest that auditory reduction arises when visual information reliably predicts AV onset across trials. Moreover, they found that the auditory response was not different in amplitude or latency between the synchronous and asynchronous AV inputs during mixed order presentations. Interestingly, a later component (i.e., the P2) showed a different pattern and was attenuated when early asynchronous and synchronous AV stimuli were fixed and varied from trial to trial. This suggests a possible dissociation between these two components of the auditory ERP. Taken together, these studies suggest that the neural effects of AV expectations depend on various factors, such as temporal synchrony, the amount of visual and auditory input, whether or not trial information was known beforehand, and the stimuli used.

One factor not considered in Vroomen and Stekelenburg (2010) was what would happen with a more naturalistic visual event such as a single object, moving in a uniform direction (i.e., a ball bouncing off a wall). Here, we fill this gap in the literature by better characterizing the neural correlates governing the anticipation of dynamic, temporally synchronous AV processing. Unidirectional dynamic visual stimuli might provide 1) the visual system more precise expectations about the collision event that elicits the anticipatory sound and/or 2) the auditory event itself may be more predictable when the accompanying visual stimulus is moving unidirectionally. Dynamic visual stimuli moving unidirectionally, in turn, may afford the visual system greater sensitivities toward small temporal AV asynchronies sooner in the auditory processing stream. Furthermore, it is currently unknown whether AV effects occur based on expectations alone, or whether visual objects and sounds need to be both present in order to affect sensory processing. Thus, the primary objective of our study was to examine how dynamic visual input—a single object moving continuously in one direction across the visual field—influences early auditory processing of a sound that is either congruent with the object's motion, and thus likely perceived as being part of the visual object, or incongruent with the object's motion. We were guided by the hypothesis that AV temporal synchrony would result in an attenuated and faster auditory response, compared to a unimodal auditory presentation—a response profile that would mimic the auditory effects seen in Stekelenburg and Vroomen (2007) and Vroomen and Stekelenburg (2010). Considering the null findings regarding the neural response toward the synchrony of dynamic AV input outlined above (i.e., 2007, 2010), we expected differences might appear because our stimuli were designed to constrain visual expectations in a single direction, perhaps affording the brain greater sensitivity toward small temporal asynchronies earlier in time. We also examined whether such auditory ERP effects only occur when a visual stimulus is presented at the same time as the sound, as predicted by multisensory integration accounts, or whether the expectation triggered by a moving visual stimulus is sufficient in influencing auditory processing. To test this hypothesis, we conducted a second experiment and examined auditory responses elicited by visual anticipatory information that becomes occluded prior to temporally congruent collision.

2. Methods

We report all data exclusions, all inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all manipulations, and all measures in the study. A statement on how sample size was determined can be found in the methods section of Experiment 2.

2.1. *Participants*

Twenty-nine college-aged adults (M_{age} = 20.48 years, *SD* = 1.76; 14 female) were recruited via an online university subject pool and received course credit for participating. Prior to the experiment, each participant reported having normal or corrected-to-normal vision, normal hearing, and no history of neuropsychological, cognitive, or developmental disorders. All participants provided written informed consent in accordance with the tenets of the 1964 Declaration of Helsinki. An additional eight adults were tested but were excluded due to equipment malfunction (*n* = 3) and excessive (>10% of trials) EEG artifact (i.e., head motion, muscle artifact etc.; *n* = 5).

2.2. *Audio-visual stimuli*

The AV stimuli used in this experiment were the same as in Werchan et al. (2018). The stimuli were created using Adobe After Effects software, while stimulus delivery was controlled by E-prime software (Psychology Software Tools, 2016) and presented on a CRT monitor (13 width × 9.5 height; in inches), with a 60 Hz refresh rate. Participants viewed the stimuli at an average distance of 71 cm. The primary object of interest was a red ball that was one inch in diameter, subtending a visual angle of about 2.05°. The ball appeared within a black rectangle (7.75 width × 5.5 height in inches; visual angle $_{width}$ = 15.8°; subtended visual angle $_{height}$ = 11.2°) that was overlaid on top of a neutral gray background (13 width × 9.5 height; in inches; visual angle $_{width}$ = 26.2°; visual angle $_{height}$ = 19.3°). The inside of the black rectangle contained a grid of small white dots that emphasized the straight, horizontal motion of the red ball. The ball's horizontal movement was constrained to occur within the black rectangle at a rate of 2.5s per motion cycle (i.e., visual object starts and returns to its origin). The sounds were presented via two speakers presented to the left and right of the monitor. The sound itself was a 50 decibel (dB) complex tone that resembled a solid object colliding with a hard surface (a knocking sound) and had a duration of 200 msec.

2.3. *Paradigm and procedure*

Each participant was seated in a dark room and was shown a randomly presented stream of four AV sensory conditions: 1) visual-only, 2) audio-only, 3) AV-synchronous, and 4) AV-asynchronous, while high-density EEG (Electrical Geodesics, Inc.) was recorded. The experimental session took part during a single lab visit, and the EEG recording lasted approximately 45 min. At the start of each trial (see Fig. 1 for a single trial diagram), a geometric, achromatic fractal video was presented for 1000 msec, and served as visual input to promote participant attention and engagement during the passive viewing task.

A small fixation cross (.75 width × .75 in height; 1.5 of visual angle) then appeared for 1000 msec, followed by the random presentation of one of the four sensory conditions previously mentioned. There were a total of 416 experimental trials (104 trials per condition), split into four blocks. The experiment was split into blocks to provide breaks to the participant as needed. The primary part of each trial (where the ball moved across the screen, described in more detail below) lasted 2000 msec. Upon completion of each trial, a single letter (.75 width × .75 in height; 1.5 subtended visual angle), out of a possible of eight, appeared randomly in the center of the screen for 500 msec. At the start of each experimental block, the participant was instructed to identify and count, using a handheld clicker, a single target letter. This secondary task served as an attention check to keep each participant engaged during the passive viewing task. All participants were above a 95% accuracy rate in the secondary task so no subjects were removed due to poor attention.

For the visual-only, AV-synchronous, and AV-asynchronous conditions, a single red ball randomly appeared on either the far left or right boundary of the black rectangle display. The ball then traversed horizontally to make contact with and bounce off the opposite boundary of the black rectangle. Participants were not explicitly told to maintain fixation but were encouraged to not track the exact motion of the ball and to take in the stimuli holistically. The EEG data was cleaned for any eye movements that occurred (see EEG data processing). The time between the start of the red ball's motion to the time it reached the opposite boundary of the display was 1200 msec. For the AV-synchronous condition, the bounce-sound occurred at 1200 msec, exactly when the ball touched the opposite boundary of the grid. During the AV-asynchronous condition, the bounce-sound occurred at 750 msec post stimulus onset, which corresponded to the ball just having passed the vertical midline as it moved toward the opposite wall, and before it contacted the wall. The visual-only condition contained the ball moving and bouncing off the

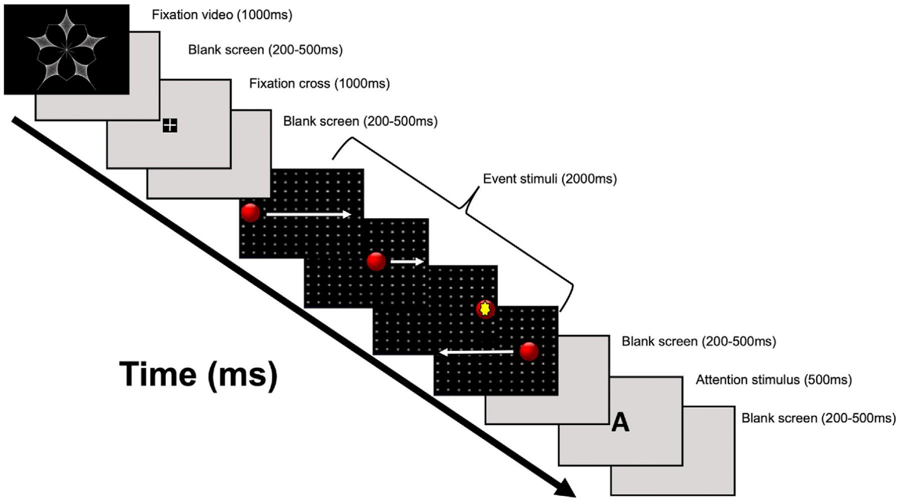

Fig. 1 – Single trial schematic depicting the AV-synchronous condition. For each trial, a small achromatic fractal video was first presented for 1000 msec (ms) to promote participant engagement during the passive viewing task. A small fixation cross then appeared, followed by the random presentation of one of four sensory conditions for a total of 416 experimental trials, 104 trials per condition. The interval labeled "Event Stimuli" contains the primary part of the trial, where participants viewed a ball that moved across the screen and may or may not be presented with auditory input. A visual depiction and description of each sensory condition is outlined within Fig. 2. Lastly, upon completion of the event stimuli, each participant was instructed to identify a single target letter among 7 distractor letters, which served as an attention check.

opposite boundary of the black rectangle with no sound. The ball for the AV-synchronous, AV-asynchronous, and visual-only conditions remained stationary at the opposite boundary for 50 msec, so the time from the start of motion in one direction and the start of motion back to its origin was 1250 msec. At 1250 msec, the ball started to move toward its origin at the same speed and the stimulus subsequently terminated at 2000 msec, well before contact with the boundary of origin. The audio-only condition contained the black rectangle and no visual input provided by the red ball, but the bounce-sound occurred at 1200 msec after the start of the trial. The duration of the sound for the audio-only, AV-synchronous, and AV-asynchronous conditions was 200 msec. Each trial occurred with equal probability and was randomly generated within each experimental block (see Fig. 2 for a visual diagram of each sensory condition).

2.4. *EEG data processing*

Continuous EEG was recorded via a 128-channel HydroCel Geodesic Sensor Net (Electrical Geodesics, Inc.; EGI). Impedances were kept below 50 kOhms in all electrodes and the raw EEG data were referenced online to the vertex (Cz) and digitized at 500 Hz. EEG data were amplified according to the default settings of an EGI internal amplifier (model type: Net Amps 300). All data were processed off-line using MATLAB (Mathworks, Inc.) and EEGLAB/ERPLAB software (Delorme & Makeig, 2004; Lopez-Calderon & Luck, 2014). A video of each participant was obtained during the EEG recording to ensure they kept their eyes on the display during the EEG session.

The raw EEG data were first digitally filtered using a .05–50 Hz bandpass (Butterworth) and 60 Hz notch filters. Data were then manually inspected for individual bad channels present throughout at least 50% of the recording, as well as electromyographic (EMG) and other movement artifacts. EEG data with evidence of egregious EMG, movement, or muscle artifacts were rejected from the analysis. Data from bad channels were replaced using a spherical spline interpolation algorithm. The cleaned EEG data were then taken through an independent component analysis (ICA), where evidence of eye artifact (i.e., eye blinks and saccades) was removed from the

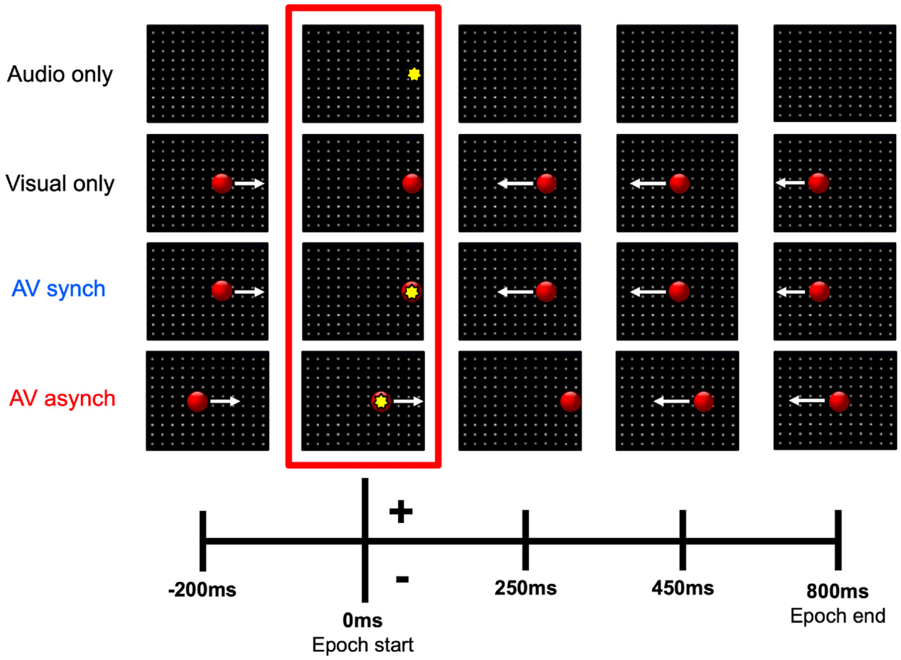

Fig. 2 – Depiction of the left-start sensory conditions (right-start not pictured). The moving spherical visual object is shown in red, the presentation of the sound is depicted as a bright yellow star, and the white arrows indicate the direction of motion of the visual object. The red rectangle drawn over a single frame of each condition reflects the point in time where we are event-locking the EEG data for each sensory condition. All ERPs were time-locked to the sound presentation, or in case of the visual-only condition, the moment when the ball bounced off the boundary. The audio-only condition contained no visual input at all. The visual-only condition contained the dynamic motion of the ball, with no audio input. For the synchronous condition, the bounce sound occurred when the ball first made contact with the boundary of the grid. For the asynchronous condition, the bounce sound corresponded to the ball just having moved past vertical midline. All conditions were equally likely to occur and were randomly intermixed across the experiment. The onset of each trial followed an inter-stimulus interval with a randomly presented jitter of 500–750 msec.

data set. To ensure that all ocular-related artifacts were eliminated by the ICA successfully, we scrolled through the entire raw EEG traces to look for any residual eye blinks or eye movements and if present, removed them by hand. This ICA procedure ensured that no eye movement artifacts were contained in the final data. Thus, while it is possible that some participants moved their eyes less in one condition than the other (e.g., audio-only vs. visual-present), this should not affect our ERP results. We also noticed ICA components in the data that resembled high-frequency harmonics and opted to remove them. The EEG data were then segmented into 1000 msec epochs (−200 to 800 msec relative to stimulus onset), and baseline corrected using mean voltage during the 200 msec prestimulus baseline period. ERPs were time-locked to the onset of the sound in all conditions except the visual-only condition in which case the ERPs were time-locked to the exact moment the ball touched the boundary. Each segmented data set was again manually inspected for excessive artifacts. Once artifact rejection was completed, the EEG data were again filtered, this time using a 30 Hz lowpass (Butterworth) filter and then re-referenced to an average reference. Grand-averaged ERPs were then obtained for each participant by averaging all available epochs for each condition.

The total number of acceptable ERP segments per participant was on average 404.28 trials (SD = 8.08): (audio-only condition: M = 101.38, SD = 2.04; visual-only condition: M = 101.45, SD = 1.97; AV-synchronous condition: M = 100.66, SD = 3.05; AV-asynchronous condition: M = 100.79, SD = 3.02). There were no significant differences between the conditions in the amount of total useable segments included in the construction of each individual ERP response, $F_{(3,\ 28)} = 1.4$, $p = .25$, $\eta^2_p = .05$.

2.5. *ERP regions & components of interest*

To test whether early sensory processing was affected by the temporal synchrony of dynamic AV events, the auditory N1 and P2 components of the auditory evoked potential were evaluated. The N1 component is the first negative going peak of the auditory evoked potential and is thought to index the early sensory processing of auditory stimuli (Godey et al., 2001; Mayhew et al., 2010; Näätänen & Winkler, 1999; Picton et al., 1974; Ponton et al., 2002). The N1 was operationalized here as the minimum peak amplitude and latency occurring within 100–200 msec after sound onset. The auditory P2 component is the second positive going peak of the auditory evoked potential and its functional significance is much less clear compared to the preceding N1. One possible hypothesis posits that the auditory P2 may be involved in matching current sensory input with past perceptual representations (Freunberger et al., 2007; Luck & Hillyard, 1994). The P2 was operationalized here as the maximum peak amplitude and latency occurring within 200–300 msec after sound onset. Both the time window and the regions of interest were selected based on our hypotheses about the timing of each ERP component (Stekelenburg & Vroomen, 2007; Vroomen & Stekelenburg, 2010) and from visual inspection using the grand averaged ERP across all participants and conditions. To quantify early processing of a sound across the entire auditory ERP, we calculated the N1–P2 peak-to-peak amplitude response, which reflects the amplitude change between the negative N1 trough and positive P2 peak. To obtain this value, we subtracted the amplitude of the P2 response from the amplitude of the N1 response for each subject. For our latency analyses, we planned to conduct individual N1 and P2 peak latency measures for both experiments. A six-channel frontal-central auditory region was constructed to evaluate differences in auditory activity between each sensory condition. The ERP data, stimuli, and scripts that support the findings of this study are available to download (Marin et al., 2021a, 2021b). Note that no part of the study's procedures or analysis plan were formally pre-registered before the research was conducted.

3. Results

Fig. 3a presents the grand averaged (n = 29) ERP waveforms, split between sensory conditions.

As can be clearly seen in Fig. 3, all conditions that included a sound elicited auditory-evoked potentials, but – as expected – the visual-only condition did not elicit an auditory response, and was thus dropped from all subsequent analyses.[1] We conducted two separate one-way within-subjects repeated measures ANOVAs with three levels (audio-only, AV-synchronous, AV-asynchronous) for the amplitude and latency responses, within frontal-central scalp regions. All statistical analyses presented below were conducted in R studio, using the ‘tidyverse’ and ‘emmeans’ plugin packages.

3.1. *N1–P2 peak-to-peak amplitude*

As can be seen in Fig. 3a and b, the N1–P2 peak-to-peak amplitude differed between conditions. Statistical analysis of the N1–P2 peak-to-peak amplitude confirmed this observation and revealed a significant main effect of condition, $F_{(2,\ 28)} = 14.2$, $p < .001$, $\eta^2_p = .34$ (see Fig. 3b). Post-hoc tests revealed that the AV-synchronous condition (M = −7.52, SD = 2.83) exhibited a smaller N1–P2 peak-to-peak amplitude that significantly differed compared to the AV-asynchronous (M = −8.44, SD = 3.75; p = .006, d = .28) and the audio-only (M = −9.05, SD = 3.78; p < .001, d = .46) responses. The AV-asynchronous response was not different from the audio-only response (p = .1).

3.2. *N1 and P2 peak latency*

To assess whether the timing of early auditory ERP was affected by the temporal synchrony of dynamic AV events, the N1 and P2 components were evaluated using analyses similar to the N1–P2 peak-to-peak. We predicted that the N1 and P2 responses toward the dynamic AV synchrony should elicit faster peak amplitudes compared AV asynchronous and audio only responses.

[1] The visual-only condition was included in the design of the experiment to keep all conditions symmetrical and not bias participants' expectations in any particular way. Because the visual-only condition contained continuous visual information but no sound, we planned to not look at the EEG data for this condition in any meaningful way.

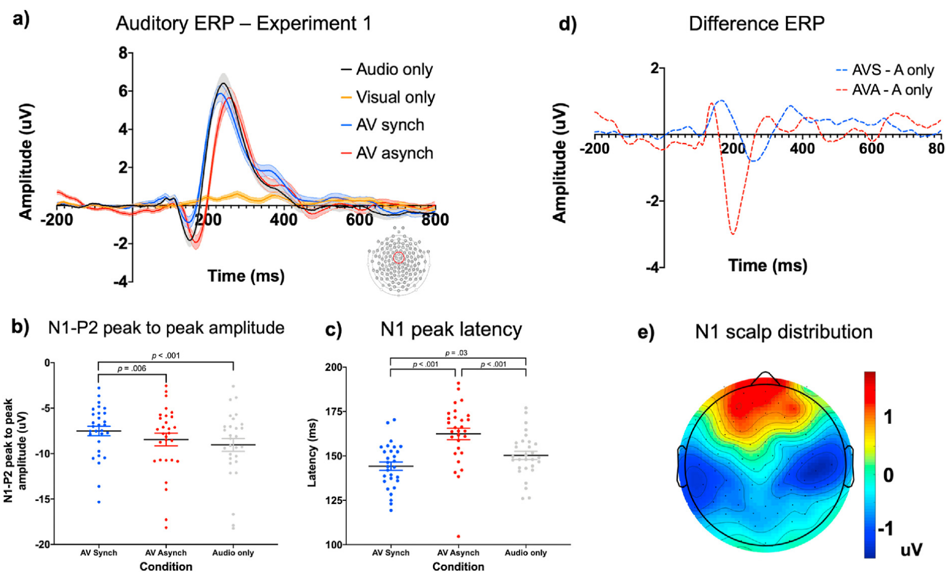

Fig. 3 – Grand averaged ERP and auditory N1–P2 peak-to-peak amplitude and N1 peak latency responses for Experiment 1. Sub-figure (a) presents the frontal-central grand-averaged ERP obtained from a six-channel auditory region of interest shown below the x-axis of the ERP figure. For the ERP figure, the y-axis reflects voltage, which is plotted positive up and the x-axis is the time in milliseconds (msec). Error bars around the ERP reflect the upper and lower bonds of one within-subject standard error of the mean (+/–). Sub-figures (b) and (c) reflect individual scatter plots for the central-frontal N1–P2 peak-to-peak amplitude and N1 peak latency responses, respectively, for 29 adults. The visual-only condition (i.e., sub-figure a, orange trace) was omitted from our main analyses due to the absence of an auditory-evoked potential. The grey color represents the audio alone condition, the blue denotes the AV-synchronous (AVS) condition, and the red is the AV-asynchronous condition (AVA). The error bars in the scatter plot figures reflect the upper and lower bounds of one within-subject standard error of the mean (+/–), and significant *p*-values are provided above each bracket comparison. Sub-figure (d) reflects the grand averaged ERP difference wave of the audio-only response subtracted from the AV-synchronous (blue dash trace) and AV-asynchronous (red dash trace) responses. Sub-figure (e) reflects the voltage distribution (in microvolts; uV) on the scalp of the average N1 activity of the AV-synchronous, AV-asynchronous, and audio only responses. The activity here reflects the mean voltage across the scalp between 100 and 200 msec after the onset of the sound.

Analysis of N1 minimum peak latency revealed a significant main effect of condition, $F_{(2,\ 28)} = 33$, $p < .001$, $\eta_p^2 = .54$ (see Fig. 3c). The N1 peaked earlier for the AV-synchronous condition (M = 144.17 msec, SD = 12.7) compared to AV-asynchronous (M = 162.39 msec, SD = 17.2; $p < .001$, d = 1.21) and audio-only (M = 150.3 msec, SD = 12.4; p = .03, d = .49) responses. Additionally, the N1 was slower for the AV-asynchronous response compared to the audio-only response ($p < .001$, d = .81).

Analysis of P2 maximum peak latency also revealed a significant main effect of condition, $F_{(2,\ 28)} = 31.98$, $p < .001$, $\eta_p^2 = .53$ (not depicted). Like the N1, the P2 peaked sooner for the AV-synchronous condition (M = 234.32 msec, SD = 14.8) compared to AV-asynchronous (M = 255.33 msec, SD = 13.2; $p < .001$, d = 1.5) and audio-only (M = 242.12 msec, SD = 17.4; p = .01, d = .48) responses. Additionally, the P2 was slower for the AV-asynchronous response compared to the audio-only response ($p < .001$, d = .86).

3.3. *Summary and discussion of Experiment 1*

We found that the auditory response was sensitive to the temporal relationship between dynamic AV input for Experiment 1. The auditory response was smaller in amplitude and occurred earlier in time when visual input was synchronously paired in time and space with an expected sound, compared to the response elicited from auditory-alone and asynchronous AV inputs. Additionally, the neural response toward asynchronous AV input was significantly delayed compared to AV-synchronous and auditory-alone presentations. Importantly, smaller auditory responses were seen even when the synchrony of the AV collision event varied unpredictably from trial to trial – a key distinction from Vroomen and Stekelenburg (2010). This pattern of results suggests that early sensitivity toward the temporal synchrony of anticipated sounds allows the brain to code for temporally congruent AV events, resulting in an attenuated (or suppressed) early auditory response. Critically, Experiment 1 underscores the role of temporal synchrony in facilitating early auditory processing, providing further evidence that the auditory effects are relevant for non-predictable inanimate objects, not just expected human actions and AV speech perception (additional theoretical implications are included in the general discussion). Taken together, the results of Experiment 1 suggest that the continuous presentation of a moving object can alter the processing of incoming auditory information within the first 200 msec of processing.

4. Experiment 2

Experiment 1 showed that a moving visual object can alter early auditory processing of a subsequent sound that is perceived as part of the same object. One interpretation of the results of Experiment 1 is that the auditory effects occurred because the sound and visual object were present at the same time during the collision event – which would be consistent with a multisensory account of sensory facilitation. An alternative is, however, that the expectation of a sound induced by a single moving visual stimulus is sufficient to alter auditory processing, even if no visual object is present at the same time the sound occurs. Thus, in Experiment 2, we tested whether continuous visual input is necessary to generate the auditory effects found during temporally synchronous AV presentations, or whether the expectation about a moving object is sufficient to modulate early auditory processing. To do this, we added a new AV condition in which we showed the visual and motion cues provided by the ball and its motion, and then removed these cues via occlusion well before the object collided with an artificial boundary and subsequently elicited an expected bounce sound. Thus, the sound appeared at the moment the ball would collide with the boundary, only the ball was not visible to participants anymore. We compared this condition to the AV synchronous, AV-asynchronous, and audio-only conditions identical to those in Experiment 1. With this new AV-occluded condition, we hoped to elicit similar visual expectancies as in the AV-synchronous condition, but to eliminate the simultaneous presentation of visual object and sound during the collision itself, to tease apart effects of expectation alone, and multisensory integration.

We expected to replicate the auditory effects seen in Experiment 1 for AV-synchronous relative to AV-asynchronous and audio-only conditions. Of particular interest was the AV-occluded condition: If the AV-occluded auditory response was most similar to audio-only activity, it would imply that temporally concordant AV input is important to elicit the auditory effects, consistent with multisensory integration. Alternatively, if the AV-occluded response looks similar to the AV-synchronous response, it would suggest that visually-induced expectations alone are sufficient to alter early auditory processing. Lastly, if the AV-occluded condition resembled the AV-asynchronous response profile, this would suggest that both conditions elicit responses possibly related to detecting AV incongruencies.

5. Methods

5.1. *Participants*

Due to relatively large effect sizes in Experiment 1, we reduced the sample in Experiment 2 to match that used by Vroomen and Stekelenburg (2010). Nineteen college-aged adults (M_{age} = 20.51 years, *SD* = 1.46; 9 female) participated in Experiment 2. An additional five adults were tested but were excluded due to excessive EEG artifact based on the removal criteria outlined in the methods section of Experiment 1.

5.2. *Audio-visual stimuli*

The AV stimuli used in this experiment were the same as in Experiment 1, with the exception of the added AV-occluded condition described below. The new AV-occluded condition contained the same AV properties as the AV-synchronous condition. We did not include the visual-only condition in this experiment. Like in Experiment 1, the AV-asynchronous tone occurred 450 msec before contacting the opposite edge.

5.3. *Paradigm and procedure*

Each participant was seated in a dark room and was shown a randomly presented stream of four AV sensory conditions: 1) audio-only, 2) AV-synchronous, 3) AV-asynchronous, and 4) AV-occluded while high-density EEG was recorded. For the new AV-occluded condition (see Fig. 4 for a single trial stimulus presentation of the timing of events), a single red ball randomly appeared on either the far left or right boundary of a black rectangle display, at the horizontal midline of the monitor.

The ball in this condition traversed along the horizontal midline, but at approximately 600 msec, it began to move through an invisible slit midway in the display and became fully occluded before contacting the opposite boundary. For this condition, the bounce-sound occurred at 1200 msec, exactly when the occluded ball would contact the opposite boundary of the rectangle display. Thus, the timing of the sound was predictable based on when the object entered the occluding area, but the visual object itself was not visible when the bouncing sound was played. After auditory onset, the invisible ball started to move back toward its origin (still occluded at this point) and became fully visible half-way through the display (after another 600 msec), then the stimuli subsequently terminated at 2000 msec. Additionally, all participants again performed the secondary task in between trials and were above a 95% accurate in task, resulting in no subjects removed due to poor attention.

5.4. *EEG data processing*

The EEG/ERP pre-processing steps were the same as Experiment 1. The total number of acceptable ERP segments per participant was on average 404 trials (*SD* = 11.6): (audio-only condition: *M* = 101.74, *SD* = 2.28; AV-synchronous condition: *M* = 100.79, *SD* = 3.33; AV-asynchronous condition: *M* = 100.79, *SD* = 3.19; AV-occluded condition: *M* = 100.68, *SD* = 3.93).

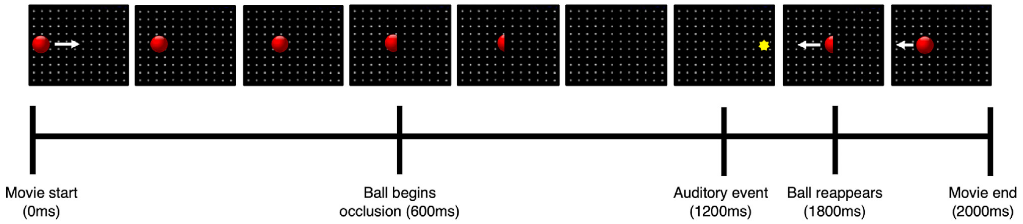

Fig. 4 – Depiction of the AV-occluded condition for Experiment 2 (right start not pictured). In this condition, a red ball appeared on the left or right side of the display and began to move to the opposite boundary. The ball began to occlude behind an invisible slit in the display when it approached the half-way point (600 msec). By the time the ball reached the opposite boundary, it was invisible, but a sound was presented that contained the same temporal characteristics as the audio-visual synchronous presentation. ERPs for this condition were time-locked to the frame labeled "Auditory event."

There were no condition differences regarding the amount of total useable segments included in the construction of each individual ERP response, $F_{(3,\ 18)} = 1.7$, $p = .18$, $\eta_p^2 = .09$.

6. Results

Fig. 5a displays the grand averaged (n = 19) auditory ERP waveforms, split between the four sensory conditions.

As seen in the grand averaged ERP for Experiment 2 (see Fig. 5a), ERP deflections between each condition were seen as early as the onset of the sound (0 msec) and by using N1–P2 peak-to-peak measures, we hope to account for these early ERP differences. The N1 and P2 peak latency analyses for Experiment 2 were conducted in the same manner as the latency analyses for Experiment 1 because the early visual amplitude drift seen between the conditions would not influence the interpretation of timing on the auditory ERP. We then conducted a within-subjects repeated measures ANOVAs with 4 levels (audio-only, AV-synchronous, AV-asynchronous, AV-occluded) for the N1–P2 peak-to-peak amplitude response and N1 and P2 peak latency responses within frontal-central scalp regions.

6.1. *N1–P2 peak-to-peak amplitude*

As shown in Fig. 5a, the auditory component differed in terms of amplitude and latency across the four conditions. Statistical analysis of N1–P2 peak-to-peak amplitude revealed a significant main effect of condition, $F_{(3,\ 18)} = 8.73$, $p < .001$, $\eta_p^2 = .33$ (see Fig. 5b). Post-hoc tests revealed that the AV-synchronous response (M = −5.79, SD = 2.09) exhibited an attenuated N1–P2 peak-to-peak amplitude compared to the AV-asynchronous (M = −7.18, SD = 2.77; p = .004, d = .57) and audio-only (M = −7.32, SD = 2.36; p = .001, d = .69) responses, while the AV-asynchronous response was not statistically different from the audio-only response (p = .98), overall replicating Experiment 1. Importantly, the AV-occluded response (M = −5.91, SD = 1.67 was significantly smaller compared to the AV-asynchronous (p = .01, d = .56) and audio-only (p = .003, d = .69) responses. The AV-occluded response was not different compared to the AV-synchronous (p = .99) response.

6.2. *N1 and P2 peak latency*

Analysis of N1 minimum peak latency revealed a non-significant main effect of condition ($F_{(3,\ 18)} = 2.08$, p = .11, $\eta_p^2 = .1$), diverging from Experiment 1. However, analysis of P2 maximum peak latency revealed a significant main effect of condition ($F_{(3,\ 18)} = 68.9$, $p < .001$, $\eta_p^2 = .79$; see Fig. 5c). The P2 was significantly delayed for the AV-asynchronous response (M = 244.8 msec, SD = 11.31) compared to the AV-synchronous (M = 221.49 msec, SD = 9.99; p < .001, d = 2.18), audio-only (M = 228.16 msec, SD = 10.78; p < .001, d = 1.51), and AV-occluded (M = 225.18 msec, SD = 10.68; p < .001, d = 1.78) responses. The P2 for the AV-synchronous response also peaked sooner compared to the audio-only response (p = .002, d = .64). All other comparisons failed to reach statistical significance (all p's > .17).

6.3. *Summary and discussion of Experiment 2*

The amplitude and latency effects seen in Experiment 1 were replicated in Experiment 2, providing further evidence that the early auditory response is sensitive to the temporal relationship between dynamic AV input. However, the effects of latency were not present at the N1, but were only seen at the P2 for Experiment 2. While we want to be careful in interpreting the N1 latency effects for Experiment 1, the perception of AV asynchrony, on average, delays the auditory response relative to each condition for both Experiments 1 and 2. Importantly, the partial replication of the effects of latency suggests they are overall less robust compared to the amplitude effects.

Of particular interest in Experiment 2 was the response pattern of the AV-occluded condition. We found that the auditory N1–P2 peak-to-peak amplitude response elicited during the AV-occluded condition was smaller compared to unimodal auditory and temporally asynchronous AV inputs, and closely mimicked the AV-synchronous response in amplitude and latency. These findings suggest that early auditory sensitivity toward the expectation of sounds can arise as the result of preceding visual input, without simultaneous audio and visual input. The AV-occluded P2 response also revealed a significant difference in speeded latency compared to the AV-asynchronous response. Importantly, the overall pattern of results suggests that the AV-occluded response closely resembled

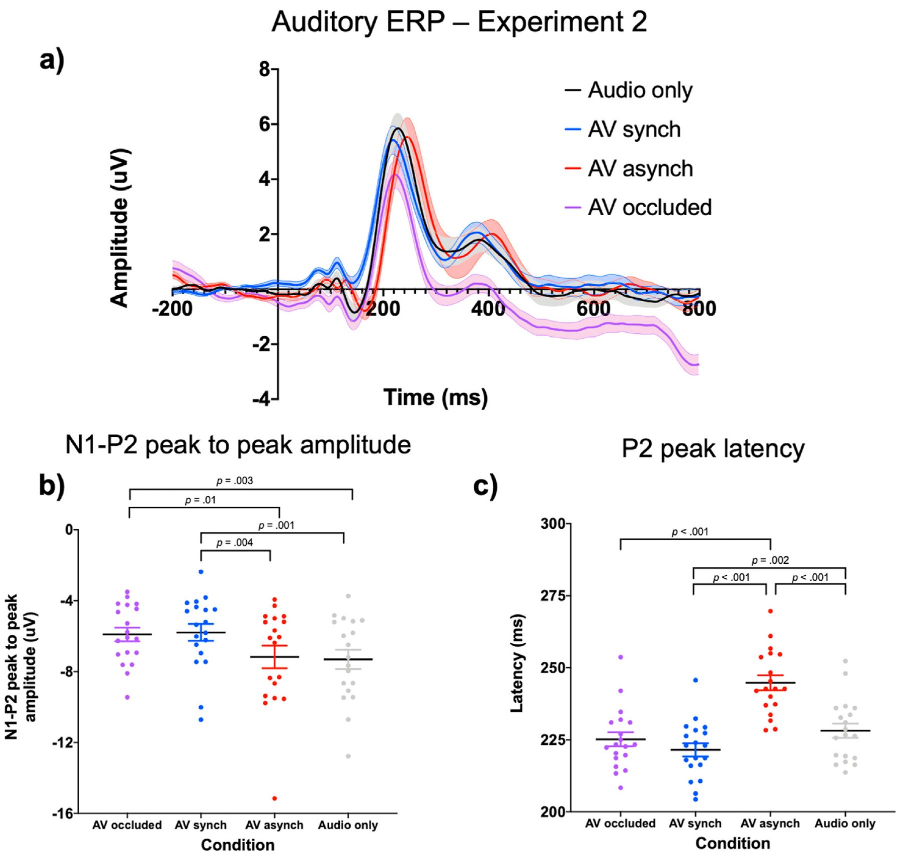

Fig. 5 – Grand averaged ERP, auditory N1–P2 peak-to-peak amplitude response, and P2 peak latency response for Experiment 2. Subfigure (a) presents the frontal-central grand-averaged ERP and individual scatter plots for (b) N1–P2 peak-to-peak amplitude (c) P2 peak latency responses for 19 adults. N1 peak latency responses did significantly differ between conditions. For the ERP figure, the y-axis reflects voltage, which is plotted positive up and the x-axis is the time in milliseconds (msec). Error bars around the ERP reflect the upper and lower bonds of one within-subject standard error of the mean (+/−). The grey color represents the audio alone condition, the blue denotes the AV-synchronous condition, the red is the AV-asynchronous condition, and the purple is the AV-occluded condition. The error bars in the scatter plot figures reflect the upper and lower bounds of one within-subject standard error of the mean (+/−), and significant *p*-values are provided above each bracket comparison.

the activity in the AV-synchronous condition. Overall, Experiment 2 demonstrated that visual expectation induced by a single moving object can facilitate early auditory processing, as most clearly indexed by the overall reduction of the auditory ERP.

7. General discussion

7.1. *Early auditory processing is attenuated for synchronous audio-visual events*

The goal of this study was to examine the electrophysiological correlates of dynamic AV temporal synchrony in the healthy adult brain. For Experiment 1, we were guided by the hypothesis that subtle manipulations of the temporal synchrony underlying dynamic AV events would result in unique patterns of neural responses underlying the auditory ERP, specifically the early auditory response. We found clear evidence that dynamic AV stimulation that differed in temporal onset synchrony subsequently altered the early sensory response (<200 msec) to sounds in fundamentally different ways. Specifically, the early auditory response to temporally synchronous AV events resulted in a pattern of reduced auditory processing (i.e., lower amplitude, faster peak latency) compared to discordant AV stimulation. A second experiment was conducted to assess early auditory responses toward visually occluded but temporally synchronous auditory input. We found that the N1–P2 peak-to-peak response toward AV-synchrony was similar to AV input that contained temporally synchronous, but visually occluded auditory input. These early sensitivities toward the temporal alignment of dynamic AV input demonstrate that general auditory processing is shaped by the temporal expectancies triggered by preceding dynamic visual input.

Our analyses for Experiment 1 revealed both an attenuated and accelerated auditory response when processing dynamic and temporally congruent AV inputs. These changes in the auditory evoked potential can be interpreted as very early auditory (<200 msec) processing being reduced during congruent AV conditions relative to incongruent AV (or audio-only) conditions, possibly indicating that participants coded the temporal synchrony of an expected sound generated by a moving stimulus very early in the auditory processing stream. Additionally, the AV-asynchronous response, where the sound occurred before it was expected, elicited a delayed auditory response compared to the response elicited from unimodal auditory events, likely reflecting a signature of detecting sensory conflict between the timing of the auditory and visual stimuli. The findings of early changes of the auditory evoked potential – amplitude reduction and shorter latency – for temporally congruent, dynamic AV stimuli provide more evidence for the idea that similar amplitude reductions (i.e., suppression) of the auditory response arises in scenarios in which preceding sensory input matches additional, yet expected sensory input (Clementz et al., 2002; D'Andrea-Penna et al., 2020; Ford et al., 2007; Ford & Hillyard, 1981; Kononowicz & van Rijn, 2014; Lange, 2009, 2013; Menceloglu et al., 2020; Schafer et al., 1981; Stekelenburg & Vroomen, 2007; van Wassenhove et al., 2005; Vroomen & Stekelenburg, 2010). Note that no direct comparisons were made between visual-only presentations and the three other sensory conditions, due to the lack of an observed auditory ERP in the visual-only condition. Given that our visual stimulus was continuous rather than a discrete event, we were also not able to observe a clear visually-evoked response of the ball bouncing; however, our data clearly showed that the visual stimulus modulated auditory processing. We recognize that there are potentially important differences between the AV-synchronous and AV-asynchronous conditions. First, the visual position of the object during sound onset is different between the AV-synchronous and AV-asynchronous conditions, where the ball is closer to the origin for the AV-asynchronous condition, which makes a direct comparison of their baselines difficult. Second, the reversal of the object after the collision in the synchronous condition may have provided the observer additional visual information that was not seen during the AV-asynchronous condition. Because of these important, yet unavoidable sensory differences between the AV-synchronous and AV-asynchronous conditions, some degree of caution is needed when interpreting these results.

The present results deviate in some ways from the findings by Vroomen and Stekelenburg (2010). Specifically, Vroomen & Stekelenburg observed differences in the auditory response only when the visual input predicted the sound with high reliability (fixed blocks only; three condition comparison: audio-only, visual-only, and AV-synchronous expectations; their Experiment 1), but when the relation became less reliable (mixed vs. fixed blocks; six condition comparison: audio-only, visual-only, AV-synchronous expectation, AV-non expectation, and early and late AV-asynchronous sound onset; their Experiment 2), these early auditory modulations only appeared during fixed-block, AV-synchronous expectation presentations. With regards to the later auditory response (~240 msec), they found an equally suppressed response toward synchronous AV and early asynchronous AV input regardless of the predictability of the audio-visual events (i.e., both in fixed and intermixed blocks). We, on the other hand, found that the auditory response was sensitive to AV asynchronies even when synchronous and asynchronous trial types were randomly intermixed, and the visual input thus did not reliably predict the timing of a sound across trials. The later auditory response appeared to show, on average, a general sensitivity to audio-visual inputs regardless of temporal synchrony, similar to Vroomen and Stekelenburg (2010). However, since we did not have clear *a priori* expectations with regards to the later auditory response in our tasks, we hesitate to strongly interpret these changes in the later auditory response.

Why do our results differ from those observed in the previous study? In the current set of experiments, a visual object provided expectations about upcoming auditory input in a single, uniform direction. In Vroomen and Stekelenburg (2010), two visual objects appeared to the left and right of a rectangle and moved toward it, eventually colliding with and bouncing off it. In our experiments, visual expectation, and therefore attention, was not divided between two objects, which may have reduced uncertainty, providing for a more accurate perceptual representation of the temporal relationships underlying the AV inputs. Additionally, the stimuli used in the current experiment were perhaps more reflective of natural sensory environments. For example, we opted to use a red sphere that appeared to move toward and bounce off a single, artificially defined barrier, eliciting a "knock" sound that resembled the sound of a ball bouncing of the wall. Vroomen and Stekelenburg used more simplified stimuli, including two white visual disks that elicited a pure auditory tone upon synchronous contact with an artificially defined barrier. Alternatively, the contrast between the results of our Experiment 1 and Vroomen and Stekelenburg (2010), where we found a difference in the neural response toward synchronous and asynchronous AV inputs during mixed trials while they did not, may have arisen due to differences in the temporal gap between discordant AV stimulation. In our experiments, the auditory onset for asynchronous AV input occurred 450 msec before visual collision. Vroomen and Stekelenburg presented auditory information 240 msec before visual collision (2010). The auditory response is thought to reflect an early sensory response that is modulated by low-level auditory characteristics like loudness and pitch (Hyde, 1997), and thus in principle could be sensitive to other low-level characteristics like systemically smaller asynchronous temporal gaps in dynamic AV input. We think it is unlikely that these small differences in the temporal onset of a sound paired with dynamic, yet discordant visual input drives the auditory response synchrony differences between the two studies.

We measured responses to AV asynchrony using a single offset in timing between the synchronous AV event and the asynchronous one. However, differences in timing between when a sound is expected based on visual input and when it actually occurs could matter for how reliable the multisensory percept is. Additional studies will also be needed to assess the auditory response while perceiving systematically smaller temporal onsets, or even small delays between discordant AV input. Such research will help further characterize whether this mechanism relies more so on a general sensitivity toward the temporal expectation elicited by a moving object itself versus one that would rely on a unique multisensory interplay between the AV inputs.

7.2. Synchronous visual expectations about the timing of sounds reduces auditory responses

In Experiment 2, we asked whether expectation provided by continuous visual input preceding the sound was sufficient to elicit an attenuated auditory response even in the absence of the visual input continuing to the point of impact. The N1–P2 peak-to-peak response to audio-only and AV-asynchronous inputs were greater compared to sensory information that provided synchronous auditory stimulation but occluded visual information at the point of collision. Additionally, the AV-occluded condition was not statistically different in amplitude or latency from the AV-synchronous response. Thus, the reduced auditory amplitude toward occluded AV stimulation provides evidence for sensitivity toward the expectation of the impending sound in that the brain's response resembled the perception of AV-synchronous input, even without a precise visual representation of the collision event. The human visual system displays a remarkable ability to represent the persistence of dynamic objects that undergo brief visual occlusion (see review, Scholl, 2007). Even six-month old human infants are able to anticipate the exit trajectory of a briefly occluded visual object in motion (Johnson et al., 2003). Additionally, the ability to visually track and identify multiple target objects that undergo brief visual occlusion is unimpaired in normal sighted individuals (Scholl & Pylyshyn, 1999). In this case, tracking a briefly occluded visual object may help reduce the computational demands of auditory brain regions, allowing for the auditory system to better coordinate in time and space the physical properties of the occluded object (i.e., rate of motion, physical boundaries, etc.) with its expected sound. In other words, the suppressed auditory response elicited by occluded and synchronous AV inputs may have resulted from the brain generating successful predictions about when an expected sound would occur. We showed this by simply presenting a dynamic moving object for a relatively short time that contained accurate temporal and spatial information. These visual cues in essence allowed the perceiver to infer the source of the sound even without simultaneous visual input. It is worth noting that the visible condition contained more precise information about when the tone occurs, and thus the dissociation between expectation and integration ought to be interpreted carefully. Nonetheless, Experiment 2 underscores the importance of visual expectations in eliciting auditory suppression, in that the brief representation of a dynamic visual object's spatial and temporal properties led to the expectation of an accompanying sound.

7.3. Summary

Early sensitivity toward the temporal synchrony of dynamic AV events is important for the successful identification of bimodal sensory signals that should be perceived as either unified or two separate sensory events. Sensitivity to the temporal relation between AV input allows the brain to either processes congruent AV events or detect asynchrony very early in the auditory processing stream. Thus, the reduction of the auditory response to AV temporal synchrony may manifest from the brain generating successful predictions about basic sensory events in the environment. In other words, the reduced auditory response seen here may have resulted from a perceptual match between top-down expectancies of a sound and correct bottom-up sensory input, like the ball's motion and its synchronous relation to the timing of the sound itself. Such mechanisms may help lay the groundwork to further understand how neural activity is shaped in later processing stages that involve higher-level cognitive processes like attention and decision making. For example, one might direct less attention toward the low-level features of temporally synchronous AV events, while exerting more effort in extracting contextual information embedded within the AV signal. Conversely, AV input that is temporally asynchronous may evoke disturbances in fundamental mechanisms designed to bind bimodal sensory information as one. These highly specialized sensory processes afford the brain the ability to exhibit sensitivities toward relatively small temporal discrepancies between dynamic AV stimulation. Such mechanisms are crucial, as the successful integration of the expectation of an accompanying sound that arise via dynamic visual stimuli results in precise scene representations. Another interpretation of the data could be that the auditory response expectancy effects resulted from the preactivation of the neural representation of the expected sound contained in the time prior to sound actually occurring (Blom et al., 2020; Kok et al., 2017). This is a very important distinction that can be addressed in future research. For example, sensory manipulations, like varying the speed of the object itself, the time spent behind occlusion, or a combination of both are needed to assess contextual occlusion influences over the visual ERP. The fact that the auditory response was not modified by occlusion but was different from unimodal input suggests that expectation plays some role in processing expected sounds based on the trajectory of a moving visual stimulus. The neural mechanism characterized in this study may be fundamental to proper AV processing more broadly, in that these early auditory responses are sensitive to temporal discrepancies between AV sensory input that differ in milliseconds.

In sum, our results provide evidence for a neural mechanism that is sensitive to the underlying temporal relation of dynamic AV input in healthy adults. Early sensitivities to temporally congruent and discordant events may help determine how the brain subsequently processes bimodal experiences in natural sensory environments. For synchronous events, the brain exhibited a reduced auditory response when the temporal predictions of an ensuing sound were in line with preceding visual input. In contrast, greater neural responses were seen in the presence of AV temporal incongruency. Early sensitivity toward the temporal synchrony of dynamic AV events, as reflected in the early auditory response modulations, may reflect a basic perceptual mechanism used to gage the plausibility of expected sensory events contained within our environment. Importantly, a moving visual object that provides accurate spatial and temporal expectations about when a sound is likely about to appear are sufficient in attenuating early auditory processing. Broadly, this suggests that visual input – or the representation of visual input – leads to more efficient auditory processing within the first few hundred milliseconds of processing. These early influences likely have important consequences for other downstream processes.

Credit author statement

Andrew Marin: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data Curation, Writing – Original draft, Visualization, Project administration. **Viola S. Störmer:** Conceptualization, Methodology, Writing – Review & editing, Supervision. **Leslie J. Carver:** Conceptualization, Methodology, Resources, Writing – Review & editing, Supervision.

Open practices

The study in this article earned Open Data and open Materials badges for transparent practices. Data and materials for this study can be found at http://dx.doi.org/10.17632/k3j772tmwk.2.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability statement

The deidentified ERP data that support the findings of this study are available to download via an online data repository. The cleaned and segmented ERP data (.set/.fdt files), N1 and P2 peak amplitude and latency data sets, all programming scripts, and stimuli used for both experiments are available to download via a publicly available online data repository (https://dx.doi.org/10.17632/k3j772tmwk.4). Raw EEG data sets are also available for download via a separate online data repository (osf.io/d245g/).

Declaration of competing interest

None.

Acknowledgements

We would first like to thank Dr. Denise Werchan and Dr. Dima Amso for their willingness to share their stimuli with us. We would also like to acknowledge Zoë and Cameron Tauxe, and Dr. David Lewkowicz for the help with stimuli creation. Lastly, we would like to thank Jonathan Keefe and Elizabeth Baker for their helpful comments and edits when preparing this manuscript.

REFERENCES

Baldeweg, T. (2007). ERP repetition effects and mismatch negativity generation: A predictive coding perspective. *Journal of Psychophysiology, 21*(3–4), 204–213. https://doi.org/10.1027/0269-8803.21.34.204

Besle, J., Fort, A., Delpuech, C., & Giard, M. H. (2004). Bimodal speech: Early suppressive visual effects in human auditory cortex. *European Journal of Neuroscience, 20*(8), 2225–2234. https://doi.org/10.1111/j.1460-9568.2004.03670.x

Blom, T., Feuerriegel, D., Johnson, P., Bode, S., & Hogendoorn, H. (2020). Predictions drive neural representations of visual events ahead of incoming sensory information. *Proceedings of the National Academy of Sciences, 117*(13), 7510–7515. https://doi.org/10.1073/pnas.1917777117

Clementz, B. A., Barber, S. K., & Dzau, J. R. (2002). Knowledge of stimulus repetition affects the magnitude and spatial distribution of low-frequency event-related brain potentials. *Audiology and Neuro-Otology, 7*(5), 303–314. https://doi.org/10.1159/000064444

D'Andrea-Penna, G. M., Iversen, J. R., Chiba, A. A., Khalil, A. K., & Minces, V. H. (2020). One tap at a time: Correlating sensorimotor synchronization with brain signatures of temporal processing. *Cerebral Cortex Communications, 1*(1), tgaa036. https://doi.org/10.1093/texcom/tgaa036

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

Ford, J. M., & Hillyard, S. A. (1981). Event-related potentials (ERPs) to interruptions of a steady rhythm. *Psychophysiology, 18*(3), 322–330. https://doi.org/10.1111/j.1469-8986.1981.tb03043.x

Ford, J. M., Roach, B. J., Faustman, W. O., & Mathalon, D. H. (2007). Synch before you speak: Auditory hallucinations in schizophrenia. *The American Journal of Psychiatry, 164*(3), 458–466. https://doi.org/10.1176/ajp.2007.164.3.458

Fort, A., Delpuech, C., Pernier, J., & Giard, M. H. (2002). Dynamics of cortico-subcortical cross-modal operations involved in audio-visual object detection in humans. *Cerebral Cortex, 12*(10), 1031–1039. https://doi.org/10.1093/cercor/12.10.1031

Freunberger, R., Klimesch, W., Doppelmayr, M., & Höller, Y. (2007). Visual P2 component is related to theta phase-locking. *Neuroscience Letters, 426*(3), 181–186. https://doi.org/10.1016/j.neulet.2007.08.062

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences, 360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *Journal of Cognitive Neuroscience, 11*(5), 473–490. https://doi.org/10.1162/089892999563544

Godey, B., Schwartz, D., de Graaf, J. B., Chauvel, P., & Liégeois-Chauvel, C. (2001). Neuromagnetic source localization of auditory evoked fields and intracerebral evoked potentials: A comparison of data in the same patients. *Clinical Neurophysiology, 112*(10), 1850–1859. https://doi.org/10.1016/S1388-2457(01)00636-8

Hyde, M. (1997). The N1 response and its applications. *Audiology and Neurotology, 2*(5), 281–307. https://doi.org/10.1159/000259253

Johnson, S. P., Amso, D., & Slemmer, J. A. (2003). Development of object concepts in infancy: Evidence for early learning in an eye-tracking paradigm. *Proceedings of the National Academy of Sciences, 100*(18), 10568–10573. https://doi.org/10.1073/pnas.1630655100

Kok, P., Mostert, P., & de Lange, F. P. (2017). Prior expectations induce prestimulus sensory templates. *Proceedings of the National Academy of Sciences,* 114(39), 10473–10478. https://doi.org/10.1073/pnas.1705652114

Kononowicz, T. W., & van Rijn, H. (2014). Decoupling interval timing and climbing neural activity: A dissociation between CNV and N1P2 amplitudes. *Journal of Neuroscience, 34*(8), 2931–2939. https://doi.org/10.1523/JNEUROSCI.2523-13.2014

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *Plos One, 2*(9), e943. https://doi.org/10.1371/journal.pone.0000943

Lange, K. (2009). Brain correlates of early auditory processing are attenuated by expectations for time and pitch. *Brain and Cognition, 69*(1), 127–137. https://doi.org/10.1016/j.bandc.2008.06.004

Lange, K. (2013). The ups and downs of temporal orienting: A review of auditory temporal orienting studies and a model associating the heterogeneous findings on the auditory N1 with opposite effects of attention and prediction. *Frontiers in Human Neuroscience, 7*. https://doi.org/10.3389/fnhum.2013.00263

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience, 8*. https://doi.org/10.3389/fnhum.2014.00213

Luck, S. J., & Hillyard, S. A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology, 31*(3), 291–308. https://doi.org/10.1111/j.1469-8986.1994.tb02218.x

Marin, A., Störmer, V. S., & Carver, L. J. (2021a). *Expectations about dynamic visual objects facilitates early sensory processing of congruent sounds* (Vol. 4). Mendeley Data. https://doi.org/10.17632/k3j772tmwk.4

Marin, A., Störmer, V. S., & Carver, L. J. (2021b). *Expectations about dynamic visual objects facilitates early sensory processing of congruent sounds (Raw Data)*. Retrieved from osf.io/d245g.

Mayhew, S. D., Dirckx, S. G., Niazy, R. K., Iannetti, G. D., & Wise, R. G. (2010). EEG signatures of auditory activity correlate with simultaneously recorded fMRI responses in humans. *NeuroImage,* 49(1), 849–864. https://doi.org/10.1016/j.neuroimage.2009.06.080

Menceloglu, M., Grabowecky, M., & Suzuki, S. (2020). Rhythm violation enhances auditory-evoked responses to the extent of overriding sensory adaptation in passive listening. *Journal of Cognitive Neuroscience, 32*(9), 1654–1671. https://doi.org/10.1162/jocn_a_01578

Meredith, M., Nemitz, J., & Stein, B. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *The Journal of Neuroscience, 7*(10), 3215–3229. https://doi.org/10.1523/JNEUROSCI.07-10-03215.1987

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory–visual interactions during early sensory processing in humans: A high-density electrical mapping study. *Cognitive Brain Research, 14*(1), 115–128. https://doi.org/10.1016/S0926-6410(02)00066-6

Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin, 125*(6), 826. https://doi.org/10.1037/0033-2909.125.6.826

Picton, T. W., Hillyard, S. A., Krausz, H. I., & Galambos, R. (1974). Human auditory evoked potentials. I: Evaluation of components. *Electroencephalography and Clinical Neurophysiology,* 36, 179–190. https://doi.org/10.1016/0013-4694(74)90155-2

Ponton, C., Eggermont, J. J., Khosla, D., Kwong, B., & Don, M. (2002). Maturation of human central auditory system activity: Separating auditory evoked potentials by dipole source modeling. *Clinical Neurophysiology, 113*(3), 407–420. https://doi.org/10.1016/S1388-2457(01)00733-7

Psychology software Tools, Inc. [E-Prime 2.08.90]. Retrieved from https://www.pstnet.com, (2016).

Schafer, E. W. P., Amochaev, A., & Russell, M. J. (1981). Knowledge of stimulus timing attenuates human evoked cortical potentials. *Electroencephalography and Clinical Neurophysiology,* 52(1), 9–17. https://doi.org/10.1016/0013-4694(81)90183-8

Scholl, B. J. (2007). Object persistence in philosophy and psychology. *Mind & Language, 22*(5), 563–591. https://doi.org/10.1111/j.1468-0017.2007.00321.x

Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology, 38*(2), 259–290. https://doi.org/10.1006/cogp.1998.0698

Spence, C. (2007). Audiovisual multisensory integration. *Acoustical Science and Technology,* 28(2), 61–70. https://doi.org/10.1250/ast.28.61

Stekelenburg, J. J., & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *Journal of Cognitive Neuroscience,* 19(12), 1964–1973. https://doi.org/10.1162/jocn.2007.19.12.1964

Stevenson, R. A., Fister, J. K., Barnett, Z. P., Nidiffer, A. R., & Wallace, M. T. (2012). Interactions between the spatial and temporal stimulus factors that influence multisensory integration in human performance. *Experimental Brain Research, 219*(1), 121–137. https://doi.org/10.1007/s00221-012-3072-1

Stevenson, R. A., Zemtsov, R. K., & Wallace, M. T. (2012). Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions. *Journal of Experimental Psychology. Human Perception and Performance,* 38(6), 1517–1529. https://doi.org/10.1037/a0027339

Vroomen, J., & Stekelenburg, J. J. (2010). Visual anticipatory information modulates multisensory interactions of artificial audiovisual stimuli. *Journal of Cognitive Neuroscience, 22*(7), 1583–1596. https://doi.org/10.1162/jocn.2009.21308

Wallace, M. T., & Stevenson, R. A. (2014). The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. *Neuropsychologia, 64*, 105–123. https://doi.org/10.1016/j.neuropsychologia.2014.08.005

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences,* 102(4), 1181–1186. https://doi.org/10.1073/pnas.0408949102

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia, 45*(3), 598–607. https://doi.org/10.1016/j.neuropsychologia.2006.01.001

Werchan, D. M., Baumgartner, H. A., Lewkowicz, D. J., & Amso, D. (2018). The origins of cortical multisensory dynamics: Evidence from human infants. *Developmental Cognitive Neuroscience,* 34, 75–81. https://doi.org/10.1016/j.dcn.2018.07.002

Chapter 1, in full, is a reprint of the material as it appears in Expectations about dynamic visual objects facilitates early sensory processing of congruent sounds in *Cortex*. Marin, Andrew; Störmer, Viola S.; Carver, Leslie J. (2021). The dissertation author was the primary investigator author of this paper.

Electrophysiological Differences Underlying the Perception of Auditory Error in Autism

Andrew Marin[1], Lucy Pearson[1,2], Mincong Wu[1], Elizabeth Baker[3], & Leslie J. Carver[1]

1) University of California, San Diego - Psychology Department

2) Cardiff University - Psychology Department

3) University of California, Riverside - School of Education

Author Note

Correspondence concerning this article should be addressed to Andrew Marin, Department of Psychology, 9500 Gilman Drive, McGill Hall, La Jolla, CA, 92093, USA. Email: amarin@ucsd.edu

Acknowledgments

We would like to thank Melaine Dratva for a thoughtful edit of this manuscript. We would like to thank Kayce Padilla and Sophia Bokovikova for coordinating this study, and also Danna Wu for their help with data collection and data management.

Abstract

Predictive coding accounts of Autism Spectrum Disorder (ASD) suggest that autistic individuals display differences in predicting sensory information. Previous research has noted neural differences in auditory predictions and error integration in autistic individuals, but which signals are most impacted remains unclear. Here, we recorded auditory event-related potentials (ERPs) from autistic and neurotypical (NT) adults who passively observed a ball that made a bounce sound upon physical collision. We compared auditory ERPs to conditions where the sound of the ball either occurred: 1) in synchrony with visual collision, 2) asynchronously before collision, 3) during synchronous, but occluded collision, or 4) the sound occurred in isolation. Across all participants, the asynchronous condition elicited greater auditory responses compared to synchronous and obscured inputs, while the activity in the obscured condition closely resembled that of the synchronous condition. We also found the differential P2 amplitude toward asynchrony in autism was greater relative to NT, while no differences in response to fully visible or occluded synchrony were observed. Exploratory brain-behavioral analyses revealed that greater neural responses corresponding to the perception of asynchronous sound was related to a decrease in self reported autistic traits. These results suggest that increased neural responses to basic perceptual errors in autism may result from differences in bottom-up sensory processing, rather than the integration of top-down expectations.

Keywords: Autism, ASD, auditory ERP, audio-visual (AV), predictive coding

Electrophysiological Differences Underlying the Perception of Auditory Error in Autism

Imagine yourself at a grocery store. Employees greet you with a smile and wave. Old, rickety carts get louder as people whiz by you. All of a sudden, a customer drops a glass jar from the shelf, prompting you to cover your ears in anticipation of it crashing to the floor. These dynamic visual experiences, occurring in both social and non-social settings, provide visual cues about when to anticipate certain sounds. Processing cross-modal events depends on the successful integration of unisensory signals entering the brain, which equip us to process and respond to naturalistic sensory information. As a result, differences in how the brain anticipates sound from vision may stem from disruptions in integrative sensory systems. Characterizing the mechanisms in neurodevelopmental populations who experience diffuse alterations in these neural systems, coupled with documented behavioral issues related to predictive processing, is one way to gain insight into how we neurologically predict sounds based on visual information. To these individuals, walking through a grocery store could be a challenging sensory experience, a sentiment commonly shared by autistic people (MacLennan et al., 2023).

Autism Spectrum Disorder (ASD) is a heterogeneous neurodevelopmental condition with a diverse range of individual symptoms characterized by differences in social communication and the presence of restrictive and repetitive behaviors (American Psychiatric Association, 2022). Though sensory processing differences are well-documented, clinicians have traditionally emphasized autism as a primarily socially orientated disorder, involving challenges in processing socially motivated stimuli and understanding others' behaviors and mental states (for review, Baron-Cohen, 2000; Chevallier et al., 2012; Klin et al., 2002). Contemporary research has tended to compartmentalize sensory and social features when studying autism, despite findings that the severity of sensory sensitivities is related to the severity of social-cognitive symptoms (Hilton et al., 2010; Thye et. al, 2018; Tavassoli et al., 2018; Zhai et al., 2023). Although empirical and clinical evidence suggests that autism is a disorder largely defined by social differences, there remains a gap between the documented differences in social cognition and the role of sensory mechanisms that may underlie social functioning.

Predictive Coding Theory (PCT) of autism is a domain-general account of autistic symptomatology, suggesting the heterogeneous symptom profile is defined by fundamental disruptions in utilizing internal predictive models and effective error processing (Cannon et al., 2021; Sinha et al., 2014). PCT outlines that the brain's principal function is to build internal mental models, or “priors,” to optimize efficient bottom-up processing of external information (Friston, 2005; Jiang & Rao, 2021). PCT argues that top-down expectations are compared to bottom-up sensory signals (Huang & Rao, 2011; Huang, 2018; Choi, 2018; Heeger, 2017), where discrepancies between what we expect and what we perceive trigger prediction errors, or a ‘surprise’ (Feldman & Friston, 2010), which in turn facilitates attention towards updating mental models (Picard and Friston, 2014; Sinclair et al., 2021). In sum, subjective inference of our environment is dependent on the delicate balance between integrating prediction errors with current mental models, or experience-dependent learning (Friston, 2012). Thus, PCT is emerging as a hypothesis to understand the cortical processing implicated in predicting real-world expectations that can be quantified at a mechanistic level.

Neurocomputational models such as PCT have been posited as an explanatory framework to understand the neural mechanisms underlying major neurodevelopmental and neuropsychiatric disorders (Huys et al., 2021; Huys et al., 2016; Kaye and Krystal, 2020). Dysfunctions in the regulation of predictive neural systems are thought to characterize differences in perceptual, attentional, or reasoning deficits, which appear to map onto clinical symptoms (Smith et al., 2021; van Schalkwyk et al., 2017). In applying the PCT framework to autism, it is commonly accepted that differences in predictive circuitry hinder the optimization of top-down prediction models, leading to a heightened reliance on sensory information. Originating from an inability to flexibly process prediction errors, core deficits appear to be consequences of a disinhibited system that may formulate an endophenotype shared across autistic individuals (Van Boxtel & Lu, 2013; Greene, 2019). However, the mechanisms underlying these predictive differences in autism remain a topic of ongoing debate.

Pellicano and Burr (2012) first promoted the view that autism etiology is primarily rooted in attenuated top-down signals (i.e., 'hypo-priors'). They posit distinct differences in the precision of priors, in which autistic people occupy 'weaker' mental models in comparison to neurotypical (NT) people. Alternatively, Van De Cruys et al. (2014) and Lawson et al. (2014) suggest that core symptomology in autism is underpinned by the inability to flexibly process bottom-up prediction errors, where errors are weighed higher than pre-existing priors. These theories converge in their argument that individuals on the spectrum have a bias towards overemphasizing bottom-up sensory information, such that the formation of mental models is compromised. However, the contribution of top-down and bottom-up signaling in shaping predictive differences is an open question. To address this, we will directly compare the mechanisms that support the integration of top-down expectations versus those that signal bottom-up prediction error in autistic participants.

Questions regarding the neural basis of stimulus prediction have particularly garnered a lot of interest within the auditory domain (for review, Lange, 2013). Event-related potentials (ERPs), which represent the averaged electroencephalogram (EEG) response time-locked to a particular event, have been employed to assess how early auditory responses are shaped by predictable sensory information. The auditory evoked potential (i.e., the N1–P2 complex) is an early sensory response that is modulated by predictable auditory input. For example, several studies have reported that the auditory response is attenuated when hearing temporally predictable sounds (Clementz et al., 2002; D'Andrea-Penna et al., 2020; Ford et al., 2007; Ford & Hillyard, 1981; Kononowicz & van Rijn, 2014; Lange, 2009; Menceloglu et al., 2020; Schafer et al., 1981). Within a predictive coding framework (Friston, 2005; Lange, 2013), auditory response reduction toward expected sounds is thought to be reflective of top-down expectations. The synchronous match between top-down and bottom-up signals is hypothesized to reduce the error signal of the predictive sound and thus attenuate the evoked ERP (Baldeweg, 2007; Lange, 2013). Moreover, sounds that violate auditory expectations result in an increased amplitude and latency in auditory responses (for review, Näätänen et al., 2007). Therefore, the auditory ERP serves as an electrophysiological marker for prediction and error signals, displaying reduced amplitudes for prediction and an amplified response for error.

Of particular interest are two studies by van Laarhoven et al., which utilized ERPs to quantify neural markers of motor-auditory prediction (2019) and auditory omission error (2020) in autistic adults. In the study assessing motor-auditory predictions, researchers analyzed how the autistic brain predicts sounds generated by movements, using the N1 component as a measure for these internal predictions. Participants were exposed to three conditions: 1) a motor-auditory condition where each participant pressed a button that resulted in a sound, 2) an audio-only condition where participants only heard the sound, and 3) a motor-only condition where each participant made a button press, but no sound was heard. Here, the neural response to self-initiated sounds was reduced compared to externally produced sounds in NT, but was absent in the autistic group. van Laarhoven et al. suggested that the motor-to-auditory predictions may be compromised in autistic individuals, subsequently limiting the ability to predict the consequences of their self-initiated movements. The inability to anticipate upcoming sensory events, especially if self-generated, provides evidence for differences in top-down predictive models in autism.

Extending this work, a further study by van Laarhoven examined autistic neural responses to error while observing expected and unexpected actions of others. Here, ERPs were recorded while adult participants passively viewed video recordings of hand claps, including trials where visual input was predictive of the expected sound and trials that contained unexpected omissions of the auditory stimulus. When the auditory stimulus was omitted, autistic participants demonstrated a larger error response compared to NT, indicating that autistic participants exhibited an increased response when top-down predictions were violated. The presence of increased sensory signaling quantified by the omission study suggests the presence of bottom-up sensory-driven predictive differences in autism.

These two findings demonstrate electrophysiological support for top-down (van Laarhoven et al., 2019) and bottom-up (van Laarhoven et al., 2020) predictive coding differences in autism, evidenced specifically within the motor-audio and audio-visual (AV) domains. However, neither of these studies distinguished between auditory signals that represent processing sound based on expected top-down information from the same sound causing bottom-up errors. The motor-auditory prediction study lacked a comparison between predictable sounds and those that violated expectations. Similarly, the handclap paradigm quantified the error signal as the absence of auditory stimulation, lacking a direct comparison of predictable sound versus sound that induces error. Despite the primary proposition of the PCT framework being that perceptual processing involves the integration of prediction errors, the differences between responses to predicted sounds and sounds that elicit error have not, to our knowledge, yet been measured within autistic participants. To advance PCT from a computational theory to a domain-general sensory processing model, dissecting the interplay between external information and prior expectations is crucial to understanding predictive differences in autism.

Extending the work of van Laarhoven et al. and taking advantage of the utility of ERP markers of prediction in the AV domain, we will use a paradigm that directly compares auditory prediction and error signaling as two distinct mechanisms of predictive auditory processing. As discussed earlier, auditory stimulation can become predictable if anticipatory visual information provides congruent spatiotemporal cues about the timing and content of the sound. For example, previous research has shown that a moving visual object, like a ball bouncing off a physical barrier, can reduce early auditory responses to congruent bounce sounds in NT adults (Marin et al., 2021; Vroomen & Stekelenburg, 2010). Bounce sounds that violate the spatiotemporal expectations of sound (i.e., sound before collision) elicit enlarged auditory responses that reflect the perception of error (Marin et al., 2021). Moreover, the reduction of the auditory ERP occurs without visual input as observed in an occluded AV condition, where a ball slipped behind an invisible occluder before synchronous collision. We propose to use these stimuli to directly compare neural responses related to the integration of basic auditory predictions versus mechanisms related to error perception in autistic adults, which to our knowledge, is the first study to do so.

Our aim is to enhance our understanding of the mechanisms that underlie predictive differences in autism, as outlined by the PCT framework. In previous work employing these stimuli in NT adults, we found that fully visible and occluded synchrony exhibited a reduced auditory response relative to asynchronous collision, and audio alone inputs – a pattern we expected to replicate here. Note that directional hypotheses regarding group-level differences in auditory ERP prediction and error signaling between autistic and NT participants were not made prior to data collection. Group-level differences in synchronous and occluded AV processing could signal differences in the integration of top-down expectations signals in autism, as found in van Laarhoven et al. (2019). Alternatively, individuals within the autistic group may exhibit differences in error signaling compared to NT, as found in van Laarhoven et al., (2020). We may also find that both signals are affected in autism compared to NT controls. Null effects between autistic and NT participants could indicate no evidence of group-level differences in processing basic auditory predictions. We will also conduct exploratory analyses to assess the relation between neural signatures of auditory prediction and the presence of autistic traits. By directly contrasting predictive and error signaling, we will be better equipped to explore fundamental questions regarding the nature of mechanisms that drive predictive coding differences in autism.

Methods

Participants

Our sample included 20 autistic and 20 NT participants. Each group included six males and 14 females, and the average age of the autistic (M_{years} = 21.8, SD_{years} = 3.7) and NT (M_{years} = 21.6; SD_{years} = 2.6) groups did not differ at the time of EEG recording (p = 0.58). All NT participants self-reported: normal or corrected-to-normal vision, normal hearing, and no history of neuropsychological, cognitive, or developmental disorders. NT participants were all college students who received course credit for participation and were recruited through an online participant pool. The autistic group was primarily recruited via flyers on college campuses in the San Diego area, or through an online participant pool. Autistic participants had the option to receive course credit or a $30 gift voucher. Of the 20 autistic participants, 17 were attending college at the time of the study, and three had previously graduated. Informed consent was obtained from all participants. For eligibility purposes, autistic participants were required to present a diagnostic report from a certified clinical psychologist or medical professional. Eight autistic participants presented documentation that included scores from a past administration of the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2012). The age of autism diagnosis ranged from age 3 to age 20, with the average age of diagnosis being around 15 years old (SD_{years} = 6.8; Median = 17). Two participants reported challenges in retrieving their documentation. As an alternative, participants offered confirmation through their college's academic office for students with disabilities as an acknowledgment of their diagnosis.

Self-report measures of autistic traits

All participants completed self-report questionnaires to measure autistic traits. The Social Responsiveness Scale, Second Edition (SRS-2) was administered to assess autistic individual differences related to social reciprocity and communication skills (Constantino & Gruber, 2012). The SRS-2 quantifies social communication behaviors related to: social awareness, social cognition, social communication, social motivation, and restricted and repetitive behaviors (RRBs). Formatted as likert-scale response options (not true, sometimes true, often true, and almost always true), raw total scores are converted into gender-normed *t*-scores. As expected, average *t*-scores on the SRS were significantly higher for autistic participants (M = 73.9, SD = 9.44) than NT participants (M = 50.4, SD = 8.29; $t(37.38)$ = 8.39, $p < 0.001$, d = 2.65).

AV stimuli

Identical stimuli and room set up used in Marin et al. (2021) was used in this study. For a visual depiction and specifications of these stimuli, refer to Marin et al. (2021).

Procedure

Each participant was first invited to schedule an online meeting where an assistant explained the study details and requirements. Autistic participants were asked to provide clinical documentation of their autism diagnosis then were sent links to complete online questionnaires. The ERP procedures were identical to experiment two in Marin et al. (2021).

EEG/ERP processing pipeline

The EEG data acquisition specifications and the ERP pipeline were identical to Marin et al. (2021). The only exception is the use of automated artifact detection criteria to help flag excessive, muscle-based artifacts. This criterion utilizes a sliding window peak-to-peak threshold method, where any 200ms window with a voltage change exceeding 75 µV, relative to a 700 ms baseline period, is flagged as an artifact. On average, participants contributed 399.6 valid ERP segments (*SD* = 11.9) across the four conditions. We conducted a 4 (condition; Ao, AVo, AVs, AVa) x 2 (group; ASD vs NT) repeated measures ANOVA to test for differences in usable ERP segments between each condition and group. Statistical analysis showed no differences in the number of usable segments across trial types (*F*(1, 38) = 0.08, *p* = 0.78). We did, however, find a marginally significant main effect of group, where the autistic group (*M* = 403.3, *SD* = 2.7) had more usable segments compared to NT (*M* = 396.2, *SD* = 3.7; *F*(1, 38) = 3.9, *p* = 0.06). There was no interaction between trial type and group (*F*(1, 38) = 0.01, *p* = 0.93).

ERP regions & components of interest

For this study, the impact of temporal synchrony underlying dynamic AV events on early sensory processing was assessed by examining the auditory N1 and P2 components. The N1 was defined as the mean peak amplitude and latency within a 100–200 msec window following the onset sound, while the P2 was defined as the mean peak amplitude and latency within a 200–300 msec window post sound onset. The selection of components, time windows, and the regions of interest were the same as described in Marin et al., (2021). The use of mean amplitude measures instead of peak amplitude measures is a deviation from our previous work in Marin et al. (2021). Mean peak amplitude was used to reduce the impact of random noise or fluctuations in the ERP signal. Averaging across multiple data points around each peak can better control for some of the variability caused by noise (Luck, 2014, Picton et al., 2000, Handy, 2005), leading to more stable and reliable measurements of the N1 and P2 responses across participants and between groups. For mean amplitudes, we averaged each sample point within the predefined N1 and P2 time windows across all usable ERP segments to create a single average for each channel and trial type, per participant. N1-P2 peak-to-peak (P2P) mean amplitude measurements were obtained by subtracting the P2 activity from the N1 response.

Results

Figure 2.1 depicts the grand averaged ERPs between NT and autistic participants for each trial type. Subfigure 1c is the ERP between groups, collapsed across trial types.

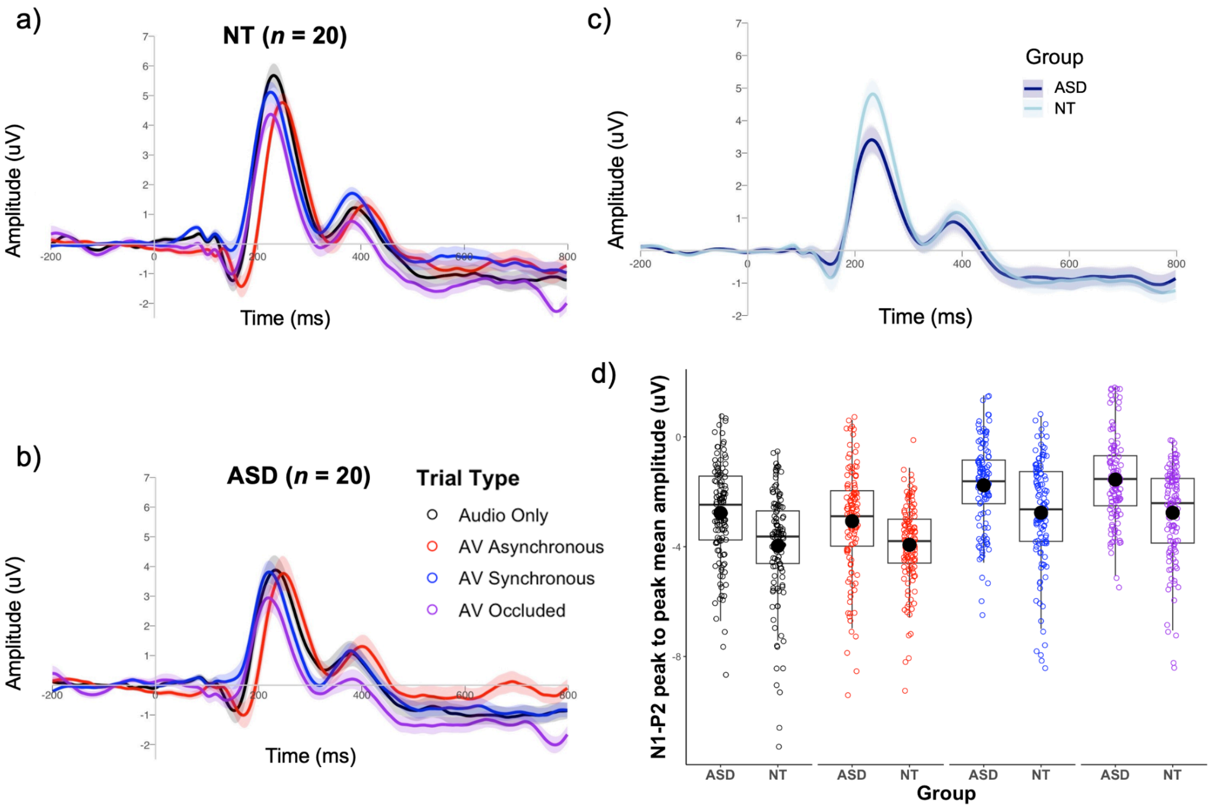

Figure 2.1. *Grand averaged ERPs between the autistic and NT groups, and N1-P2 peak-to-peak mean amplitude across trial types.* Subfigures a and b represent the grand averaged ERPs for each trial type, split between the NT (fig. 1a) and autistic (fig. 1b) groups. The black ERP trace corresponds to the Ao condition, the blue trace represents the AVs condition, the red trace represents the AVa condition, and the purple trace represents the AVo condition. Subfigure c is the grand averaged ERP for NT (in light blue) and autistic participants (in dark blue), collapsed across the four trial types. On each ERP figure, the y-axis represents voltage, with positive values plotted upwards, while the x-axis represents time in milliseconds (msec). The error bars around the ERP indicate the upper and lower bounds of one within-subject standard error of the mean (+/-). Subfigure d is a scatter plot illustrating N1-P2 P2P mean amplitude for each participant. In this plot, each scatter point represents the single-channel level activity from a six-channel central-frontal auditory region for individual participants. Each boxplot shows the interquartile range (IQR) between the first and third quartiles, with black horizontal lines indicating median averages. Whiskers extend to the minimum and maximum values within the IQR. The black point within each box plot represents model estimates of the marginal means for N1-P2 P2P mean amplitude for each trial type. Note that the formatting conventions for the ERP figures and scatter plots will remain consistent across all subsequent figures.

We opted to employ linear mixed effects modeling in lieu of traditional ANOVA approaches used in Marin et al. (2021). Linear mixed effects modeling is often more flexible over traditional

ANOVA approaches because it handles nested or repeated data better, allows for random effects, and offers increased power (Hox, 2010; West et al., 2014). For each ERP analysis reported below, we tested the interaction between group (2-level between-subjects factor; autism vs. NT) and trial type (4-level within-subjects factor; Ao, AVs, AVa, AVo), which was defined by the following code in R: lmer(ERPmeasure ~ Trial Type*Group + (1|Participant:Channel). We used a nested effects structure to model the random effects, considering amplitudes and latencies from the six EEG channels contained within a single frontal-central auditory region, nested within each participant. We then assessed model fit using an analysis of deviance, comparing the Wald Chi-Square tests for each main effect, as well as their interaction.

We will conduct a likelihood ratio test to assess model fit via a reduction in the Akaike Information Criterion (AIC). Here, the interaction model will be compared to additive main effect and null models to assess fit. A significant decrease in AIC suggests that a new model provides a better balance between goodness of fit and complexity compared to the previous model. We will then perform post-hoc tests on the model that fits the data best, as determined by significant decreases in AIC for each analysis. All statistical analyses were conducted in R (version 4.3.2), using the ‘tidyverse’, ‘emmeans’, and ‘lmerTest’ packages. Regarding post-hoc tests for significant interactions between trial type and group, our main focus, as hypothesized, is on analyzing the differences between groups for each trial type (e.g., NT response to AVa vs. ASD response to AVa). Significant group differences between different trial types (e.g., NT response to AVo vs. ASD response to AVa) were not expected based on our hypotheses, and will not be interpreted as meaningful. For reference, *p*-values for these comparisons can be found in Appendix A.

N1–P2 peak-to-peak mean amplitude

The analysis of model fit found that trial type ($\chi2(3) = 282.2$, $p < .001$) and group ($\chi2(1) = 27.8$, $p < .001$) were significant predictors of N1-P2 P2P mean amplitude. The interaction between trial type and group was also marginally significant ($\chi2(3) = 7.3$, $p = .06$). A likelihood ratio test shows that the interaction model with trial type and group as interacting predictors marginally explained more variation (AIC = 3045.6) in the observed data compared to a main effects model with trial type and group as additive predictors (AIC = 3046.9, $p = .06$). The interaction model did significantly explain more variation compared to a null model (AIC = 3452.4, $p < .001$). We conducted post hoc tests for the trending group by trial type interaction using the emmeans package in R. Table 2.1 depicts the model estimated marginal means of N1-P2 P2P mean amplitude and N1/P2 latency responses between the autistic and NT groups and across the four trial types.

Table 2.1
Marginal means of peak amplitude and latency between each group, across trial types

	N1-P2 mean P2P amp. (uV)		N1 latency (ms)		P2 latency (ms)	
	Autism	NT	Autism	NT	Autism	NT
	M(SE)	*M(SE)*	*M(SE)*	*M(SE)*	*M(SE)*	*M(SE)*
Trial Type						
Ao	-2.8(.2)	-4(.2)	150(2.1)	150(2.1)	233(1.3)	233(1.3)
AVs	-1.8(.2)	-2.8(.2)	138(2.1)	139(2.1)	223(1.3)	227(1.3)
AVa	-3.1(.2)	-3.9(.2)	164(2.1)	158(2.1)	248(1.3)	248(1.3)
AVo	-1.6(.2)	-2.8(.2)	136(2.1)	140(2.1)	227(1.3)	227(1.3)

Post-hoc tests for the trending interaction effect between trial type and group revealed the autistic group had smaller N1-P2 P2P mean amplitude responses across all four trial types compared to NT (all *t*'s > 3.8, all *p*'s < .005, all *d*'s > 1.02). Moreover, there were a number of significant differences between the groups, and across the four trial types. We omitted the significant results of these interaction comparisons due to the absence of explicit hypotheses, but detailed descriptions for each ERP measure can be found in Appendix B. The pattern of N1-P2 mean amplitudes between the four trial types was the same within each group. Within both groups, post-hoc tests revealed that the AVs and AVo responses were significantly reduced compared to the AVa (all *t'*s < -10.8, all *p'*s < .001, all *d*'s < -1.39) and Ao (all *t'*s < -11.1, all *p'*s < .001, all *d*'s < -1.2). There were no differences between the AVs and AVo responses (both *p'*s > .54). In both groups, the AVa response was not statistically different compared to the Ao response (ASD, *p* = 0.1; NT, *p* = 1.0).

N1 peak latency

The analysis of model fit found the main effect of trial type was a significant predictor of N1 peak latency (χ2(3) = 187.1, *p* < .001). The main effect of group was not a significant predictor of N1 peak latency (χ2(1) = 0.001, *p* = .97), but the interaction between trial type and group was (χ2(3) = 8.5, *p* = .04). The analysis of model fit for N1 peak latency found that the interaction model with trial type and group as predictors explained significantly more variation (AIC = 8557.7) in the observed data compared to an additive main effects model (AIC = 8560.1, *p* = .04) and to a null model (AIC = 8786, *p* < .001; see Appendix C for a N1 latency figure).

Notably, post-hoc tests revealed no group differences in N1 peak latency between the four trial types (all *p*'s > .57). N1 peak latency post-hoc tests revealed similar patterns within both groups. The AVs response peaked sooner compared to AVa (both *t'*s > 8.6, both *p'*s < .001, both *d*'s > 1.1) and Ao (both *t'*s > 4.9, both *p'*s < .001, both *d*'s > 0.63). There were no differences in N1 peak latency between AVs and AVo processing (*p* > .98). The AVo response also peaked earlier compared to AVa (both *t'*s > 8.1, both *p'*s < .001, both *d*'s > 1) and Ao (both *t'*s > 4.4, both *p'*s < .001, both *d*'s > 0.6). The AVa response also exhibited a delay in N1 peak latency compared to Ao (both *t'*s < -3.7, both *p'*s < .006, both *d*'s < -0.5).

P2 peak latency

The analysis of model fit found the main effect of trial type was a significant predictor of P2 peak latency (χ2(3) = 633, *p* < .001). The main effect of group was not a significant predictor of P2 peak latency (χ2(1) = 0.03, *p* = .86), but the interaction between trial type and group was significant (χ2(3) = 11.8, *p* = .008). A likelihood ratio test revealed that the interaction model with trial type and group as predictors explained significantly more variation (AIC = 7338) in the observed data compared to the main effects model with trial type and group as additive predictors (AIC = 7343.8, *p* = .008). The interaction model also explained more variation compared to a null model (AIC = 8012.6, *p* < .001; see Appendix C for a P2 latency figure).

Post-hoc tests revealed no group differences in P2 peak latency between the four trial types (all *p*'s > .38). Within both groups, the overall pattern was virtually the same. Responses toward AVs resulted in faster P2 peak latencies compared to AVa (both *t's* < 19.5, both *p's* < .001, both *d*'s > 2.5) and Ao processing (both *t's* > 5, both *p's* < .001, both *d*'s > 0.65). AVa responses resulted in a delayed P2 peak with respect to Ao (both *t's* < -14.3, both *p's* < .001, both *d*'s < -1.84) and AVo (both *t's* > 19.6, both *p's* < .001, both *d*'s > 2.5). The AVo P2 response also peaked sooner compared to the Ao response (both *t's* > 5.1, both *p's* < .001, both *d*'s > .66). The only difference within the groups regarding P2 latency occurred between the AVs and AVo trial types. For the autism group, the P2 peaked sooner for the AVs compared to AVo (*t* = -3.7, *p* = .005, *d*'s = -0.5), but this difference was absent in NT (*p* = 1.00).

Auditory ERP subtraction technique to account for group-level amplitude differences

As demonstrated by the four condition N1-P2 P2P amplitude ERP analysis, there were overall differences in amplitudes between the two groups, where the autistic group showed smaller auditory responses compared to NT. To control for group differences in auditory processing, we opted to subtract auditory-related activity from each AV condition since the sound used in the Ao condition was identical to the sounds used in the other three AV conditions. To do this, we took the averaged Ao response from each individual participant and subtracted it from their averaged AV responses. This allowed us to accurately consider and adjust for any low-level auditory processing differences between the groups. Figure 3.2 displays the ERP difference waves between trial types, which reveal comparable amplitude effects between the groups.

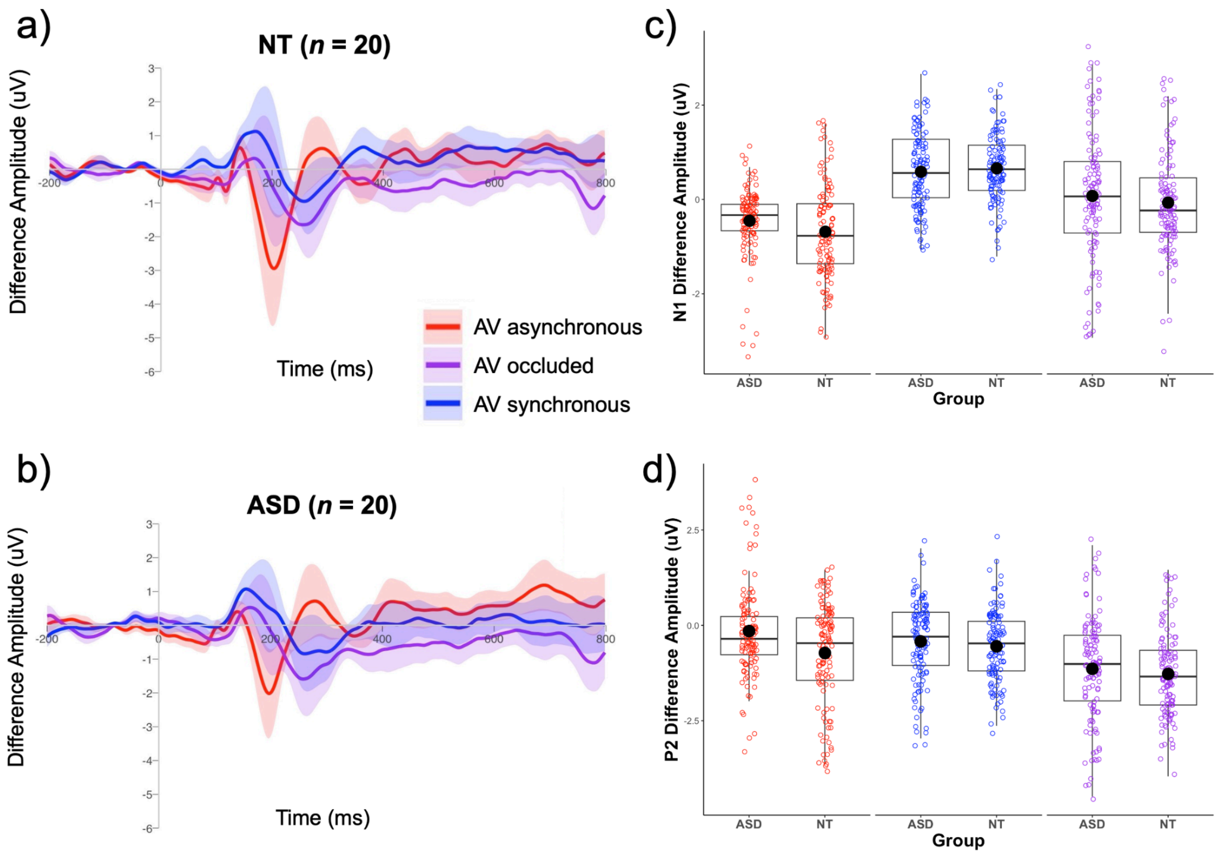

Figure 2.2. *AV Difference ERPs between autistic and NT groups.* The figure above depicts the grand-averaged difference ERPs from a frontal-central auditory region. The Ao response for each participant has been subtracted from each AV condition. Sub-figure a) represents the difference ERPs for the autistic group, while subfigure b) depicts NT. The error bars around the ERP indicate the upper and lower bounds of one within-subject standard error of the mean (+/-). On the ERP figure, the y-axis represents difference amplitude, while the x-axis represents time in milliseconds (ms). Subfigure 2c displays a scatter plot illustrating the non-significant interaction for N1 mean difference amplitude, split between trial type and group. On the N1 scatter plot, the y-axis represents the difference amplitude, where positive values indicate reduced amplitudes relative to Ao processing, while the x-axis is trial type. Subfigure 2d displays scatter plots illustrating the P2 mean difference amplitude. Here, the y-axis represents the difference amplitude, with more positive values indicating greater amplitudes relative to Ao processing.

We averaged the mean amplitude for each participant across the three difference wave ERPs, which was our main dependent variable. We did not consider N1 and P2 latency analyses because of the auditory subtraction technique employed.

N1–P2 peak-to-peak mean differential amplitude

As depicted in Figure 3.2, the grand averaged auditory difference ERP exhibited variations in N1-P2 P2P mean amplitude across the three trial types and the two groups. For this model, we tested the interaction between two fixed factors: AV trial type (3-level within-subjects factor AVs, AVa, AVo) and group (2-level between-subjects factor; Autism and NT). The analysis of model fit found a non-significant interaction between trial type and group on N1-P2 P2P mean differential amplitude ($\chi2(2) = 4.3$, $p = .12$). Due to the presence of a non-significant interaction, we chose to conduct main effect tests of model fit using the additive main effects model. Here, we found significant main effects of trial type ($\chi2(2) = 306$, $p < .001$) and group ($\chi2(1) = 3.8$, $p = .05$) on N1-P2 P2P mean differential amplitude. A likelihood ratio test for N1-P2 P2P mean differential amplitude suggests the trial type by group interaction model (AIC = 2084.7) did not explain the data better compared to the additive main effects model (AIC = 2085, $p = .12$). The interaction and additive main effect models did significantly explain the data better compared to a null model (AIC = 2320.3, both *p*'s < .001; see Appendix C for a figure of the non-significant interaction). Table 2.2 provides model estimated marginal means for the N1-P2 P2P difference amplitude, N1 difference amplitude, and P2 difference amplitude across trial type and group.

Table 2.2

Marginal means of differential mean peak amplitude between both groups, across trial types

	N1-P2 P2P difference amp. (uV)		N1 mean difference amp. (uV)		P2 mean difference amp. (uV)	
	Autism	NT	Autism	NT	Autism	NT
	M (SE)	*M (SE)*	*M (SE)*	*M (SE)*	*M (SE)*	*M (SE)*
Trial Type						
AVs	1(.1)	1.2(.1)	0.6(.09)	0.7(.09)	-0.4(.11)	-0.6(.11)
AVa	-0.3(.1)	0.04(.1)	-0.5(.09)	-0.9(.09)	-0.2(.11)	-0.7(.11)
AVo	1.2(.1)	1.2(.1)	0.1(.09)	-0.1(.09)	-1.1(.11)	-1.3(.11)

Post-hoc tests of the main effect of trial type revealed that the AVa response exhibited greater differential amplitudes compared to the AVs and AVo (both *t*'s < -14.5, both *p*'s < .001, both *d*'s < -2.70). There were no differences between the AVo and AVs (*p* = .45). For the significant main effect of group, we found that the autistic N1-P2 P2P differential mean amplitude response collapsed across the three AV trial types (*M* = 0.64, 95% CI [0.51, 0.77]) was smaller compared to NT (*M* = 0.82, 95% CI [0.69, 0.95]; *t* = -2, *p* = .05, *d* = -0.39).

Analysis of individual N1/P2 amplitude responses

In addition to N1-P2 P2P difference amplitudes, we chose to analyze difference amplitudes across the individual components of the auditory response. Analyzing individual peaks in ERP data offers several advantages over traditional P2P analyses. ERP components exhibit variability in morphology, latency, and amplitude across individuals and experimental conditions. By focusing on individual peaks, we can better capture this variability and accurately account for individual differences in ERP responses. In clinical settings, individual peak analyses can be particularly informative for identifying differences in ERP components associated with neurological or psychiatric conditions.

N1 mean differential amplitude

The analysis of model fit found a non-significant interaction between trial type and group on N1 mean differential amplitude ($\chi2(2) = 3.8$, *p* = .15). Due to the presence of a non-significant interaction, we chose to conduct main effect tests of model fit using the additive main effects model. Here, we found a significant main effect of trial type ($\chi2(2) = 216.9$, *p* < .001), but not group ($\chi2(1) = 1.4$, *p* = .23). Analyses of model fit for N1 differential mean amplitude suggest that the trial type by group interaction model (AIC = 1994.4) did not explain the data better compared to the additive main effects model (AIC = 1994.2, *p* = .15). The interaction and additive main effects models did explain the data better compared to a null model (AIC = 2169.2, both *p*'s < .001; see figure 2.2c for the interaction between trial type and group). Post-hoc tests of the significant main effect of trial type revealed that the AVa response exhibited greater N1 differential mean amplitudes compared to AVs (*t* = -14.7, *p* < .001, *d* = -1.34) and AVo (*t* = -7.1, *p* < .001, *d* = -0.65). In addition, the AVs response was reduced relative to the AVo response (*t* = 7.7, *p* < .001, *d* = 0.7).

P2 mean differential amplitude

The analysis of model fit found that trial type (χ2(2) = 53.2, *p* < .001) and group (χ2(1) = 14, *p* < .001) were significant predictors of P2 mean differential amplitude. The analysis also revealed a significant interaction between trial type and group (χ2(2) = 6.6, *p* = .04; see figure 2.2d). Analyses of model fit suggest the AV condition by group interaction model (AIC = 2282) explains the data better compared to an additive main effects model (AIC = 2284.6, *p* = .04) and to a null model (AIC = 2356, *p* < .001). Post-hoc tests for the P2 differential mean amplitudes revealed that the autistic group had larger P2 differential mean amplitude responses for the AVa condition compared to NT response toward AVa (*t* = 3.7, *p* = .003, *d* = .53). There were no significant group-level differences between the AVs (*p* = .96) and AVo (*p* = .95) responses. The overall pattern of P2 mean differential amplitude responses within both groups were the same. For both groups, the differential AVo response was reduced compared to the AVs (both *t*'s > 5.1, both *p*'s < .001, both *d*'s > .66) and AVa (both *t*'s > 3.9, both *p*'s < .001, both *d*'s > .51) responses. There were no significant differences in P2 differential amplitude between the AVs and AVa responses in both groups (both *p*'s > .39).

Exploratory ERP associations with autistic traits

We conducted a series of additional exploratory analyses to better understand the connection between predictive neural signals that revealed group level differences (i.e., N1-P2 P2P mean amplitude, AVs/AVo P2 peak latency difference, and P2 mean differential AVa amplitudes) and behaviors related to self-reported autistic traits. For ERP measures, we calculated participant level y-intercepts, whereas for autistic traits, we obtained the total SRS *t*-score and each of the five *t*-scored subscales (i.e., RRBs, social cognition, social communication, social awareness, and social motivation). We obtained ERP intercepts for each individual participant and trial type using the following code in R: lmer (ERP measure ~ Group*Trial Type + (1| Participant:Trial Type). We opted to run correlations within the autistic group because of the presence of group-level differences in ERP and behavioral measures. For this set of exploratory analyses, we tested 36 unique correlations using the intercepts from each ERP measure that revealed group differences and the six *t*-score measures of the SRS. Thus, after applying Bonferroni correction, *p*-values below .001 will be considered statistically significant.

First, we tested correlations between N1-P2 P2P mean amplitudes and SRS *t*-scores across the four trial types, within the autistic group. We found that more positive intercepts (i.e., greater amplitudes) in response to AVa was related to a decrease in SRS RRB *t*-scores in the autistic group, *r*(18) = -0.48, *p* = .03. The same association was not seen within each of the four other subdomains of the SRS (all *p*'s > .47), the SRS total *t*-score (*p* = .48), nor was the RRB *t*-score correlated with amplitude responses toward Ao (*p* = .09), AVs (*p* = .18), and AVo (*p* = .27). There were no associations between each of AV responses and RRB *t*-scores in the NT group (*p* = .37). Figure 2.3 presents the correlations between the N1-P2 amplitudes for each trial type and the SRS RRB *t*-scores within the autistic group.

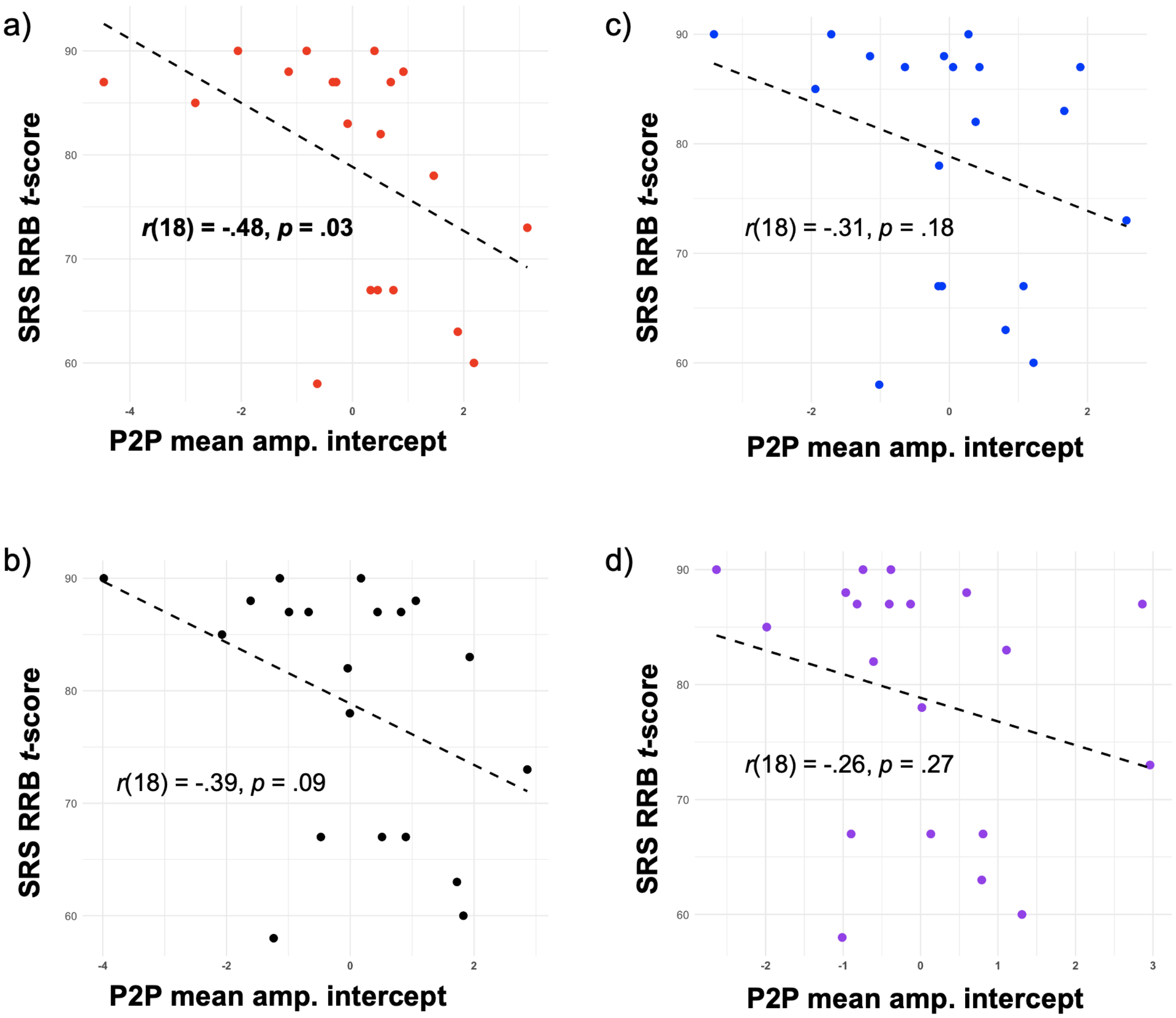

Figure 2.3. *Correlations between the SRS RRB t-score and N1-P2 peak-to-peak mean amplitude intercepts within the autistic group.* The figure above depicts the correlations between the SRS RRB *t*-scores and AVa (fig. 3a), Ao (fig. 3b), AVs (fig. 3c), and AVo (fig. 3d) intercepts. Each scatter point, color coded by trial type, represents a participant's N1-P2 P2P mean amplitude intercept, plotted on the x-axis, and SRS RRB *t*-score, plotted on the y-axis for each trial type. A black dashed line of best fit was drawn for visualization purposes.

No significant correlations were found between SRS measures and the difference in P2 latency between the AVs and AVo intercepts (all *p*'s > .28), or the P2 differential amplitude response toward AVa (all *p*'s > .1). Note that the correlations reported here failed to reach significance after correcting for multiple comparisons.

Discussion

The current study measured auditory ERPs between autistic and age-matched NT adults to evaluate neural signatures associated with prediction and error. There has been no attempt to directly differentiate neural signals that represent the processing of a sound based on predictable top-down information from sounds that elicit bottom-up error in autistic adults. Here, participants observed a moving object generate synchronous, asynchronous, or occluded collision sounds while ERPs were recorded. Auditory ERPs showed a consistent response profile across all participants, with both fully visible and occluded collisions resulting in reduced auditory responses compared to isolated and asynchronous sounds, replicating Marin et al. (2021). Autistic participants also had reduced sensory responses to sounds across all trial types relative to NT. To account for group-level auditory differences, we subtracted out neural activity related to unimodal auditory processing from each of the three AV scenarios, which resulted in comparable ERP responses between the groups.

Analyses of differential mean amplitude revealed the presence of bottom-up sensory differences, where larger P2 amplitudes toward asynchronous collision sounds were observed in autistic participants. Critically, autistic ERP responses to fully visible and occluded synchrony were comparable to NT, thereby suggesting top-down auditory predictions were functional in this sample, even in the absence of visual stimulation. P2 latency responses were also different in autism, where the P2 peaked faster toward fully visible AV synchrony compared to occluded synchrony - a response not seen in NT. Exploratory correlational analyses revealed that overall amplitudes toward asynchrony were related to self-reported variability in autistic traits related to RRBs, but this association did not remain significant after adjusting for multiple comparisons. Our findings indicate that early auditory responses are affected by the alignment of dynamic AV input, highlighting neural distinctions between autistic and NT participants. Here, we provide mechanistic evidence for predictive coding differences in autism that are driven by early auditory differences related to the perception of error, and not the integration of top-down expectations.

P2 amplitudes toward asynchronous processing were enlarged in ASD relative to NT

Notably, we found that the autistic P2 differential response elicited by asynchronous sound was enlarged relative to NT. Group differences were not seen in responses that originated from synchronous and occluded AV input. These results suggest that group-level differences were driven by responses related to the perception of bottom-up sensory error, and not the integration of top-down expectations in autism. This suggests that bottom-up error processing is what drives predictive processing differences between our sample of autistic and NT participants. Our findings mirror those of van Laarhoven et al. (2020), who observed enlarged error responses from autistic participants when perceiving an absence of sound elicited by a person clapping their hands. The amplified error signaling measured in this study builds upon the finding of increased error responses in autism, which extends beyond the scope of social cognitive processing to include basic perceptual predictions.

Not all studies examining error and prediction yield results consistent with those reported here. A recent meta-analysis by Chen et al. (2020) reviewed auditory ‘oddball’ mismatch negativity (MMN) responses in autistic populations. Across studies, autistic participants consistently show lower MMN amplitudes and delayed latencies compared to NTs. In flanker tasks, autistic adults (South et al. 2010), as well as autistic children (Sokhadze et al., 2010) show reduced amplitudes in error-related negativity when viewing incompatible target/distractor pairings, indicating possible error attenuation. In reward processing, autistic adults show reduced neural responses to socially contingent rewards compared to NT individuals (Scott-Van Zeeland et al., 2010; Delmonte et al., 2012; Kohls et al., 2013). These studies collectively indicate reduced neural responses in reward anticipation integration and error perception, contrasting with our study and van Laarhoven et al. (2020). Whether hypo- or hyper-sensory processing is present, both indicate that sensory differences impact perception in autism. These differences, regardless of directionality, may affect mental model formation by improperly weighting prediction errors in natural perception.

In our study, we found enlarged error responses at the P2 component, while van Laarhoven et al. (2020) reported larger N1 responses. It could be that the unexpected omission of sound introduces earlier amplitude differences compared to errors related to the synchrony between sound and vision. Alternatively, the differences in component location may have been influenced by dissimilar stimuli used between the two studies. The visual of a handclap was used in van Laarhoven et al. (2020), so the ERPs measured there can be interpreted as intrinsically social. Human hands convey socially meaningful signals that reflect thoughts and intentions (Hoppe et al., 2020; Fausey et al., 2016), and autistic individuals struggle with processing socially relevant stimuli. For instance, Amoruso et al. (2018) found that lower-support autistic adults are less likely than NT to integrate top-down contextual expectations with kinematics when observing actions. Perhaps social processing complexities could cause earlier amplitude differences toward error in autism compared to more non-social perceptual phenomena.

Larger P2 responses toward asynchrony suggest that auditory error signaling differences may occur in slightly later sensory stages in autism. The auditory P2 response is a multifaceted component that contributes to auditory perception, attention, language processing (Lewandowska, 2008; Bolt et al., 2023, Tremblay et al., 2014), stimulus classification (Key et al., 2005) as well as the maintenance of auditory information in working memory tasks (Rader et al., 2008). In autism, the auditory P2 has been associated with social deficits and attention switching difficulty (Chien et al., 2019), AV speech perception (Borgolte et al, 2021), sensitivity toward sound omissions (Foss-feig et al., 2018), and multisensory integration (Russo et al. 2010; Stefanou et al., 2020). The finding of enlarged P2 error amplitude aligns with theoretical accounts, particularly Van De Cruys (2014) and Lawson et al. (2014), who argue that an improper weighting of prediction error is a key driver of predictive differences in autism. Regardless of component location, these results suggest that the perception of error can influence the early sensory processing of sound (< 300ms).

Overall auditory amplitude reduction in ASD relative to NT

Auditory processing amplitudes of the N1-P2 complex were reduced in autism relative to NT. Group-level amplitude differences can hinder the interpretation of meaningful group comparisons. However, we implemented an effective way to limit the influence of group differences by subtracting out auditory-related activity from each AV ERP. Even after implementing the subtraction technique, we observed small group differences in overall N1-P2 peak-to-peak differential amplitudes across the three AV responses. Yet, we discovered that while overall ASD amplitudes were smaller mathematically, P2 amplitude was larger in autism.

Reduced auditory processing in autism may suggest the presence of broader processing variations within the phenotype. For example, a recent meta-analysis by Williams et al. (2020) found reduced and delayed N1 amplitudes to pure tones in autism, where there were no observed group-level differences in P2 responses. It is worth noting that in our sample, larger amplitudes toward asynchrony were linked to reduced RRBs in ASD, but this association was not significant after adjusting for multiple comparisons. Thus, the observed reduction in amplitudes may underscore a crucial aspect to consider when studying sensory processing in autistic individuals, potentially offering clinically meaningful markers of comparison.

After controlling for auditory processing differences, we still found that the differential N1-P2 peak-to-peak response was reduced in autism, but these differences were absent when assessing each component in isolation. Importantly, group level differences toward the perception of error at the P2 were larger in autism. It could be reasonably assumed that larger P2 error responses in autism may be driven by preceding N1 responses, which were smaller on average but not statistically different between the groups (see fig. 2c). However, it is important to note that smaller N1 responses associated with error perception may contribute, at least in part, to some of the reported N1 reduction findings in the literature (Williams et al., 2020). It is often challenging to compare overall amplitude differences, as these could be driven by factors such as skull/hair thickness, statistical power, or potentially more meaningful factors like autism symptom heterogeneity. Regardless of potential reduction occurring at the N1, we report larger P2 amplitudes toward the perception of error that occur early in the processing stream (< 300ms) that map onto other research of larger error responses in autism (van Laarhoven et al., 2020).

Comparable ERP pattern for each trial type within each group

We observed that the overall pattern of mean amplitude and latency responses were similar within groups. Specifically, fully visible and occluded collision resulted in a reduction of the auditory response relative to sound occurring slightly before contact, and to sound that did not contain visual expectations. By analyzing auditory prediction and error signaling, we showed that expected perceptual phenomena are integrated normatively via the presence of reduced auditory amplitudes toward expected sound, reflective of prior integration, in autistic participants. This contrasts with van Laarhoven et al. (2019), who found an absence of auditory reduction to predictable motor-auditory events, thereby concluding that motor-to-auditory predictions may be compromised in autism. In addition, our finding challenges some predictive coding accounts of autism, notably one of the original accounts by Pellicano and Burr (2012) in which predictive differences are attributed to ‘weaker’ prior models in comparison to NTs, positing that top-down components of the autistic predictive circuitry are impacted. While we recognize not all autistic individuals have stable prior models across psychological domains, our research highlights their ability to integrate simple auditory predictions effectively. Findings such as ours are important to highlight that autistic sensory processing is not always a deviation from NT, which is crucial in autism research and benefits the autistic community more widely.

Our sole finding in the difference of overall response patterns between the groups was that autistic individuals exhibited faster P2 latency responses towards occluded AV stimuli compared to fully visible AV synchrony. The occlusion condition can be seen to represent ‘pure’ prediction, as no visual input requires one to rely on expectation alone. Thus, there may be a small multisensory facilitation effect of fully visible synchrony seen in autism. These P2 latency differences did not map onto variability in autistic traits, suggesting this small facilitation effect may not contain significant clinical relevance. Replication efforts are needed.

General Conclusions and Future Directions

Considering the characteristics of our sample is important here, especially when comparing our results to those of van Laarhoven et al. (2019), who found an absence of auditory response reduction toward predictable sound. In contrast, we found that autistic adults do integrate simple auditory predictions, as evidenced by a significant reduction in amplitude toward predictable sound. The discrepancy between the two studies may be partly driven by differences in support levels between the two samples. van Laarhoven et al., (2019) enrolled high-support autistic adults who were residing in long-term care facilities, and who were facing severe mental health challenges. Our sample represents a low-support group, who were recruited primarily from an undergraduate population. Hence, each study may not be broadly representative of the diverse range of symptoms that characterize autism. Therefore, as we observe small differences in this low-support group, there are reasonable hypotheses to suggest that we may see larger effects in populations with higher-support needs. This is not to argue that neural indices have the same stability in all individuals of the spectrum with similar support needs, as we have suggested that there is likely wide heterogeneity in predictive models. However, it appears that within our low-support sample, auditory expectations were integrated normatively in the context of simple perceptual associations. As this is a novel finding, it will be essential to replicate these results using larger and more diverse autistic samples, including those with varying support needs. Thus, replication is necessary to confirm such findings, and to delve deeper into the diverse phenotypic heterogeneity within the autism spectrum.

It is worth noting that, although informative, predictions made from simple AV stimuli do not encapsulate the multi-modal social predictions we make within our daily lives. It is likely that as stimuli become more complex and socially determined, predictions become increasingly difficult to employ. Challenges in social communication are well documented in autism. Social processing challenges may explain earlier error enhancement in autism during unexpectedly omitted sound (van Laarhoven et al., 2019). It could be that more socially demanding predictions could exhibit greater influence over sensory processing in ASD. Studies targeting predictive differences across social contexts are critical to understanding the mechanisms that support the instantiation of predictions across socially relevant domains of psychological functioning.

It is also important to acknowledge that our research findings are constrained by the specific attributes of our sample group, which may impact generalizability. One challenge was independently confirming clinical diagnoses, which introduces some uncertainty regarding the reliability of diagnoses in the autistic sample. Our predominantly female sample also contradicts the 3:1 male-to-female ratio typical in ASD population estimates (Loomes et al., 2017), suggesting that the sex characteristics of our sample may not fully represent the broader autistic population. Yet, research seems to point towards more female-specific profiles in sensory processing challenges in autism for both child (Osório et al., 2021; Kumazaki et al., 2015) and adult (Cardon et al., 2023) populations. For example, Lai et al. (2011) investigated behavioral differences in autistic men and women and found that women reported more lifetime sensory symptoms but fewer social-cognitive challenges. Therefore, our female-heavy sample, although not directly generalizable, may be important in understanding autistic sensory differences in underrepresented female cohorts.

In sum, our results provide evidence for neural differences underlying the perception of error in autism relative to NT. We found that autistic adults, like NTs, exhibited reduced neural responses toward synchronous collision sound relative to sound without visual input. Our results support the idea of dysregulated neural systems in autism concerning the perception of sensory-driven perceptual errors rather than systems related to the integration of top-down expectations. In a broader sense, these findings have significant sensory-based implications for predictive coding theories of autism. These implications are likely important in understanding the heterogeneity within autism and its effects on other related processes, like social cognition.

References

American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.). https://doi.org/10.1176/appi.books.9780890425787

Baldeweg, T. (2007). ERP repetition effects and mismatch negativity generation: A predictive coding perspective. *Journal of Psychophysiology*, *21*(3–4), 204–213. https://doi.org/10.1027/0269-8803.21.34.204

Baron-Cohen, S. (2000). Theory of mind and autism: A review. *International Review of Research in Mental Retardation, 23,* 169–184. https://doi.org/10.1016/S0074-7750(00)80010-5

Bolt, N. K., & Loehr, J. D. (2023). The auditory P2 differentiates self- from partner-produced sounds during joint action: Contributions of self-specific attenuation and temporal orienting of attention. *Neuropsychologia*, *182*, 108526. https://doi.org/10.1016/j.neuropsychologia.2023.108526

Borgolte, A., Roy, M., Sinke, C., Wiswede, D., Stephan, M., Bleich, S., Münte, T. F., & Szycik, G. R. (2021). Enhanced attentional processing during speech perception in adult high-functioning autism spectrum disorder: An ERP-study. *Neuropsychologia*, *161*, 108022. https://doi.org/10.1016/j.neuropsychologia.2021.108022

Cannon, J., O'Brien, A. M., Bungert, L., & Sinha, P. (2021). Prediction in autism spectrum disorder: A systematic review of empirical evidence. *Autism Research: Official Journal of the International Society for Autism Research*, *14*(4), 604–630. https://doi.org/10.1002/aur.2482

Cardon, G., McQuarrie, M., Calton, S., & Gabrielsen, T. P. (2023). Similar overall expression, but different profiles, of autistic traits, sensory processing, and mental health between young adult males and females. *Research in Autism Spectrum Disorders*, *109*, 102263. https://doi.org/10.1016/j.rasd.2023.102263

Chen, T.-C., Hsieh, M. H., Lin, Y.-T., Chan, P.-Y. S., & Cheng, C. H. (2020). Mismatch negativity to different deviant changes in autism spectrum disorders: A meta-analysis. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *131*(3), 766–777. https://doi.org/10.1016/j.clinph.2019.10.031

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in Cognitive Sciences*, *16*(4), 231–239. https://doi.org/10.1016/j.tics.2012.02.007

Chien, Y. L., Hsieh, M. H., & Gau, S. S. F. (2019). P50-N100-P200 sensory gating deficits in adolescents and young adults with autism spectrum disorders. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, *95*, 109683. https://doi.org/10.1016/j.pnpbp.2019.109683

Clementz, B. A., Barber, S. K., & Dzau, J. R. (2002). Knowledge of stimulus repetition affects the magnitude and spatial distribution of low-frequency event-related brain potentials. *Audiology & Neuro-Otology*, *7*(5), 303–314. https://doi.org/10.1159/000064444

Constantino J. N. and Gruber C. P. (2012). Social Responsiveness Scale–Second Edition (SRS-2). Torrance, CA: Western Psychological Services.

D'Andrea-Penna, G. M., Iversen, J. R., Chiba, A. A., Khalil, A. K., & Minces, V. H. (2020). One tap at a time: Correlating sensorimotor synchronization with brain signatures of temporal processing. *Cerebral Cortex Communications*, *1*(1), tgaa036. https://doi.org/10.1093/texcom/tgaa036

Delmonte, S., Balsters, J. H., McGrath, J., Fitzgerald, J., Brennan, S., Fagan, A. J., & Gallagher, L. (2012). Social and monetary reward processing in autism spectrum disorders. *Molecular Autism*, *3*(1), 7. https://doi.org/10.1186/2040-2392-3-7

Fausey, C. M., Jayaraman, S., & Smith, L. B. (2016). From faces to hands: Changing visual input in the first two years. *Cognition*, *152*, 101–107. https://doi.org/10.1016/j.cognition.2016.03.005

Ford, J. M., & Hillyard, S. A. (1981). Event-related potentials (ERPs) to interruptions of a steady rhythm. *Psychophysiology*, *18*(3), 322–330. https://doi.org/10.1111/j.1469-8986.1981.tb03043.x

Ford, J. M., Roach, B. J., Faustman, W. O., & Mathalon, D. H. (2007). Synch before you speak: Auditory hallucinations in schizophrenia. *The American Journal of Psychiatry*, *164*(3), 458–466. https://doi.org/10.1176/ajp.2007.164.3.458

Foss-Feig, J. H., Stavropoulos, K. K. M., McPartland, J. C., Wallace, M. T., Stone, W. L., & Key, A. P. (2018). Electrophysiological response during auditory gap detection: Biomarker for sensory and communication alterations in autism spectrum disorder? *Developmental Neuropsychology*, *43*(2), 109–122. https://doi.org/10.1080/87565641.2017.1365869

Friston, K. (2012). Prediction, perception and agency. *International Journal of Psychophysiology*, *83*(2), 248–252. https://doi.org/10.1016/j.ijpsycho.2011.11.014

Greene, R. K., Zheng, S., Kinard, J. L., Mosner, M. G., Wiesen, C. A., Kennedy, D. P., & Dichter, G. S. (2019). Social and nonsocial visual prediction errors in autism spectrum disorder. *Autism Research: Official Journal of the International Society for Autism Research*, *12*(6), 878–883. https://doi.org/10.1002/aur.2090

Heeger, D. J. (2017). Theory of cortical function. *Proceedings of the National Academy of Sciences*, *114*(8), 1773–1782. https://doi.org/10.1073/pnas.1619788114

Hilton, C. L., Harper, J. D., Kueker, R. H., Lang, A. R., Abbacchi, A. M., Todorov, A., & LaVesser, P. D. (2010). Sensory responsiveness as a predictor of social severity in children with high functioning autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *40*(8), 937–945. https://doi.org/10.1007/s10803-010-0944-8

Hox, J., Moerbeek, M., & van de Schoot, R. (2010). Multilevel analysis: Techniques and applications, Second Edition (2nd ed.). Routledge. https://doi.org/10.4324/9780203852279

Huang, C.-C., Amini, B., & Bitmead, R. R. (2019). Predictive coding and control. *IEEE Transactions on Control of Network Systems*, *6*(2), 906–918. https://doi.org/10.1109/TCNS.2018.2882190

Huang, Y., & Rao, R. P. N. (2011). Predictive coding. *WIREs Cognitive Science*, *2*(5), 580–593. https://doi.org/10.1002/wcs.142

Jiang, L. P., & Rao, R. P. N. (2022). Predictive coding theories of cortical function. *Oxford Research Encyclopedia of Neuroscience.* https://doi.org/10.1093/acrefore/9780190264086.013.328

Kaye, A. P., & Krystal, J. H. (2020). Predictive processing in mental illness: Hierarchical circuitry for perception and trauma. *Journal of Abnormal Psychology*, *129*(6), 629–632. https://doi.org/10.1037/abn0000628

Key, A. P. F., Dove, G. O., & Maguire, M. J. (2005). Linking brainwaves to the brain: An ERP primer. *Developmental Neuropsychology*, *27*(2), 183–215. https://doi.org/10.1207/s15326942dn2702_1

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Defining and quantifying the social phenotype in autism. *The American Journal of Psychiatry*, *159*, 895–908. https://doi.org/10.1176/appi.ajp.159.6.895

Kohls, G., Schulte-Rüther, M., Nehrkorn, B., Müller, K., Fink, G. R., Kamp-Becker, I., Herpertz-Dahlmann, B., Schultz, R. T., & Konrad, K. (2013). Reward system dysfunction in autism spectrum disorders. *Social Cognitive and Affective Neuroscience*, *8*(5), 565–572. https://doi.org/10.1093/scan/nss033

Kononowicz, T. W., & Rijn, H. van. (2014). Decoupling interval timing and climbing neural activity: A dissociation between CNV and N1 P2 Amplitudes. *Journal of Neuroscience*, *34*(8), 2931–2939. https://doi.org/10.1523/JNEUROSCI.2523-13.2014

Kumazaki, H., Muramatsu, T., Kosaka, H., Fujisawa, T. X., Iwata, K., Tomoda, A., Tsuchiya, K., & Mimura, M. (2015). Sex differences in cognitive and symptom profiles in children with high functioning autism spectrum disorders. *Research in Autism Spectrum Disorders*, *13–14*, 1–7. https://doi.org/10.1016/j.rasd.2014.12.011

Lange, K. (2013). The ups and downs of temporal orienting: A review of auditory temporal orienting studies and a model associating the heterogeneous findings on the auditory N1 with opposite effects of attention and prediction. *Frontiers in Human Neuroscience*, *7*, 1-7. https://doi.org/10.3389/fnhum.2013.00263

Lewandowska, M., Bekisz, M., Szymaszek, A., Wrobel, A., & Szelag, E. (2008). Towards electrophysiological correlates of auditory perception of temporal order. *Neuroscience Letters*, *437*(2), 139–143. https://doi.org/10.1016/j.neulet.2008.03.085

Loomes, R., Hull, L., & Mandy, W. P. L. (2017). What is the male-to-female ratio in autism spectrum disorder? A systematic review and meta-analysis. *Journal of the American*

Academy of Child and Adolescent Psychiatry, *56*(6), 466–474. https://doi.org/10.1016/j.jaac.2017.03.013

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

MacLennan, K., Woolley, C., @21andsensory, E., Heasman, B., Starns, J., George, B., & Manning, C. (2023). “It is a big spider web of things”: Sensory experiences of autistic adults in public spaces. *Autism in Adulthood*, *5*(4), 411–422. https://doi.org/10.1089/aut.2022.0024

Marin, A., Störmer, V. S., & Carver, L. J. (2021). Expectations about dynamic visual objects facilitates early sensory processing of congruent sounds. *Cortex*, *144*, 198–211. https://doi.org/10.1016/j.cortex.2021.08.006

Menceloglu, M., Grabowecky, M., & Suzuki, S. (2020). EEG state-trajectory instability and speed reveal global rules of intrinsic spatiotemporal neural dynamics. *PLOS ONE*, *15*(8), e0235744. https://doi.org/10.1371/journal.pone.0235744

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *118*(12), 2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026

Osório, J. M. A., Rodríguez-Herreros, B., Richetin, S., Junod, V., Romascano, D., Pittet, V., Chabane, N., Jequier Gygax, M., & Maillard, A. M. (2021). Sex differences in sensory processing in children with autism spectrum disorder. *Autism Research: Official Journal* *of the International Society for Autism Research*, *14*(11), 2412–2423. https://doi.org/10.1002/aur.2580

Pellicano, E., & Burr, D. (2012). When the world becomes "too real": A bayesian explanation of autistic perception. *Trends in Cognitive Sciences*, *16*(10), 504–510. https://doi.org/10.1016/j.tics.2012.08.009

Picard, F., & Friston, K. (2014). Predictions, perception, and a sense of self. *Neurology*, *83*(12), 1112–1118. https://doi.org/10.1212/WNL.0000000000000798

Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, R., Miller, G. A., Ritter, W., Ruchkin, D. S., Rugg, M. D., & Taylor, M. J. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology*, *37*(2), 127–152. https://doi.org/10.1111/1469-8986.3720127

Russo, N., Foxe, J. J., Brandwein, A. B., Altschuler, T., Gomes, H., & Molholm, S. (2010). Multisensory processing in children with autism: High-density electrical mapping of auditory-somatosensory integration. *Autism Research: Official Journal of the International Society for Autism Research*, *3*(5), 253–267. https://doi.org/10.1002/aur.152

Schafer, E. W., Amochaev, A., & Russell, M. J. (1981). Knowledge of stimulus timing attenuates human evoked cortical potentials. *Electroencephalography and Clinical Neurophysiology*, *52*(1), 9–17. https://doi.org/10.1016/0013-4694(81)90183-8

Scott-Van Zeeland, A. A., Dapretto, M., Ghahremani, D. G., Poldrack, R. A., & Bookheimer, S. Y. (2010). Reward processing in autism. *Autism Research: Official Journal of the International Society for Autism Research*, *3*(2), 53–67. https://doi.org/10.1002/aur.122

Sinclair, A. H., Manalili, G. M., Brunec, I. K., Adcock, R. A., & Barense, M. D. (2021). Prediction errors disrupt hippocampal representations and update episodic memories.

Proceedings of the National Academy of Sciences of the United States of America, *118*(51), e2117625118. https://doi.org/10.1073/pnas.2117625118

Smith, R., Badcock, P., & Friston, K. J. (2021). Recent advances in the application of predictive coding and active inference models within clinical neuroscience. *Psychiatry and Clinical Neurosciences*, *75*(1), 3–13. https://doi.org/10.1111/pcn.13138

Sokhadze, E., Baruth, J., El-Baz, A., Horrell, T., Sokhadze, G., Carroll, T., Tasman, A., Sears, L., & Casanova, M. F. (2010). Impaired error monitoring and correction function in autism. *Journal of Neurotherapy*, *14*(2), 79–95. https://doi.org/10.1080/10874201003771561

South, M., Larson, M. J., Krauskopf, E., & Clawson, A. (2010). Error processing in high-functioning autism spectrum disorders. *Biological Psychology*, *85*(2), 242–251. https://doi.org/10.1016/j.biopsycho.2010.07.009

Stefanou, M. E., Dundon, N. M., Bestelmeyer, P. E. G., Ioannou, C., Bender, S., Biscaldi, M., Smyrnis, N., & Klein, C. (2020). Late attentional processes potentially compensate for early perceptual multisensory integration deficits in children with autism: Evidence from evoked potentials. *Scientific Reports*, *10*(1), 16157. https://doi.org/10.1038/s41598-020-73022-2

Tavassoli, T., Miller, L. J., Schoen, S. A., Jo Brout, J., Sullivan, J., & Baron-Cohen, S. (2017). Sensory reactivity, empathizing and systemizing in autism spectrum conditions and sensory processing disorder. *Developmental Cognitive Neuroscience*, *29*, 72–77. https://doi.org/10.1016/j.dcn.2017.05.005

Thye, M. D., Bednarz, H. M., Herringshaw, A. J., Sartin, E. B., & Kana, R. K. (2018). The impact of atypical sensory processing on social impairments in autism spectrum disorder.

Developmental Cognitive Neuroscience, *29*, 151–167. https://doi.org/10.1016/j.dcn.2017.04.010

Tremblay, K. L., Ross, B., Inoue, K., McClannahan, K., & Collet, G. (2014). Is the auditory evoked P2 response a biomarker of learning? *Frontiers in Systems Neuroscience*, *8*, 28. https://doi.org/10.3389/fnsys.2014.00028

Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: Predictive coding in autism. *Psychological Review*, *121*(4), 649–675. https://doi.org/10.1037/a0037665

van Laarhoven, T., Stekelenburg, J. J., Eussen, M. L. J. M., & Vroomen, J. (2019). Electrophysiological alterations in motor‐auditory predictive coding in autism spectrum disorder. *Autism Research*, *12*(4), 589–599. https://doi.org/10.1002/aur.2087

van Laarhoven, T., Stekelenburg, J. J., Eussen, M. L., & Vroomen, J. (2020). Atypical visual-auditory predictive coding in autism spectrum disorder: Electrophysiological evidence from stimulus omissions. *Autism, 24*(7), 1849–1859. https://doi.org/10.1177/1362361320926061

Vroomen, J., & Stekelenburg, J. J. (2010). Visual anticipatory information modulates multisensory interactions of artificial audiovisual stimuli. *Journal of Cognitive Neuroscience*, *22*(7), 1583–1596. https://doi.org/10.1162/jocn.2009.21308

West, B.T., Welch, K.B., & Galecki, A.T. (2014). Linear Mixed Models: A Practical Guide Using Statistical Software, Second Edition (2nd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/b17198

Williams, Z. J., Abdelmessih, P. G., Key, A. P., & Woynaroski, T. G. (2021). Cortical auditory processing of simple stimuli Is altered in autism: A meta-analysis of auditory evoked responses. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *6*(8), 767–781. https://doi.org/10.1016/j.bpsc.2020.09.011

Appendix A

Table of Post-Hoc Tests for each ERP Analysis

Appendix A includes post-hoc tests for each of the six ERP analyses, covering both the four-condition and three-condition differential analyses. The first row of the table displays the *p*-values for the interaction between trial type and group. Post-hoc *p*-values for the trial type main effect are also provided when there are non-significant interactions.

Table 2.A1

Post-hoc comparisons for each ERP measure's interaction between trial type & group

Group	Trial type comparison	N1-P2 mean peak-to-peak (*p* = .06)	N1 peak latency (*p* = .04)	P2 peak latency (*p* = .008)	N1-P2 mean peak-to-peak audio out (*p* = .12; ns)	N1 mean peak audio out (*p* = .15; ns)	P2 mean peak audio out (*p* = .04)
ASD	Ao v. AVa	.1	<.0001	<.0001	–	–	–
	Ao v. AVs	<.0001	<.0001	<.0001	–	–	–
	Ao v. AVo	<.0001	<.0001	<.0001	–	–	–
	AVa v. AVs	<.0001	<.0001	<.0001	<.0001	<.0001	.38
	AVa v. AVo	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
	AVs v. AVo	.54	.9824	.005	.45	<.0001	<.0001
NT	Ao v. AVa	1.000	.006	<.0001	–	–	–
	Ao v. AVs	<.0001	<.0001	<.0001	–	–	–
	Ao v. AVo	<.0001	.0003	<.0001	–	–	–
	AVa v. AVs	<.0001	<.0001	<.0001	<.0001	<.0001	.91
	AVa v. AVo	<.0001	<.0001	<.0001	<.0001	<.0001	.002
	AVs v. AVo	1.000	.9998	1.000	.45	<.0001	<.0001
	Ao ASD v. Ao NT	<.0001	1.000	1.000	–	–	–

Table 2.A1

Post-hoc comparisons for each ERP measure's interaction between trial type & group (continued)

ASD v. NT	Ao ASD v. AVa NT	<.0001	.12	<.0001	–	–	–
	Ao ASD v. AVs NT	1.000	.006	.05	–	–	–
	Ao ASD v. AVo NT	1.000	.02	.04	–	–	–
	AVa ASD v. Ao NT	.002	.0001	<.0001	–	–	–
	AVa ASD v. AVa NT	.005	.57	1.000	ns	ns	.007
	AVa ASD v. AVs NT	.88	<.0001	<.0001	ns	ns	<.0001
	AVa ASD v. AVo NT	.88	<.0001	<.0001	ns	ns	<.0001
	AVs ASD v. Ao NT	<.0001	.005	<.0001	–	–	–
	AVs ASD v. AVa NT	<.0001	<.0001	<.0001	ns	ns	.52
	AVs ASD v. AVs NT	.0004	1.000	.38	ns	ns	.967
	AVs ASD v. AVo NT	.0004	.9998	.42	ns	ns	<.0001
	AVo ASD v. Ao NT	<.0001	.0003	.07	–	–	–
	AVo ASD v. AVa NT	<.0001	<.0001	<.0001	ns	ns	.05
	AVo ASD v. AVs NT	<.0001	.993	1.000	ns	ns	.003
	AVo ASD v. AVo NT	<.0001	.9462	1.000	ns	ns	.992

Note that "–" denotes the exclusion of Ao conditions from the N1-P2 difference ERP. Bold font for the group difference section indicates planned comparisons between trial type and group.

Appendix B

Descriptions of the Post-Hoc Tests Between Groups, and *Across* Trial Types

Appendix B describes significant post-hoc tests applied to analyses showing significant effects across trial type and group.

Four trial type N1-P2 peak-to-peak mean amplitude

The NT N1-P2 peak-to-peak mean amplitude response to AVa processing was larger compared to the autistic response toward Ao ($t = 5.1$, $p < .001$, $d = -0.05$), AVs ($t = 9.5$, $p < .001$, $d = -1.39$) and AVo ($t = 10.4$, $p < .001$, $d = -1.39$). The NT response to AVo was also larger compared to the autistic response toward AVs ($t = 4.4$, $p < .001$, $d = 1.19$). The NT response to AVs was also larger compared to the autistic response elicited toward AVo input ($t = 5.3$, $p < .001$, $d = 0.002$). The NT response to Ao processing was also significantly larger compared to the autistic response to AVa ($t = 4$, $p = .002$, $d = 1.07$), AVs ($t = 9.7$, $p < .001$, $d = 2.63$), and AVo ($t = 10.6$, $p < .001$, $d = 2.88$). All other N1-P2 peak-to-peak amplitude comparisons between group and across trial type failed to reach significance (all *p*'s > .88).

N1 Peak latency

The Ao N1 peak latency response in the autistic group was significantly delayed compared to the NT response to AVs ($t = 3.7$, $p = .006$, $d = 0.63$) and AVo ($t = 3.3$, $p = .02$, $d = 0.57$). There were no significant N1 latency differences between the autistic Ao response and NT AVa response ($p = .12$). The N1 latency AVa response in the ASD group was also significantly delayed with respect to the NT response toward AVs ($t = 8.3$, $p < .001$, $d = 1.43$), AVo ($t = 8$, $p < .001$, $d = 1.37$), and Ao ($t = 4.6$, $p < .001$, $d = 0.80$). We also found that the autistic N1 peak latency response toward AVs occurred sooner compared to the NT response to Ao ($t = -3.8$, $p =$

.005, *d* = -0.65) and AVa (*t* = -6.5, *p* < .001, *d* = -1.12). There were no significant differences in N1 latency between the ASD response toward AVs and the NT N1 response toward AVo (*p* = .99). The N1 AVo response in the ASD group also peaked sooner compared to the NT N1 latency response to AVa (*t* = -7.2, *p* < .001, *d* = -1.24) and Ao (*t* = -4.5, *p* < .001, *d* = -0.77). Lastly, there were no significant differences between the N1 AVo response in the ASD group and the NT N1 response to AVs (*p* = .99).

P2 Peak latency

The Ao P2 peak latency response in the autistic group was significantly delayed compared to the NT response to AVs (*t* = 3, *p* = .05, *d* = 0.69) and AVo (*t* = 3.1, *p* = .04, *d* = 0.70), but peaked sooner compared to AVa processing (*t* = -8, *p* < .001, *d* = -1.82). The AVa response in the ASD group was delayed compared to the NT response to AVs (*t* = 11.2, *p* < .001, *d* = 2.53), AVo (*t* = 11.2, *p* < .001, *d* = 2.55) and Ao (*t* = 8.3, *p* < .001, *d* = 1.88). We found that the autistic P2 peak latency response toward AVs occurred sooner compared to the NT response to Ao (*t* = -5, *p* < .001, *d* = -1.14) and AVa (*t* = -13.2, *p* < .001, *d* = -3). There were no differences in the autistic P2 peak latency response toward AVs and the NT P2 response toward AVo (*p* = .42). The AVo P2 response in the ASD group also peaked sooner compared to the NT response to AVa (*t* = -11.1, *p* < .001, *d* = -2.52) and Ao (*t* = -2.9, *p* = .07, *d* = -0.66), but was not different compared to the NT response to AVs (*p* = 1.0).

P2 Differential Mean Amplitude

We found that the autistic P2 differential mean amplitude response toward AVa was significantly larger compared to the NT response toward AVo (*t* = 7.3, *p* < .001, *d* = 1.03), but was not different compared toward the NT response to AVs (*p* = .1). The autistic P2 differential response toward AVs was larger compared to the NT response to AVo (*t* = 5.6, *p* < .001, *d* =

0.79), but there were no differences compared to the NT response toward AVa (*p* = .35). Lastly, the autistic P2 differential response toward AVo was reduced compared to the NT response to AVs (*t* = -3.9, *p* = .002, *d* = -0.54) and AVa (*t* = -2.7, *p* = .07, *d* = -0.38) processing.

Appendix C

Addition Figures of N1/P2 Latency and N1-P2 Peak-to-Peak Differential Mean Amplitude

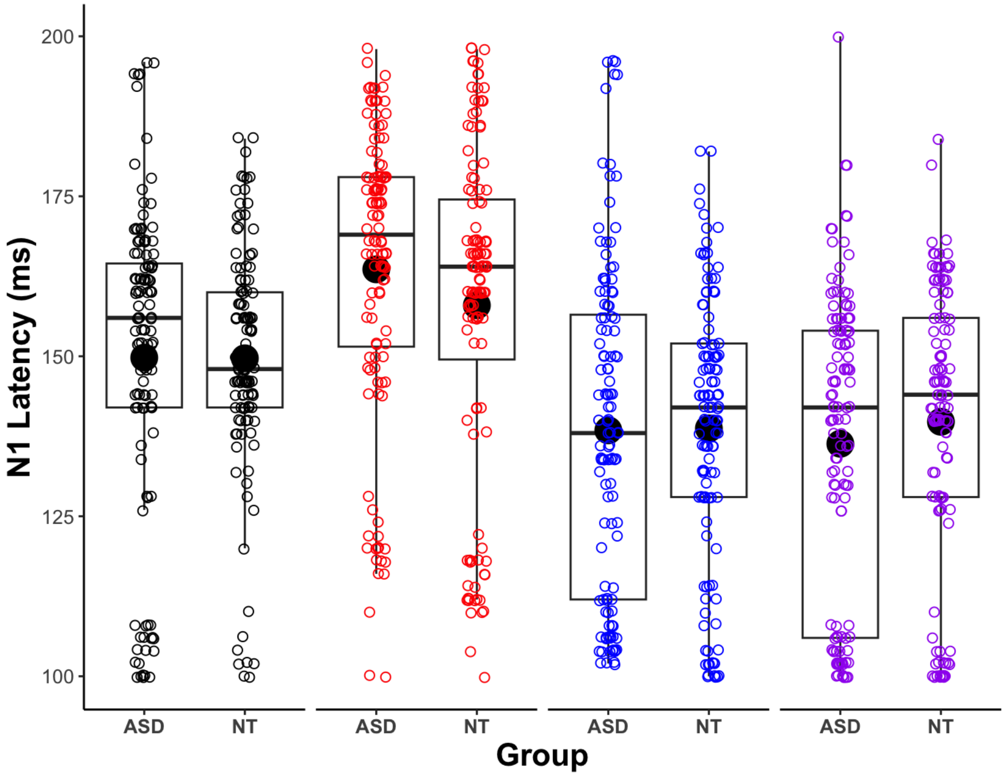

Figure 2.C1. *N1 peak latency between autistic and NT groups, and across trial types.* For the latency scatter plot, the y-axis represents time in milliseconds (ms), while the x-axis is trial type, split between groups.

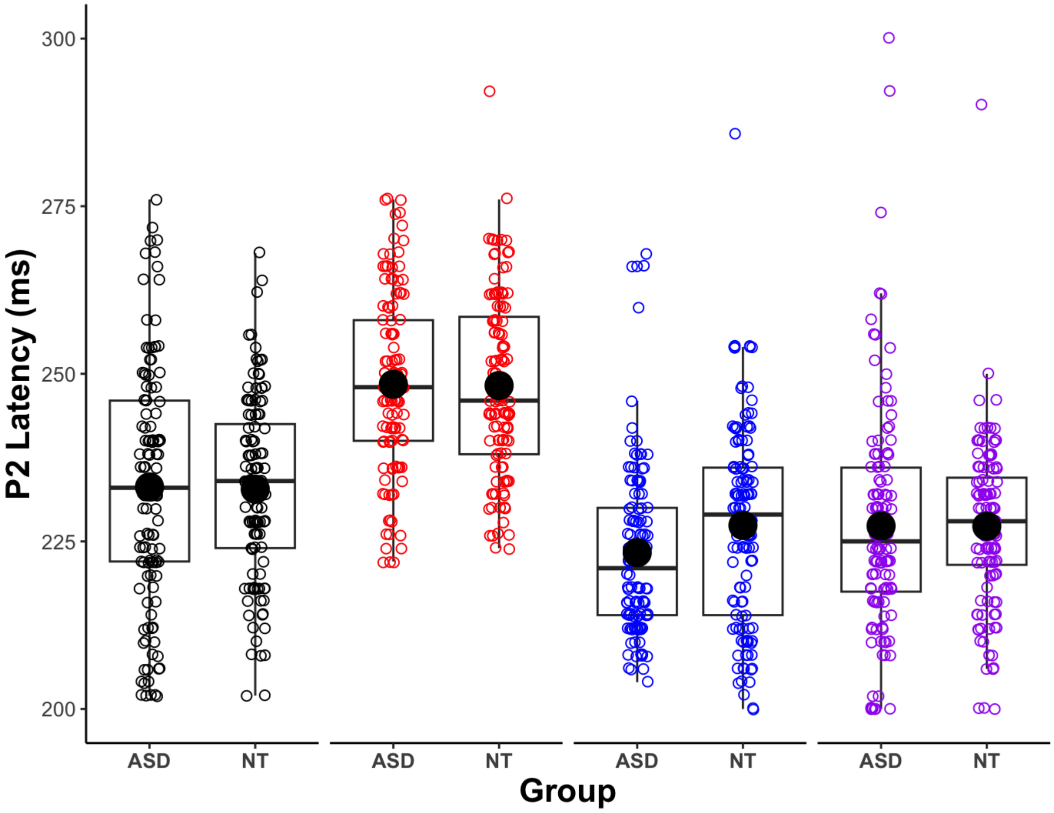

Figure 2.C2. *P2 peak latency split between autistic and NT groups, and across trial types.*

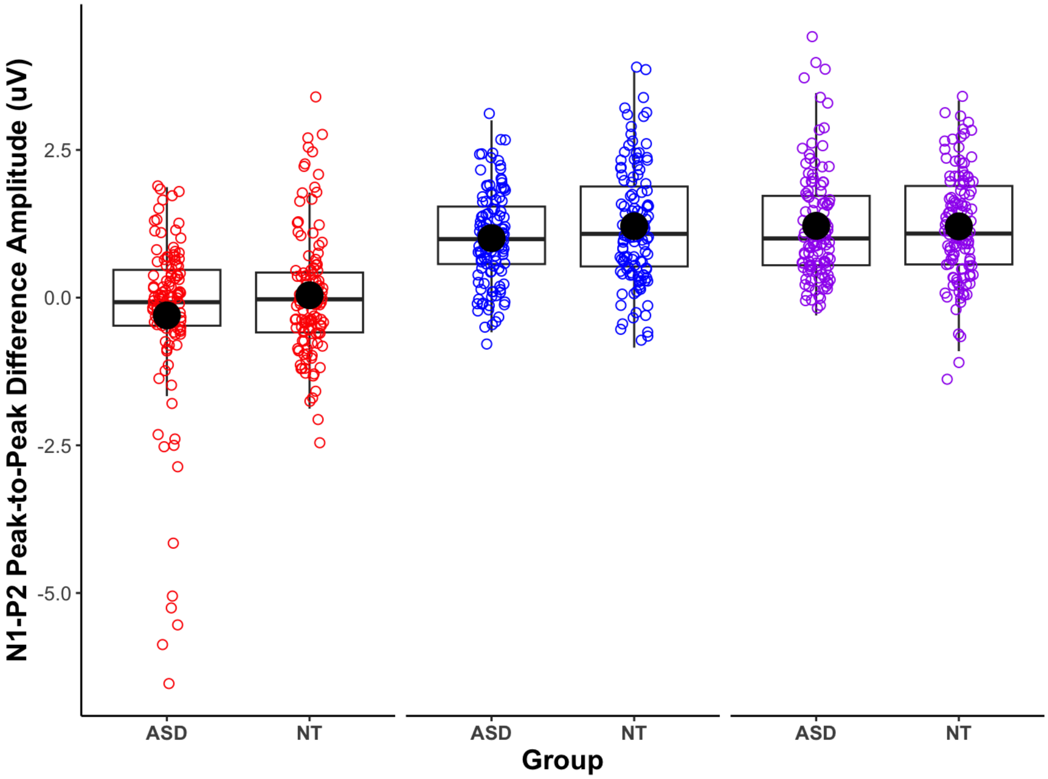

Figure 2.C3. *N1-P2 peak-to-peak mean difference amplitude between autistic and NT groups and across the three AV trial types.* This figure displays scatter plots illustrating the N1–P2 peak-to-peak mean difference amplitude for each participant. On the scatter plot figure, the y-axis represents the difference amplitude, with more positive values indicating greater response reduction relative to audio-only processing, while the x-axis represents trial type. Note that this figure depicts the non-significant interaction between trial type and group.

Chapter 2, in part is currently being prepared for submission for publication of the material. Marin, Andrew; Pearson, Lucy; Wu, Mincong; Baker, Elizabeth; Carver, Leslie J. The dissertation author was the primary investigator and author of this material.

Infant Sensitivity Toward the Timing of Sounds Predicted by a Dynamic Visual Object

Andrew Marin[1], Melanie Dratva[1], Zara Fearns[1], Lindsey J. Powell[1], Viola S. Störmer[1,2], &
Leslie J. Carver[1]

[1] University of California, San Diego (UCSD) – Psychology Department

[2] Dartmouth College – Department of Brain and Psychological Sciences

Author Note

Correspondence concerning this article should be addressed to Andrew Marin, Department of Psychology, 9500 Gilman Drive, McGill Hall, La Jolla, CA, 92093, USA. Email: amarin@ucsd.edu

Acknowledgements

We would like to thank Alexis Smith-Flores for a thoughtful edit of this manuscript and helpful feedback when designing this study. We would like to thank Gabriela Bernal for coordinating this study, and Alexander Nam, Carmen Chen, and Xiaoyang Liu for their help with data collection and behavior coding.

Abstract

Everyday perceptual experiences include dynamic objects that move and give rise to expected sounds. In this study, we asked if four- to five-month-old infants are sensitive to the temporal synchrony of sounds generated by a moving visual object. In a moderated, online study, we showed twenty-four infants a novel experimental paradigm featuring a moving 2-D ball and bounce sounds. We presented each infant with alternating trials from two conditions that manipulated the temporal congruence of the ball's motion and the sound: one in which the bounce sound occurred simultaneously upon physical collision between the ball and a barrier (audio-visual; AV-synchronous), and one in which the sound came slightly before collision (AV-asynchronous). Using offline behavioral coding of recorded looking time, we show that infants looked longer toward AV asynchronous trials compared to synchronous ones. This data demonstrates that young infants have expectations regarding the temporal alignment of visual and auditory properties of physical events over the course of a single familiarization phase. The presence of early cross-modal expectations may provide a foundation for physical inference and learning in natural sensory environments.

Keywords: Infancy, expectation, audio-visual, synchrony, cross-modal

Infant Sensitivity Toward the Timing of Sounds Predicted by a Dynamic Visual Object

The natural world is filled with dynamic visual information that is strongly linked with associated sounds. Imagine moving fingers striking the keys of a piano, the back and forth bounces of a ball in a tennis match, or a person clapping their hands: Each of these visual experiences predict accompanying sound via the integration of top-down knowledge structures related to physical environments. In the case of a ball bounce, the dynamic visual movement of the ball generates strong auditory expectations of a bounce sound as it moves toward physical collision with a hard surface. Past computational research has found that neurotypical adults use their understanding about the physical world to inform their inferences about past (Smith & Vul, 2014; Gerstenberg et al., 2018) and present (Gerstenberg et al., 2021) events, and to anticipate events in the future (Smith et al., 2013; Battaglia et al., 2013). Adults can also reconstruct what could have happened in the past based on integrating visual and sound information (Gerstenberg et al., 2012) and can infer motion trajectories of falling objects colliding with angled surfaces (Little & Firestone, 2021). Thus, abstracting about the plausibility of physical events - within and across modalities - seems to occur rapidly and relatively effortlessly for adults; however, it is not yet understood how and when these abilities emerge.

Previous work suggests that early knowledge structures about the physical world emerge during infancy (Wellman & Gelman, 1992; Baillargeon et al. 2012; White, 2014; Saxe & Carey, 2006). This includes developing sensitivities toward the physical properties of objects (for review see, Spelke and Kinzler, 2007), which may provide foundational skills for infants to form higher order inferences about cause and effect relations in real-world sensory environments (Spelke et al., 1992; Ullman & Tenenbaum, 2020). For example, early representations of the movement of dynamic objects help to organize the infant's understanding of how objects behave within the physical world. Infants as young as three months of age make more accurate and faster anticipatory eye movements toward predictable visual patterns (i.e., flashes of light moving left to right) versus irregular ones (i.e., flashes of light moving randomly; Haith et al., 1988). Importantly, the infant's ability to anticipate motion trajectories improves with experience (Johnson et al. 2003). Infants also spend more time attending to perceptual events that violate their expectations of how physical objects behave (Spelke, 1985; see review, Margoni et al., 2023). Infants are aware that objects remain cohesive, even when partly occluded (see review, Spelke, 1990; Kellman & Spelke, 1983) and that solid objects cannot pass through space occupied by other solid objects (Baillargeon et al., 1985; Spelke et al., 1992); therefore eliciting higher looking times towards events in which physical properties are violated. These findings suggest that infants have expectations of objects' behavioral dynamics due to sensitivities towards their physical properties (for review, Spelke & Kinzler, 2007) and that these skills gain sophistication throughout development (for review, Bremner et al., 2014). Infants then use these knowledge structures to build more complex abstractions about how objects behave in natural sensory environments.

In adults, inferences about behavioral dynamics of objects are tied to psychological determinants that govern the perception of causal launching events. Michotte (1946, 1963) first demonstrated this by showing adults a moving object which contacted a stationary object to launch it into motion. When the timing or spacing of the launch event varied, adults did not ascribe physical collision to be the cause of the launch event. Similar findings have been extended to young infants (e.g., four months of age; for review, Saxe & Carey, 2006) such that they look longer toward physically implausible, non-contact launching events after being habituated toward physically plausible launches (Leslie, 1984; Ball, 1973). Furthermore, sevento ten-month-old infants also look longer when encountering a temporal delay between physical collision of two objects and the subsequent launching event (for review, Scholl & Tremoulet, 2000; Kotovsky & Baillargeon, 2000), and towards launching events where the relative speed of one object does not match the second object's speed when it is subsequently launched into motion (Kominsky et al., 2017). Each of these studies suggested that a violation of the infants' learned expectations elicited longer look times toward implausible events, which in turn demonstrates that infants have pre-existing representations of visual object interactions. However, perceptual causality signals are not limited to visual collision events; as dynamic visual objects signal predictable, casual sound upon impact with another object or a boundary.

The ability to predict sound based on dynamic visual cues, like collision, depends on lower-level perceptual skills dedicated to detecting the alignment of audio-visual (AV) signals across time and space. For example, research in adults has shown that AV signals closely aligned in both space and time are more likely to be perceived as integrated compared to those that are not (Wallace et al., 2003; Körding et al., 2007; Parise et al., 2012). Infants are also able to integrate AV signals soon after they are born. Four-month-old infants, when habituated to two objects that create different rhythmic sounds, will look longer toward the object responsible for each sound upon hearing it, suggesting that they can match the rhythms of sounds that have been contingently paired with objects that create them (Spelke, 1976). Infants also prefer to look toward AV input that display synchronous temporal and spatial alignment within the first few months (Spelke, 1979; Bahrick, 1983), or even weeks (Bahrick, 2001) of their life. As discussed in Bahrick and Lickliter (2000), the coordinated timing of sensory inputs across different modalities serves as an attentional cue and facilitates perceptual learning in infancy, especially during language acquisition. Infants sensitivity toward the correspondence between AV speech signals occurs within the first six months of life (Kuhl & Meltzoff, 1982, 1984), including that infants prefer to attend to synchronous lip movement-to-voice presentations rather than those out of synchrony (Dodd, 1979), expect the spatial output of their mothers' voice to match the location of their mothers' face (Aronson & Rosenbloom, 1971), and demonstrate the McGurk effect, where the specific movement of the lip influences which phoneme they expect to hear (Rosenblum et.al., 1997). Furthermore, infants are able to remember word-object relations better when information is presented synchronously, rather than asynchronously (Gogate & Bahrick, 2001), and selectively attend to mouths when speech is synchronous (Hillairet de Boisferon et. al., 2017), and can detect AV asynchrony pertaining to speech even without rhythmic cues (Lewkowicz, 2003).

Synchrony biases like these have been theorized to facilitate cross modal learning across language and cognitive domains in infancy (see review, Oakes 2010; Poli et.al., 2020). Aside from noticing synchrony of visual input and sound when learning language, four-month-olds have shown a synchrony bias when beginning to reason about the composition of moving objects, expecting rigid versus elastic objects to behave differently upon collision (Bahrick, 1983). Furthermore, one study found that newborn infants exhibit greater looking time towards an object which is moving in a direction that matches an accompanying sound (i.e., an increase in volume when moving towards them), compared to misaligned AV information (Orioli, 2018). These studies suggest that AV integration, in the form of a synchrony bias, is a crucial part of infant learning when they are being exposed to unfamiliar sensory associations. Neural evidence also supports the emergence of infant synchrony biases as they learn to integrate AV input with experience, such that repeated exposure to AV contingencies attenuates neural responses compared to unimodal input (Kersey & Emberson, 2017). Although the precise mechanisms driving synchrony biases in infancy remains unclear, the bias toward AV synchrony can be contextualized within the framework of infant-learned associations, which contrasts the perceptual biases toward asynchrony that stem from expectation violations (Roder et al., 2000; Colombo & Mitchell, 2009).

In infancy, synchronous and asynchronous AV perceptual biases also emerge when processing dynamic objects interacting within physical environments. For example, it has been demonstrated that infants are sensitive to small asynchronies when viewing ball bounce collision events (Lewkowitz, 1996). Here, infants were first habituated to a 2-D ball that moved up and down, and made a sound when the ball made physical contact with the bottom boundary of the display. During the test phase, infants were shown the same visual event with varying spatiotemporal offsets (i.e., 250 milliseconds (ms), 300ms, and 350ms) between sound onset and visual collision, and were found to dishabituate toward the AV stimuli when presented with the 300ms offset interval. Four-month-olds are also sensitive toward violations of expected relations between object-sound numerosity, in that they look longer toward perceptual events where the number of dynamic visual objects (i.e., one ball bouncing) differed from the number of sounds elicited (i.e., two bounce sounds heard; Smith et al., 2017). These findings suggest that infants form expected associations about the number of visual objects and sounds based on cues of harmonicity, and when this expectation is violated, they look longer at the mismatching stimuli. These behavioral results also appear to reflect a contrast to the contexts where infant synchrony biases emerge. When abstracting about new associations about physical environments that infants have not had much experience with, infants prefer synchronous AV events (Bahrick, 1983; Orioli, 2018), whereas when viewing perceptual events that violate previously held expectations (Smith et al., 2017; Lewkowicz, 1996), infants show the opposite effect and prefer to look at perceptual events that violate their physical expectations.

One factor not considered in Smith et al. (2017) and in Lewkowitz (1996) was whether infants are able to detect a mismatch between AV inputs solely through brief exposure to lower-level spatial and temporal cues, without prior exposure. Therefore, the primary objective of the present study was to test whether infants have pre-existing expectations about whether moving objects elicit an immediate sound upon collision with a boundary. We were specifically interested in whether this sensitivity is present naturally, without introducing a habituation phase within the same experiment. Additionally, whereas Smith et al. (2017) and Lewkowitz (1996) both used stimuli where a ball appeared to be acted on by gravity (i.e., traveled downwards and then bounced back up), our stimuli was designed to mimic naturalistic movement that the infants would have to anticipate, as the ball traveled within the display in a self-propelled manner at a consistent speed. To investigate this, we measured the looking time of four- to five-month-old infants as they viewed a 2-D ball that moved along a physically plausible motion path to bounce off and make a sound upon colliding with a boundary. For synchronous AV trials, sound accompanied the ball's dynamic motion precisely when it collided with the edges of the display, while for the asynchronous AV trials, the sound occurred slightly before visual collision. We expected misalignment between visual and auditory input to elicit a violation of expectation, as the perception of temporal and spatial simultaneity of discrete AV events plays a large role in determining if the two sensory events will be perceptually integrated as one, or perceived as two separate events (Wallace et al., 2003; Körding et al., 2007; Parise et al., 2012). We were guided by the pre-registered prediction that infants would show greater looking time on average toward AV asynchronous presentations compared to synchronous ones, due to a violation of their expectations about the spatiotemporal relations between dynamic visual objects and collision sounds. Alternatively, greater looking times toward the AV synchronous presentations would indicate the presence of a familiarity bias toward the learned regularities of dynamic AV events, which could indicate that the infants are still learning about the alignment between visual collision and sound. No looking time differences could indicate that the infants are too young to notice the small differences in spatiotemporal information we used in our design, and therefore this association may occur with more experience. Additional analyses ruled out order effects, side biases, and color preferences driving these results.

Methods

Participants

Our pre-registered sample consisted of 24, four-to-five-month-old infant participants (age range: 4 months 1 day - 4 months 29 days; M_{age} = 4.39 months; *SD* = 0.26; 15 female) who participated in this virtual experiment with their primary caregiver over a Zoom call. An additional six infants were recruited and participated in the study, but were excluded from analysis based on our pre-registered exclusion criteria (see looking time processing for specific exclusion criteria). Each primary caregiver and their infant were recruited using social media advertisements and were not compensated for participating. All remote data collection procedures took place in San Diego, California between April 2022 and July 2022. Participants came from California (*n* = 9), Ohio (*n* = 2), Massachusetts (*n* = 2), Washington (*n* = 1), New York (*n* = 2), North Carolina (*n* = 1), Louisiana (*n* = 1), South Carolina (*n* = 1), Maine (*n* = 1), New Mexico (*n* = 1), Nebraska (*n* = 1), Colorado (*n* = 1), and Indiana (*n* = 1). The ethnic breakdown, primary caregiver level of education, and total household income of our infant sample can be seen in table one.

Table 3.1
Frequency of Demographic Categories

	Frequency (*n* = 24)	Percentage
Infant Race		
White	16	66.7
Hispanic or Latino	4	16.7
Asian or Pacific Islander	1	4.2
Black	0	0.0
Mixed Race	3	12.5
Primary Caregiver Level of Education		
High School Diploma	0	0.0
Associate Degree	3	12.5
Bachelor’s or Undergraduate Degree	9	37.5
Master’s Degree	7	29.2
Doctorate	5	20.8
Total Household Income		
Not Reported	1	4.2
$40,000 - $60,000	3	12.5
$60,000 - $80,000	2	8.3
$80,000 - $100,000	3	12.5
$100,000 or more	15	62.5

We screened each infant via primary caregiver self-report prior to their participation to confirm that the infant had normal hearing, corrected-to-normal or normal vision, no known neuropsychological, intellectual, developmental or genetic disorders, and experienced no significant prematurity (i.e. at least 37 full weeks of gestation). We also ensured that the primary caregiver joined the experiment from a laptop or desktop computer with a forward-facing webcam at the top of their display. All procedures were approved by a local institutional review board (IRB) and each primary caregiver provided informed consent on behalf of themselves and the infant.

Audio-Visual Stimuli

The AV stimuli for this experiment were created using Blender 2-D Animation (Community, 2018; version 2.8) and Movavi Video Editor Plus 21.3.0 softwares, and were exported with a frame rate of 24 frames per second using a 1920x1080 pixel resolution. The visual stimuli consisted of a single green 2-D ball that moved continuously to bounce off artificially defined boundaries of a black rectangle, which was overlaid on top of a neutral gray background (see Fig. 1). For each trial, the ball appeared at either the top left or right side of the black rectangle and began to move diagonally toward the bottom of the black rectangle. The green ball then appeared to bounce off upon visual collision with the bottom boundary and continued its dynamic motion along a path that depended on the trajectory of the previous bounce. The ball traveled across the rectangular display with a trajectory designed to avoid corners, so that on average the ball collided with one of the four boundaries every 2.05 seconds (*SD* = 0.93s), as the distance traveled varied by bounce.

Visual angles of the stimuli were estimated using 1) a viewing distance of 50 cm, which was the approximate viewing distance for our participants, and 2) a standard laptop monitor size of 30.4 cm x 21.2 cm, which was the typical computer set up used for each session. Note that we were unable to obtain an exact viewing distance and monitor size for each infant, and that visual viewing angle between each infant varied. Based on these approximations, the green ball, which was 5.43 cm in diameter, subtended a visual angle of about 2.9 degrees. The ball's movement was constrained to occur within the boundary of a black rectangle which measured approximately 7.93 inches in width and 5.3 inches in height (approximate visual angle $_{\text{height}}$ = 15.1 degrees; visual angle $_{\text{width}}$ = 11.4 degrees; subtended rectangular visual angle $_{\text{height*width}}$ = 3.01 degrees). Auditory stimuli were then embedded into the 2-D animation (using Movavi) to provide the perception of bounce sound when the green ball collided with one of the four boundaries of the black rectangle. The sound itself was a 50 decibel complex tone that resembled a solid object colliding with a hard surface (i.e., a knocking sound) and had a duration of 110 ms.

The experiment contained three conditions in total: AV synchronous, AV asynchronous, and AV surprise (see Figure 1 for a visual depiction of each condition).

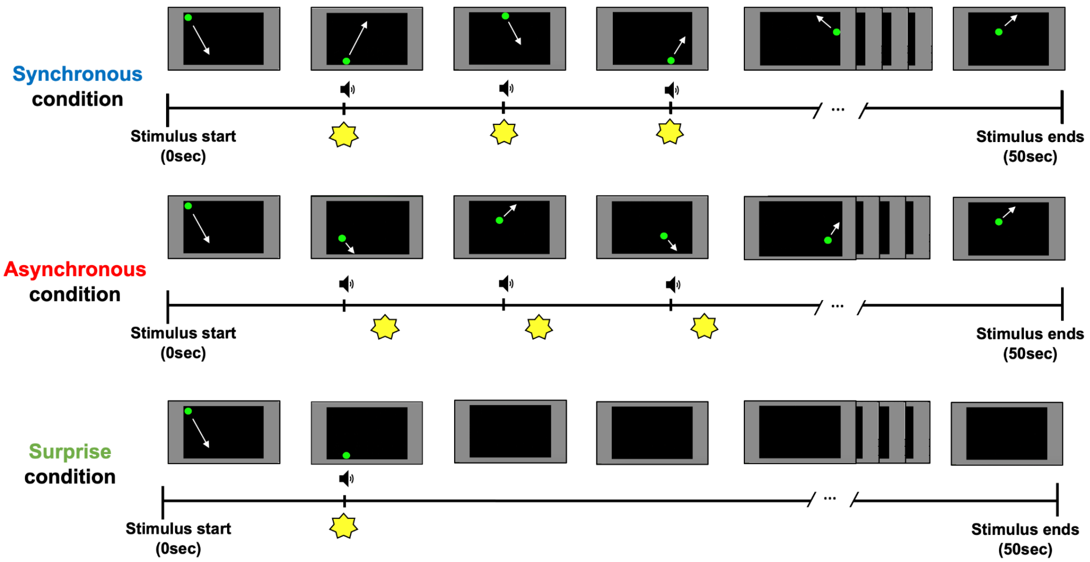

Figure 3.1. *Visual depiction of each audio-visual (AV) condition.* Each condition contained a 2-D green circle, which initially appeared at the top left or right corner for each stimulus condition (left-side start depicted above). The dynamic movement of the ball is illustrated by the white arrows (white arrows were not present in the actual experimental stimuli). The timeline below each condition shows the timing of the relevant visual and auditory stimuli. Sound onset is represented by the audio icon while visual collision between the ball and boundary is depicted by the yellow stars. In the surprise condition, the ball continued along its trajectory instead of bouncing, thus appearing to pass through the boundary upon first collision and disappearing behind the display.

For AV synchronous trials, the ball's dynamic motion was accompanied by a bounce sound precisely when it collided with one of the boundaries of the black rectangle. For the AV asynchronous condition, the bounce sound occurred approximately 250ms (+ or - 25ms jitter) before visual collision with a boundary. Although Lewkowicz (1996) found that infants are sensitive to AV asynchronies with a minimum offset of 300ms, Knopp et al. (2014) found neural evidence that infants are sensitive to AV asynchronies using a 200ms offset. Thus, we opted to use 250ms (+ or - 25ms) AV offsets in our study. For the AV surprise trials, a single bounce sound occurred at the point of collision with the lower boundary of the rectangle, but the ball continued along its downward motion path through the lower boundary and disappeared instead of bouncing. For the AV surprise trials, after the ball disappeared, it would not reappear for the remaining duration of the trial. During piloting, we observed that infants were not engaged with the AV surprise condition due to the ball disappearing. However, the novelty of the surprise trials did result in regained interest during the latter half of the experiment. Based on these preliminary observations, we opted not to analyze looking times using this condition, but kept the AV surprise trials in the experiment to improve attention during the second half of the presentation. Each AV stimulus video was up to 50 seconds long, but the trial length during the experiment was infant-controlled (see procedure section below). Additionally, the initial starting location for the ball was counterbalanced so that the ball could initially appear on either the left or right side of the black rectangle. This was accomplished by mirroring each video so that each condition had a left and a right version with the ball following the same, but reflected, motion path.

Procedure

This experiment had two run orders which both contained 12 trials (five AV synchronous, five AV asynchronous, and two AV surprise trials). Each infant participant was randomly assigned into one of two run orders to control for initial presentation effects, where 12 infants saw an AV synchronous presentation first, and the other 12 saw an AV asynchronous trial first. Trials alternated between AV synchronous and AV asynchronous presentations, with two AV surprise trials inserted in the 5th and 9th position in both run orders.

This experiment was conducted virtually over Zoom, where two experimenters were present with the infant participant and their primary caregiver. Upon joining the Zoom session, the primary experimenter welcomed each infant participant and their caregiver and provided verbal instructions to the caregiver. The experimenter explicitly instructed the caregiver to minimize external distractions in the environment in which the infant would be viewing the stimuli (e.g., closing windows, keeping pets and other people out of the room, moving toys out of the infant's reach, turning off their computer notifications, etc.). The infant participants watched the stimuli at a normal viewing distance (i.e., approximately 50cm) and the caregiver's were given the option to choose out of our recommended viewing placements: seated on the lap of their primary caregiver (*n* = 18), standing on the caregiver's lap (*n* = 1), seated in a carrier seat or high chair (*n* = 3), or seating on the ground in front of the device (*n* = 2). The caregiver was also instructed to keep the infant as still as possible, but not to distract the infant or redirect the infant's attention if the infant looked away.

The virtual experimental sessions followed the same infant controlled looking time procedure as outlined in Smith-Flores et al. (2021). In this procedure, two experimenters, who were in different locations but on a phone call with each other, joined a video call with the infant participant and caregiver. The primary experimenter controlled the stimuli presentation via a web-based presentation program (Slides.com), and the secondary experimenter, who was blind to run order and trial type, live coded infant looking time while only viewing the video stream of the infant using PsychoPy (Peirce, 2007) and PyHab (Kominsky, 2019). At the start of the experiment, each infant viewed a 5-point calibration video where a rotating disk accompanied by sound directed the attention of the infant to the center and the four corners of the display. The calibration established the boundaries of the participant's display so that the experimenters were able to distinguish whether the infant looked towards the edges of the display or away from it. Prior to the presentation of each AV trial, an attention grabber appeared. This fixation stimulus consisted of a small spinning blue star presented in the center of the screen that made a chime sound and played continuously. The fixation stimulus was designed to promote infant engagement between trials and to orient their gaze to the center of the display. When the primary experimenter determined that the infant had looked at the fixation stimulus for approximately two seconds, the experimenter advanced the slide. This ended the fixation stimulus and its accompanying chime sound, and the experimental stimuli appeared (see Figure 2 for a schematic of the experimental time course).

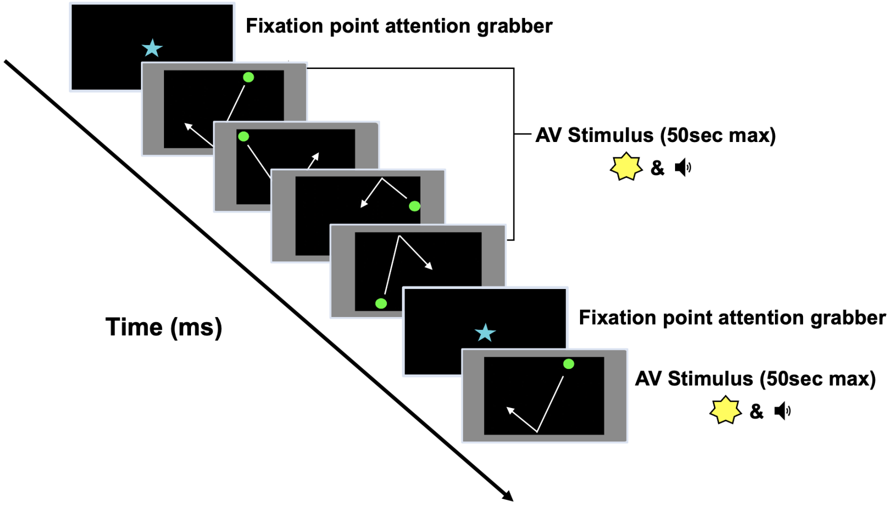
Fixation point attention grabber
AV Stimulus (50sec max)
&
Time (ms)
Fixation point attention grabber
AV Stimulus (50sec max)
&

Figure 3.2. *Diagram of the time course between experimental trials.* Each experimental trial began with an attention grabber consisting of a spinning blue star at the center of the screen that made a chime sound. The AV stimulus was presented after the attention grabber ended. The duration of the attention grabbers and experimental trials were both infant controlled. The attention grabber ended upon 2s of infant gaze as determined by the primary experimenter and the AV trial ended upon 2s of infant inattention as determined by the secondary experimenter. Sounds are represented by the audio icon while visual collision between the ball and boundary is depicted by the yellow stars. The maximum duration of each experimental trial was 50 seconds.

The cessation of the fixation chime sound served as a cue to the secondary experimenter to initiate a button press on their keyboard while viewing a live feed of the infant, indicating the start of stimulus presentation for a given AV trial. The secondary experimenter held down this key the entire time the infant was looking at the screen, and if the infant looked away from the screen, the secondary experimenter released the key. If the secondary experimenter’s key remained unpressed for 2 seconds, their computer generated an external sound that was heard by the primary experimenter through the phone call with the secondary experimenter. Upon hearing the tone, the primary experimenter ended the AV trial and advanced to the next fixation stimulus, and the procedure repeated itself.

Looking Time Processing

The primary looking time measure was determined by a separate offline coder, who was not involved in the collection of the live looking time data, and was blinded to condition and run order. This offline coder viewed the recording of each session and coded the infant's looking time using Datavyu (Datavyu Team, 2014). Looking time in seconds was calculated separately for each AV trial. Raw looking time per AV trial could be up to 50 seconds if the infant faced the screen throughout the entire trial, or the total duration they looked at the stimuli until a two-second look-away, which ended the trial. Any periods of infant look-aways that lasted less than two-seconds were treated as inattention and subtracted from the total trial looking time. As recommended for looking-time data (Csibra et al., 2016), each infant's cumulative looking time in seconds for each trial was log transformed. For each participant, we then averaged the logged looking time across all valid trials in a condition. A separate trained research assistant also coded each video to ensure reliability of the single offline coder. Both coders displayed high inter-rater reliability when determining if a given trial was good or bad based on the exclusion criteria outlined below (Kappa = .83, overall agreement = 93%). We also assessed inter-rater reliability between the two coders in determining average looking times in seconds collapsed across all the usable synchronous and asynchronous conditions in the experiment for each infant. The analysis revealed average looking time determinations between the two coders were highly correlated, $r(22) = 0.97, p < 0.001$.

Individual trials were excluded if: the infant cried for more than 10 seconds within a trial (n = 3 trials removed); the caregiver or other entity interfered with the infant's looking (e.g., a sound from the caregiver distracts the infant, caregiver points back to the screen, etc.; *n* = 7 trials removed); there were technical issues (*n* = 2 trials removed); or newly introduced environmental stimuli (e.g., a siren outside, a dog barking, etc.) captured the attention of the infant for longer than two seconds (*n* = 3 trials removed). For a given participant, individual trials were also excluded if the total accumulated look time was less than two seconds (*n* = 15 trials removed). Furthermore, if the offline coder determined that the live experimenters ended the trial before infant inattention (e.g., stimulus was advanced during online coding before the infant looked away for two seconds), that trial was dropped from analysis (*n* = 4 trials removed). Five infants did not complete the full 12 trial exposure phase due to global fussiness issues, but were included in the analysis. A total of 11 trials were not presented across these five infants. Thus, all individual trial exclusions (*n* = 34) resulted in a total of 195 (of 229 total possible trials; 85.2% retention rate) usable trials collapsed across the 24 infants in our sample.

After excluding individual trials, participants were only included in the analysis if they completed a minimum of two valid trials of both the AV synchronous and AV asynchronous conditions (after Smith et al., 2017). Infants (*n* = 3) who did not complete at least two valid trials per condition were excluded from the analysis and replaced with a new participant. Additionally, participants who experienced any technical difficulties (*n* = 3) which impacted the experimental presentation or the recording quality of the Zoom session (e.g. internet connectivity, display issues, low-resolution recordings, etc.) and lead to uncertainty in behaviorally coding the infant's gaze, were excluded from analysis and replaced with a new participant.

There were a total of 98 valid AV asynchronous trials and 97 valid AV synchronous trials collapsed across the 24 infants in our sample. The number of valid trials collapsed across the two conditions of interest from the 24 infants in our sample ranged from 5 to 10 trials ($M_{usable\ trials}$ =

8.13, $SD_{usable\ trials}$ = 1.6). We also conducted a pre-registered independent samples *t*-test to examine differences in the number of usable trials between the two run orders. We found that the total number of usable trials was not significantly different between the infants who saw the AV asynchronous trial first ($M_{usable\ trials}$ = 8.08, $SD_{usable\ trials}$ = 1.68), compared to those who saw the AV synchronous first ($M_{usable\ trials}$ = 8.17, $SD_{usable\ trials}$ = 1.64; $t(21.9) = 0.12$, $p = 0.9$). We also conducted a pre-registered paired samples *t*-test to assess whether there were significant differences between the total number of usable AV synchronous and AV asynchronous trials used in computing each infant's averages. This analysis revealed non-significant differences between the amount of AV synchronous ($M_{usable\ trials}$ = 4.04, $SD_{usable\ trials}$ = 0.86) and AV asynchronous trials ($M_{usable\ trials}$ = 4.08, $SD_{usable\ trials}$ = .97; $t(23) = 0.24$, $p = 0.81$). Thus, any differences in subsequent analyses were not driven by the number of trials seen between the two run orders and between the two conditions of interest.

Results

Our primary dependent measure was the logged average looking time for valid AV synchronous and AV asynchronous presentations. We conducted a pre-registered 2x2 mixed factorial ANOVA to examine the prediction that infants will look longer towards AV asynchronous presentations compared to AV synchronous ones. In our ANOVA model, we treated the averaged logged looking times for the AV synchronous and AV asynchronous presentations as a two level within-subjects factor, while the two run orders consisted of a two level between-subjects factor. As predicted, we found a trending main effect between AV trial type, $F(1, 22) = 3.61$, $p = 0.07$, partial $\eta^2 = .14$, where infants spent greater time looking toward the AV asynchronous ($M = 1.06$, $SD = 0.26$) presentations compared to AV synchronous trials ($M = 0.98$, $SD = 0.27$; see Figure 3).

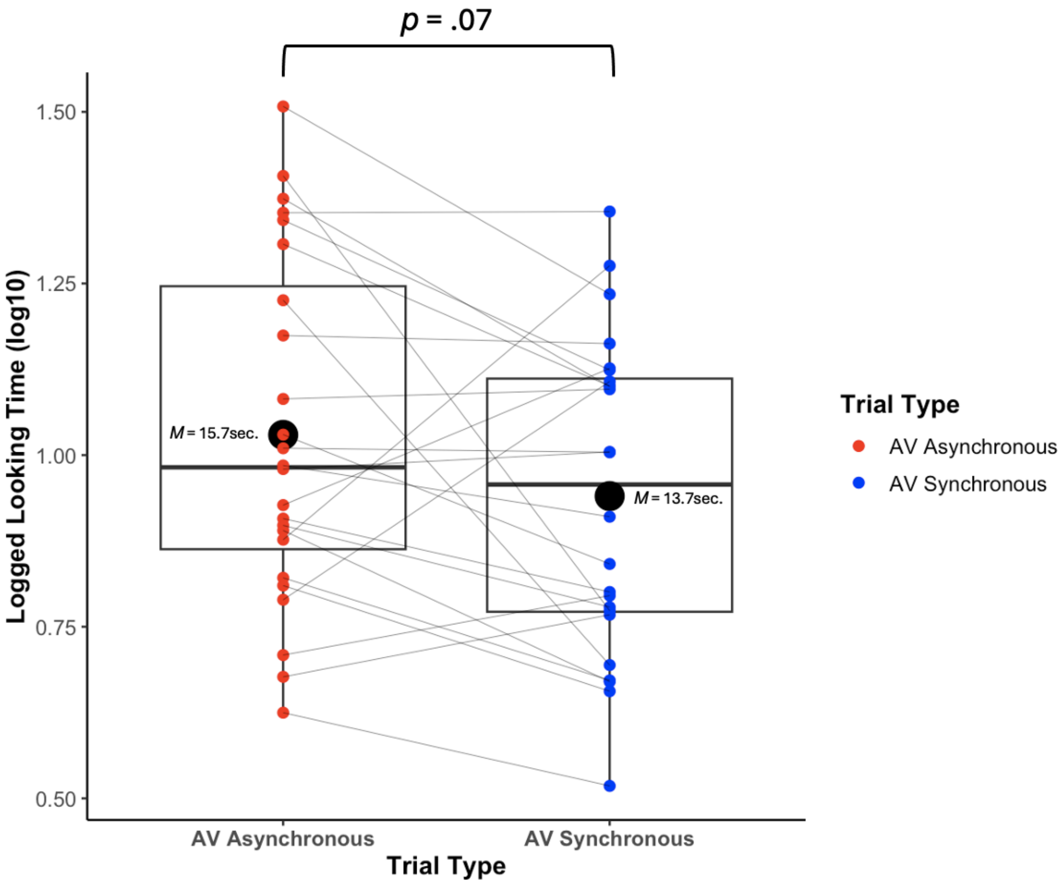

Figure 3.3. *Scatter plot of looking time averages (log10 transformed) between the AV asynchronous and AV synchronous conditions.* The above figure depicts that on average, four-to-five-month-old infants look longer toward temporally asynchronous AV input compared to synchronous AV presentations. The y-axis represents average looking time in logged seconds, while the x-axis represents the two conditions of interest. The individual average logged looking time values for the AV asynchronous condition are depicted by the red circles and the AV synchronous values are presented by the blue circles. The solid gray lines connecting each pair of red and blue circles represents the average logged looking times between AV conditions for each infant participant. The black horizontal lines within each boxplot are group level median averages of logged look time for the AV asynchronous and AV synchronous conditions. The large black scatter point represents the mean logged reaction time for each condition. Raw average values in seconds for each condition are provided next to the large scatter points. The *p*-value presented above the bracket was obtained using the log transform of average looking time in seconds between the AV asynchronous and AV synchronous conditions.

Our ANOVA also showed that there was no significant main effect of run order, *F*(1, 22) = 0.16, *p* = 0.69, nor was there a significant interaction between AV trial type and run order, *F*(1, 22) = 1.56, *p* = 0.23. We also conducted a pre-registered paired samples *t*-test to assess whether the initial appearance of the ball stimulus (i.e., top-left, top-right) across trials influenced the looking time results, similar to Smith et al. (2017). We found no significant differences in average logged looking times between left (M_{left} = 1.01, SD_{left} = 0.29) and right side onset trials (M_{right} = 1.02, SD_{right} = 0.25; t(23) = -0.17, p = 0.86).

Exploratory Linear Mixed Effects Analysis of Logged Looking Time Across Individual Trials

We also opted to conduct an exploratory linear mixed effects analysis to examine the interaction between trial type (two level factorial variable) and trial number (ordered 10 level factorial variable; 1 through 12, surprise trials 5 and 9 dropped) on logged looking time. The model included trial type and trial number as fixed effects and participant as a random effect to account for individual variability. The analysis was performed using the lmerTest package in R software (Version 4.3.2). We tested the interaction between trial type and trial number on logged looking times using the following code: lmer(Log Duration ~ Trial Number * Trial Type + (1|participant)).

The analysis of model fit found that trial number (β = -0.37, 95% CI [-0.055, -0.2], p < .001) and trial type (β = -0.09, 95% CI [-0.17, -0.01], p = .04) were significant predictors of logged looking times. As expected, logged looking times decreased as trial number increased. Logged looking times were also significantly longer for the asynchronous condition (M = 1.03, 95% CI [0.93, 1.14]) compared to synchrony (M = 0.95, 95% CI [0.85, 1.05]; t = 2.03, p = .04). We found a significant interaction between trial type and trial number (β = -0.38, 95% CI [-0.3, 0.2], p = .05). To follow up, we ran the model separately for each trial type, using ordered trial number as a fixed factor and participant as a random effect. We found that the synchronous condition had a steeper slope of decline (β = -0.44, 95% CI [-0.63, -0.25], p < .001) for logged looking times compared to asynchrony (β = -0.38, 95% CI [-0.55, -0.21], p < .001; see figure 4).

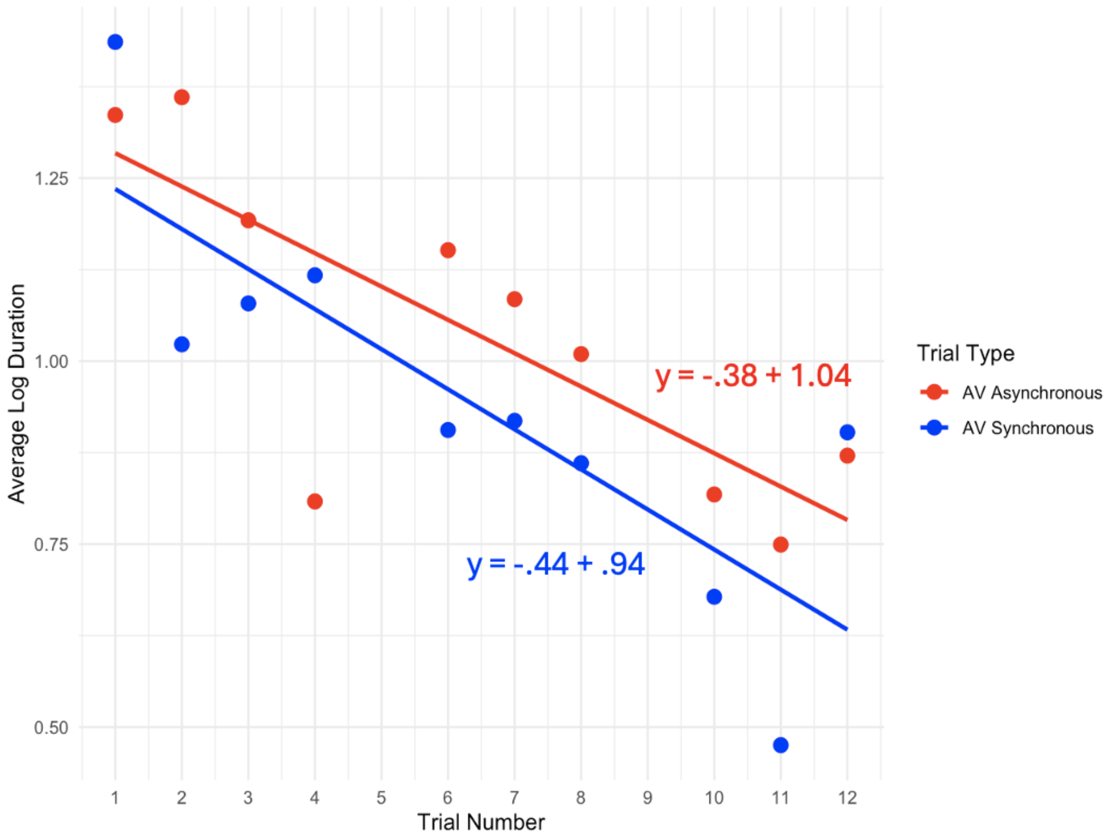

Figure 3.4. *Model estimates of logged looking times plotted across trial number, split between the AV asynchronous and AV synchronous conditions.* The above figure depicts the estimated marginal means obtained from the model for each trial number, split between trial type. The y-axis represents the logged looking time (log10) in seconds, while the x-axis represents trial number. The solid red and blue lines represent the line of best fit for each condition. Trials five and nine were the AV-surprise trials and were not included in the analysis or figure.

A likelihood ratio test shows that the interaction effect model with trial number and trial type as predictors explained significantly more variation (AIC = 138.1; BIC = 210.1) in the observed data compared to main effects model with trial number and trial type as additive predictors ($\chi2(9) = 17.9$, p = .04, AIC = 138; BIC = 180.5) and compared to a null model ($\chi2(19) = 98.4$, p < .001, AIC = 198.5; BIC = 208.3).

Discussion

Our goal for this study was to assess if infants as young as four-to-five months of age are sensitive to the spatiotemporal congruency between dynamic visual input and associated sounds. We predicted that AV asynchronous spatiotemporal information (i.e., sound preceding visual contact with a physical barrier) should result in increased looking time relative to synchronous representations due to infants' expectations about alignment between the timing of visual input and the onset of a sound. Our results supported this prediction: as a group, infants looked significantly longer toward asynchronous AV events. These findings expand our knowledge of the crossmodal perceptual abilities of infants, demonstrating that they are sensitive to AV associations over the course of a brief exposure phase. This is an important extension of previous work, such as the study by Smith et al. (2017), where infants looked longer toward events that contained a mismatch between the number of visual objects seen and sounds heard (i.e., one ball bouncing, two sounds elicited). Our data reveals that infants also display an early sensitivity toward the spatiotemporal alignment of dynamic visual input and sounds.

Past computational research has found that young children are able to integrate visual and auditory information to make inferences about past events (Outa et al., 2022). Infants’ early sensitivity to low-level temporal properties of AV associations may provide a foundation for attending to and learning about the causes underlying multimodal events. Since we did not utilize a habituation or familiarization/test procedure, the infants’ sensitivity to distinguish between AV asynchronous and synchronous inputs suggests that the infants have prior expectations about visual objects and their accompanying sounds. The lack of a familiarization/test procedure is also in contrast to the launching studies previously described (Michotte, 1946, 1963; Ball, 1973; Leslie, 1984; Kotovsky & Baillargeon, 2000; Saxe & Carey, 2006; Kominsky et al., 2017) in which infants were first habituated to a plausible launching event (i.e., a moving object collided with another static object, which propels it into motion), and were then presented with a violation of the previous event (i.e., uninitiated contact between the two objects results in launching). Although extremely useful, in habituation designs, an infant's ability to detect a violation of expectation is contingent on learning the environmental constraints of the experiment during the initial exposure phase. In contrast, our study showed that as a group, infants were able to notice a difference in AV spatiotemporal information between the two conditions based on previously held expectations of how their physical world operates.

The findings reported here suggest the presence of mechanisms responsible for associating sounds with the behavioral dynamics of a moving visual object very early in development. This raises the question: what type of perceptual experiences are useful in facilitating these skills? Dynamic AV sensory information is ubiquitously available to the infant at birth (i.e., hand clapping, objects dropping, balls bouncing, people speaking, etc.), which may be one reason why infants are able to detect changes in the temporal alignment of sounds caused by preceding visual input so early in life. Mandler (2012) argued that infant inferences about launching events also coincides with increased experiences with forces acting on the body in natural environments. For example, newborn infants often experience pressure and resistance because they are pressed against things in their immediate environment (e.g., resting in a cradle, being held, or swaddled by a blanket). As the infant develops, so do their fine and gross motor skills and their abilities to interact with objects in their immediate environment. Infants then gain more opportunities to manipulate objects, realizing that some objects can be picked up while others cannot, or that people are self-propelled agents and can send non-animate objects into motion. Mandler (2012) concluded that inferring launch causality emerges through the interaction between domain-general processes that generate concepts from attended perceptual events and the ability to manipulate a range of objects. On the other hand, the perception of crossmodal associations underlying AV input may not be dependent on the acquisition of fine and gross motor skills, but rather the maturation of visual and auditory systems, which develop sooner. Since AV information is readily available in naturalistic settings, the early ability to assess the congruence of the onset of sounds and the behavioral dynamics of a visual object may be the result of earlier capacities for the infant to extract regularity from objects, independent from motor abilities. As such, AV information may provide the infant opportunities to employ perceptual predictions in the real world earlier than launching events would.

The finding that infants have early developing abilities to notice small spatiotemporal incongruencies suggests an early sensitivity to the crossmodal structure of sound anticipated by a moving visual object. Detecting the spatiotemporal incongruence between sound onset and the movement of visual object relies on a host of different processes, including visuospatial memory (Orioli et al., 2018), crossmodal integration (Lewkowitz, 1996; Grossman et al., 2006), attention (Lewkowitz & Hansen-Tift, 2012), and the eventual inference about the timing of the sound itself, which assumes a great deal of processing power. The idea that infants integrate such complex perceptual input may suggest that these skills are critical for normative perceptual development (Meltzoff, 1990; Bahrick & Lickliter, 2000), as correctly inferring the source of expected sensory information is vital to the successful navigation of environments. We would also like to preface that the data presented here does not address *how* early these skills may develop. In fact, a recent study suggests that neonates are sensitive to low-level spatiotemporal cues that determine the perception of launching events (Mascalzoni et al., 2013). Because of this, assumptions cannot be made about *when* crossmodal anticipation of sound emerges. Yet, the fact that such inconsistencies can be noticed so early in development suggests the presence of fundamental skills used to judge the spatiotemporal properties of moving objects in natural sensory environments.

The results of this study pose an important outstanding question: is infant sensitivity toward expected sounds driven by high-level perceptual inferences of causality, or are the infants simply discriminating between conditions using low-level spatiotemporal cues? For example, an alternative explanation for differences in looking time towards synchronous versus asynchronous collision events could be that infants are attending to change in the stimulus rather than demonstrating a sensitivity toward the timing of expected sound caused by collision. On the one hand, Michotte argued that the low-level representation of perceptual causality serves as the basis of higher level causal inference (1964). However, many studies in the causal launching literature have shown that infants are able to infer causality of perceptual events such as launching, even if the launch event itself was visually occluded (for review, see Saxe and Carey, 2006). Would the infant notice temporal discrepancies of sounds without precise visual representations? Trials in which infants viewed an object moving behind an occluder prior to hearing an associated sound would require the infant to infer the expected timing of the sound in the absence of precise visual cues. Alternatively, investigating physically impossible state changes of a dynamic object undergoing collision (i.e., an object changes color upon collision) is another way to test whether infants are inferring causality of sound based on collision cues, or if they are attending to a misaligned collision because it is simply linked to a stimulus change. These studies would be critical in assessing how limited sensory information may dictate the formation of higher level inferences about anticipated sounds that are associated with the behavioral dynamics of objects.

Anticipating sounds is ubiquitous in natural sensory environments and can occur in a variety of visual contexts. One key difference between this study and previous work (Smith et. al., 2017, Lewkowitz, 1996) involves the use of an abstract dynamic visual object, rather one that moves under the constraint of gravity. Here, we show that with relatively sparse and abstract perceptual input, that infants expect AV synchrony during novel events, demonstrating that infants' early cross modal expectations are quite flexible. This may relate to how more complex visual information, originating from animate beings in the environment, may influence an infants' ability to anticipate sounds. For example, perceiving a hand clap introduces the added complexity of processing hands as animate visual objects, which have been found to influence the perception of causality underlying state changes of non-animate visual objects (Muentener & Carey, 2010). Examining further scenarios with more abstract crossmodal representations would be interesting to see if more perceptually complex AV interactions would show similar effects.

As mentioned in the introduction, familiarity biases toward AV input seem to emerge in the context of learning novel crossmodal AV associations, where infants look longer toward objects upon hearing sounds that are associated with them (Spelke et al., 1976) and use temporally synchronous, redundant visual input to help facilitate low-level auditory discriminations (Bahrick & Lickliter, 2000). Differential AV associations manifesting in the form of familiarity or novelty biases may point to the presence of different crossmodal mechanisms in infancy. For example, how do learned probabilistic AV pairings (i.e., a doorbell rings, a dog toy squeaks), where familiarity biases may emerge, differ from more deterministic physical relations, in which novelty biases emerge? Neuroimaging studies may be ideal to further unpack potential mechanistic differences between the two. For example, many EEG studies have reported that the auditory response is attenuated when hearing temporally predictable sounds in neurotypical adults (for review, see Lange, 2013), and similar findings have been reported in early infancy (Hyde et al., 2009, Emberson et al., 2015, Kopp, 2014). Within-subjects comparisons of early-life neural responses to predictable versus random AV pairings may help to distinguish mechanisms responsible for predicting sensory information in natural environments.

Lastly, future research investigating how lower level processing abilities support the acquisition of higher level perceptual skills related to prediction is also needed. For example, discoveries related to understanding how infants learn to anticipate sound from visual objects may provide foundational tools related to the acquisition of skills related to language and social cognition. Early sensitivity toward the statistical regularity between syllabus in natural speech is a critical temporal feature underlying the acquisition of language (for review, Saffran, 1996). Recent EEG evidence suggests that 3-month-old infants who display mature neural responses to the low-level discrimination of pitch were found to also exhibit neural responses related to the discrimination of novel transitional probabilities of syllables embedded in a continuous speech stream. This same pattern of pitch and syllable discrimination was reproduced in adults, which suggests that this neural language learning mechanism that persists into adulthood is present within the first few months of life (Mueller et al., 2012). Dynamic spatial information, like the movement of the mouth, also influences the perception of the resulting syllable (i.e., the McGurk effect) in neurotypical adults (MacDonald & McGurk, 1978; Munhall, 1996) and infants (Kushnerenko et al. 2008; Rosenblum et al., 1997). Clearly, the perception of low level spatiotemporal information is a fundamental feature embedded across different psychological domains, such as the perception of physical and social AV associations, or even natural language. Within-subject studies that investigate how individuals process anticipated sounds across various visual contexts can help determine the extent to which the development of basic perceptual skills contributes to the capacity to predict AV interactions across cognitive, social, and language domains.

In closing, the present results suggest that infants as young as four-to-five months of age are sensitive to temporal alignment of sounds associated with a moving visual object. Early sensitivities to dynamic crossmodal spatiotemporal input, like the perception of a ball bounce, may provide the infant foundational learning opportunities to anticipate more complex sensory events during everyday perception. Infants, as a group, displayed longer looking times toward bounce sounds that occurred before the ball made visual collision with a boundary, compared to when the same sound occurred during the exact moment of visual collision. Detecting violations of perceptual expectations toward multisensory physical events likely has developmental implications for how infants form expectations about their natural sensory environments.

References

Aronson, E., & Rosenbloom, S. (1971). Space perception in early infancy: Perception within a common auditory-visual space. *Science*, *172*(3988), 1161-1163. https://doi.org/10.1126/science.172.3988.1161

Bahrick, L. E. (1983). Infants' perception of substance and temporal synchrony in multimodal events. *Infant Behavior & Development, 6*(4), 429–451. https://doi.org/10.1016/S0163-6383(83)90241-2

Bahrick, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental Psychology, 36*(2), 190–201. https://doi.org/10.1037/0012-1649.36.2.190

Bahrick, L. E. (2001). Increasing specificity in perceptual development: Infants' detection of nested levels of multimodal stimulation. *Journal of Experimental Child Psychology, 79*(3), 253–270. https://doi.org/10.1006/jecp.2000.2588

Baillargeon, R., Spelke, E. S., & Wasserman, S. (1985). Object permanence in five-month-old infants. *Cognition*, *20*(3), 191-208. https://doi.org/10.1016/0010-0277(85)90008-3

Baillargeon, R., Stavans, M., Wu, D., Gertner, Y., Setoh, P., Kittredge, A. K., & Bernard, A. (2012). Object individuation and physical reasoning in infancy: An integrative account. *Language Learning and Development*, *8*(1), 4-46. https://doi/10.1080/15475441.2012.630610

Ball W. A. (1973, April). *The perception of causality in the infant* [Paper presentation]. Society for Research on Child Development, Philadelphia.

Battaglia, P. W., Hamrick, J. B., & Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*, *110*(45), 18327-18332. https://doi.org/10.1073/pnas.1306572110

Bremner, J. G., Slater, A. M., & Johnson, S. P. (2015). Perception of object persistence: The origins of object permanence in infancy. *Child Development Perspectives*, *9*(1), 7-13. https://doi.org/10.1111/cdep.12098

Colombo, J., & Mitchell, D. W. (2009). Infant visual habituation. *Neurobiology of Learning and Memory*, *92*(2), 225-234. https://doi.org/10.1016/j.nlm.2008.06.002

Community, B. O. (2018). Blender - a 3D modeling and rendering package. Stichting Blender Foundation, Amsterdam. Retrieved from http://www.blender.org

Csibra, G., Hernik, M., Mascaro, O., Tatone, D., & Lengyel, M. (2016). Statistical treatment of looking-time data. *Developmental Psychology*, *52*(4), 521-536. https://doi.org/10.1037/dev0000083

Datavyu Team. (2014). Datavyu: A video coding tool. Databrary Project, New York University. Retrieved from http://datavyu.org.

Dodd, B. (1979). Lip reading in infants: Attention to speech presented in- and out-of-synchrony. *Cognitive Psychology, 11*(4), 478–484. https://doi.org/10.1016/0010-0285(79)90021-5

Emberson, L. L., Richards, J. E., & Aslin, R. N. (2015). Top-down modulation in the infant brain: Learning-induced expectations rapidly affect the sensory cortex at 6 months. *Proceedings of the National Academy of Sciences*, *112*(31), 9585-9590. https://doi.org/10.1073/pnas.1510343112

Gerstenberg, T., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2012). Noisy newtons: Unifying process and dependency accounts of causal attribution. *Proceedings of the 34th*

Annual Conference of the Cognitive Science Society (pp. 378-383). Cognitive Science Society. https://escholarship.org/uc/item/1496k860

Gerstenberg, T., Siegel, M., & Tenenbaum, J. (2018). What happened? Reconstructing the past through vision and sound. Proceedings of the annual meeting of the cognitive science society (Vol. 40), Madison, Wisconsin, July 25 - 28, 2018.

Gerstenberg, T., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2021). A counterfactual simulation model of causal judgments for physical events. *Psychological Review, 128*(5), 936–975. https://doi.org/10.1037/rev0000281

Gogate, L. J., & Bahrick, L. E. (2001). Intersensory redundancy and 7-month-old infants' memory for arbitrary syllable-object relations. *Infancy*, *2*(2), 219-231. https://doi.org/10.1207/S15327078IN0202_7

Grossmann, T., Striano, T., & Friederici, A. D. (2006). Crossmodal integration of emotional information from face and voice in the infant brain. *Developmental Science*, *9*(3), 309-315. https://doi.org/10.1111/j.1467-7687.2006.00494.x

Haith, M. M., Hazan, C., & Goodman, G. S. (1988). Expectation and anticipation of dynamic visual events by 3.5-month-old babies. *Child Development*, *59*(2), 467–479. https://doi.org/10.2307/1130325

Haith, M. M., & McCarty, M. E. (1990). Stability of visual expectations at 3.0 months of age. *Developmental Psychology, 26*(1), 68–74. https://doi.org/10.1037/0012-1649.26.1.68

Hillairet de Boisferon, A., Tift, A. H. Minar, N. J., & Lewkowicz, D. J. (2017). Selective attention to a talker's mouth in infancy: Role of audiovisual temporal synchrony and linguistic experience. *Developmental Science*, *20*(3), 10.1111/desc.12381. https://doi.org/10.1111/desc.12381

Hyde, D. C., Jones, B. L., Porter, C. L., & Flom, R. (2010). Visual stimulation enhances auditory processing in 3-month-old infants and adults. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, *52*(2), 181-189. https://doi.org/10.1002/dev.20417

Johnson, S. P., Bremner, J. G., Slater, A., Mason, U., Foster, K., & Cheshire, A. (2003). Infants' perception of object trajectories. *Child Development*, *74*(1), 94-108. https://doi.org/10.1111/1467-8624.00523

Kellman, P. J., & Spelke, E. S. (1983). Perception of partly occluded objects in infancy. *Cognitive Psychology*, *15*(4), 483-524. https://doi.org/10.1016/0010-0285(83)90017-8

Kersey, A. J., & Emberson, L. L. (2017). Tracing trajectories of audio-visual learning in the infant brain. *Developmental Science*, *20*(6), e12480. https://doi.org/10.1111/desc.12480

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2014). The goldilocks effect in infant auditory attention. *Child Development*, *85*(5), 1795-1804. https://doi.org/10.1111/cdev.12263

Kim, I. K., & Spelke, E. S. (1992). Infants' sensitivity to effects of gravity on visible object motion. *Journal of Experimental Psychology: Human Perception and Performance, 18*(2), 385–393. https://doi.org/10.1037/0096-1523.18.2.385

Kim, I. K., & Spelke, E. S. (1999). Perception and understanding of effects of gravity and inertia on object motion. *Developmental Science*, *2*(3), 339-362. https://doi.org/10.1111/1467-7687.00080

Kominsky, J. F. (2019). PyHab: Open-source real time infant gaze coding and stimulus presentation software. *Infant Behavior & Development*, *54*, 114-119. https://doi.org/10.1016/j.infbeh.2018.11.006

Kominsky, J. F., Strickland, B., Wertz, A. E., Elsner, C., Wynn, K., & Keil, F. C. (2017). Categories and constraints in causal perception. *Psychological Science, 28*(11), 1649-1662. https://doi.org/10.1177/0956797617719930

Kopp, F. (2014). Audiovisual temporal fusion in 6-month-old infants. *Developmental Cognitive Neuroscience, 9,* 56–67. https://doi.org/10.1016/j.dcn.2014.01.001

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLOS ONE*, *2*(9), e943. https://doi.org/10.1371/journal.pone.0000943

Kotovsky, L., & Baillargeon, R. (2000). Reasoning about collisions involving inert objects in 7.5- month-old infants. *Developmental Science, 3*(3), 344–359. https://doi.org/10.1111/1467-7687.00129

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal development of speech in infancy. *Science*, *218*, 1138-1141. https://doi.org/0.1126/science.7146899

Kuhl, P. K., & Meltzoff, A. N. (1984). The intermodal representation of speech in infants. *Infant Behavior and Development*, *7*(3), 361-381. https://doi.org/10.1016/S0163-6383(84)80050-8

Kushnerenko, E., Teinonen, T., Volein, A., & Csibra, G. (2008). Electrophysiological evidence of illusory audiovisual speech percept in human infants. *Proceedings of the National Academy of Sciences*, *105*(32), 11442-11445. https://doi.org/10.1073/pnas.0804275105

Lange, K. (2013). The ups and downs of temporal orienting: a review of auditory temporal orienting studies and a model associating the heterogeneous findings on the auditory N1 with opposite effects of attention and prediction. *Frontiers in Human Neuroscience*, *7,* 263. https://doi.org/10.3389/fnhum.2013.00263

Leslie, A. M. (1984). Spatiotemporal continuity and the perception of causality in infants. *Perception, 13*(3), 287–305. https://doi.org/10.1068/p130287

Lewkowicz, D. J. (1996). Perception of auditory–visual temporal synchrony in human infants. *Journal of Experimental Psychology: Human Perception and Performance, 22*(5), 1094–1106. https://doi.org/10.1037/0096-1523.22.5.1094

Lewkowicz, D. J. (2003). Learning and discrimination of audiovisual events in human infants: The hierarchical relation between intersensory temporal synchrony and rhythmic pattern cues. *Developmental Psychology, 39*(5), 795–804. https://doi.org/10.1037/0012-1649.39.5.795

Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences*, *109*(5), 1431-1436. https://doi.org/10.1073/pnas.1114783109

Lewkowicz, D. J., Leo, I., & Simion, F. (2010). Intersensory perception at birth: Newborns match nonhuman primate faces and voices. *Infancy*, *15*(1), 46-60. https://doi.org/10.1111/j.1532-7078.2009.00005.x

Little, P. C., & Firestone, C. (2021). Physically implied surfaces. *Psychological Science, 32*(5), 799–808. https://doi.org/10.1177/0956797620939942

MacDonald, J., & McGurk, H. (1978). Visual influences on speech perception processes. *Perception & Psychophysics*, *24*(3), 253-257. https://doi.org/10.3758/BF03206096

Mandler, J. M. (2012). On the spatial foundations of the conceptual system and its enrichment. *Cognitive Science, 36*(3), 421–451. https://doi.org/10.1111/j.1551-6709.2012.01241.x

Margoni, F., Surian, L., & Baillargeon, R. (2024). The violation-of-expectation paradigm: A conceptual overview. *Psychological Review, 131*(3), 716–748. https://doi.org/10.1037/rev0000450

Mascalzoni, E., Regolin, L., Vallortigara, G., & Simion, F. (2013). The cradle of causal reasoning: Newborns' preference for physical causality. *Developmental Science*, *16*(3), 327-335. https://doi.org/10.1111/desc.12018

Meltzoff, A. N. (1990). Towards a developmental cognitive science: The implications of cross-modal matching and imitation for the development of representation and memory in infancy. *Annals of the New York Academy of Sciences, 608,* 1–37. https://doi.org/10.1111/j.1749-6632.1990.tb48889.x

Michotte, A. (1946/ English transl. 1963). *The perception of causality,* Basic Books.

Mueller, J. L., Friederici, A. D., & Männel, C. (2012). Auditory perception at the root of language learning. *Proceedings of the National Academy of Sciences*, *109*(39), 15953-15958. https://doi.org/10.1073/pnas.1204319109

Muentener, P., & Carey, S. (2010). Infants' causal representations of state change events. *Cognitive Psychology*, *61*(2), 63-86. https://doi.org/10.1016/j.cogpsych.2010.02.001

Munhall, K. G., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the McGurk effect. *Perception & Psychophysics*, *58*, 351-362. https://doi.org/10.3758/BF03206811

Oakes, L. M. (2010). Using habituation of looking time to assess mental processes in infancy. *Journal of Cognition and Development*, *11*(3), 255-268. https://doi.org/10.1080/15248371003699977

Orioli, G., Bremner, A. J., & Farroni, T. (2018). Multisensory perception of looming and receding objects in human newborns. *Current Biology*, *28*(22), 1283–1295, https://doi.org/10.1016/j.cub.2018.10.004

Outa, J., Zhou, X. J., Gweon, H., & Gerstenberg, T. (2022). Stop, children what's that sound? Multi-modal inference through mental simulation. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 44(44), 1359-1366. https://escholarship.org/uc/item/0jb2c02j

Parise, C. V., Spence, C., & Ernst, M. O. (2012). When correlation implies causation in multisensory integration. *Current Biology*, *22*(1), 46-49. https://doi.org/10.1016/j.cub.2011.11.039

Peirce, J. W. (2007). PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1-2), 8-13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Poli, F., Serino, G., Mars, R. B., & Hunnius, S. (2020). Infants tailor their attention to maximize learning. *Science Advances*, *6*(39), eabb5053. https://doi.org/10.1126/sciadv.abb5053

Roder, B. J., Bushnell, E. W., & Sasseville, A. M. (2000). Infants' preferences for familiarity and novelty during the course of visual processing. *Infancy*, *1*(4), 491-507. https://doi.org/10.1207/S15327078IN0104_9

Rosenblum, L. D., Schmuckler, M. A., & Johnson, J. A. (1997). The McGurk effect in infants. *Perception & Psychophysics*, *59*(3), 347-357. https://doi.org/10.3758/BF03211902

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926-1928. https://doi.org/10.1126/science.274.5294.1926

Saxe, R., & Carey, S. (2006). The perception of causality in infancy. *Acta Psychologica, 123*(1-2), 144–165. https://doi.org/10.1016/j.actpsy.2006.05.005

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, *4*(8), 299-309. https://doi.org/10.1016/S1364-6613(00)01506-0

Smith, K. A., Battaglia, P., & Vul, E. (2013). Consistent physics underlying ballistic motion prediction. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 3426–3431). Cognitive Science Society.

Smith, K. A., & Vul, E. (2014). Looking forwards and backwards: Similarities and differences in prediction and retrodiction. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 1467–1472). Cognitive Science Society.

Smith, N. A., Folland, N. A., Martinez, D. M., & Trainor, L. J. (2017). Multisensory object perception in infancy: 4-month-olds perceive a mistuned harmonic as a separate auditory and visual object. *Cognition*, *164*, 1-7. https://doi.org/10.1016/j.cognition.2017.01.016

Smith-Flores, A. S., Perez, J., Zhang, M. H., & Feigenson, L. (2021). Online measures of looking and learning in infancy. *Infancy*, *27*(1), 4-24. https://doi.org/10.1111/infa.12435

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review, 99*(4), 605–632. https://doi.org/10.1037/0033-295X.99.4.605

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*(1), 89-96. https://doi.org/10.1111/j.1467-7687.2007.00569.x

Spelke, E. S. (1976). Infants' intermodal perception of events. *Cognitive Psychology*, *8*(4), 553-560. https://doi.org/10.1016/0010-0285(76)90018-9

Spelke, E. S. (1979). Perceiving bimodally specified events in infancy. *Developmental Psychology, 15*(6), 626–636. https://doi.org/10.1037/0012-1649.15.6.626

Spelke, E. S. (1985). Preferential-looking methods as tools for the study of cognition in infancy. In G. Gottlieb & N. A. Krasnegor (Eds.), *Measurement of audition and vision in the first year of postnatal life: A methodological overview* (pp. 323–363). Ablex Publishing.

Spelke, E. S. (1990). Principles of object perception. *Cognitive Science, 14*(1), 29–56. https://doi.org/10.1207/s15516709cog1401_3

Ullman, T. D., & Tenenbaum, J. B. (2020). Bayesian models of conceptual development: Learning as building models of the world. *Annual Review of Developmental Psychology*, *2*, 533-558. https://doi.org/10.1146/annurev-devpsych-121318-084833

Wallace, M. T., Roberson, G. E., Hairston, W. D., Stein, B. E., Vaughan, J. W., & Schirillo, J. A. (2004). Unifying multisensory signals across time and space. *Experimental Brain Research*, *158*, 252-258. https://doi.org/10.1007/s00221-004-1899-9

Wellman, H. M., & Gelman, S. A. (1992). Cognitive development: Foundational theories of core domains. *Annual Review of Psychology, 43,* 337–375. https://doi.org/10.1146/annurev.ps.43.020192.002005

White, P. A. (2014). Singular clues to causality and their use in human causal judgment. *Cognitive Science, 38*(1), 38–75. https://doi.org/10.1111/cogs.12075

Chapter 3, in part is currently being prepared for submission for publication of the material. Marin, Andrew; Fearns, Zara; Dratva, Melanie; Powell, Lindsey J.; Störmer, Viola S.; Carver, Leslie J. The dissertation author was the primary investigator and author of this material.

Conclusion

With this dissertation, I examined how the brain anticipates sound via collision events, such as a bouncing ball. Grounded in predictive coding theory, I showed that the dynamic movement of objects can modulate auditory responses based on whether the object’s motion was congruent or incongruent with timing of a collision sound (chapter one). I then tested these same methods, but in a sample of autistic adults, finding that early responses towards incongruent collision sounds were larger in autism relative to neurotypical (NT) controls (chapter two). To examine the developmental origins of these skills, I demonstrated that infants as young as four months of age are able to detect small incongruencies about the timing of collision sounds (chapter three). In predictive coding, predictions are actively compared with incoming sensory inputs, and any discrepancies or "prediction errors" between the predicted and actual inputs are used to update and refine the brain's internal models. A key aspect of predictive coding theory involves two distinct processes: integrating predictions and detecting errors. Yet actual mechanistic explanations of these processes remain largely unexplored.

In chapter one, I implemented novel methods to test how visual expectations can influence early sensory processing of congruent or incongruent collision sounds. In NT adults, sounds that were incongruent with visual expectations resulted in enlarged neural responses, which reflect the integration of prediction error. Conversely, sounds that were in line with visual expectations resulted in a reduction of neural responses relative to processing sound in isolation, which reflect the integration of top-down predictions with congruent sensory information. Here, I demonstrated that the processes related to representing error and predictions are mechanistically distinct. Although these processes are different, disruptions to either could equally impact predictive neural systems. Difficulties in appropriately contextualizing previous expectations in otherwise unpredictable sensory environments could serve as a promising target to test in clinical populations who report difficulties in implementing predictions.

Predictive coding theory has also been proposed as a framework used to explain the diverse symptom profile seen in autism. It has been theorized that autistic people may exhibit atypical responses to sensory stimuli because predictive mechanisms might be less accurate or differently tuned compared to NT. In chapter two, I sought out to utilize the methods outlined in chapter 1, but in a sample of autistic adults. Although predictive coding lends a parsimonious explanation of autistic symptoms, more research is needed to fully elucidate the mechanisms that may be impacted. To date there have been no attempts to directly compare the mechanisms that reflect the integration of predictions versus those that signal error in autism. Using the same methods outlined in chapter one, I found that the auditory response elicited when hearing incongruent sounds was enlarged in autism relative to NT. Importantly, auditory responses to fully visible and occluded synchrony were not different between the groups. These findings suggest that autistic predictive differences in response to simple auditory expectations are driven by the perception of error, and not the integration of predictions.

Predictions provide us with invaluable tools to construct mental models about the world. Not only do our expectations inform our understanding of the physical properties underlying natural scenes, but they are also sensitive to the detection of faces that signal expressive emotions and social cues. The anticipation of a person's movements in relation to objects in naturalistic environments communicates intention and requests. Social beings are dynamic but far less predictable than non-social objects. Larger brain responses to error may cause the autistic individual to pay attention to uninformative contingencies in the environment at the expense of social information. That same person might experience significant activation in error networks when engaging in social situations because social beings are less predictable. This may lead that individual to self-select away from social environments, or even counteract confusing information by engaging in restricted and repetitive behaviors. By self-selecting away from unpredictable sensory environments, the individual limits their opportunities to engage in dynamic interactions, social or non-social, which have downstream cognitive and social consequences. Thus, isolating the processes involved in representing predictions and error may provide clues to how they govern perceptual and social experiences in autistic individuals.

Because autism is a neurodevelopmental condition, any explanation of autism using predictive coding as its basis must explain how disruptions in predictive systems manifest into the fundamental symptoms that characterize autism. An initial step in evaluating these developmental questions concerning autism involves investigating whether NT infants can perceive the timing of anticipated sounds. In chapter three, I explored the developmental origins of these skills by examining infant looking time behavior when viewing temporally predictable sound versus sound that violated expectations (chapter three). Here, we modified the methods used in chapters one and two for use in an infant looking time study. I found that 4- to 5-month old NT infants look longer toward collision events that violate expectations of when sound should occur, suggesting that infants are sensitive to the low-level properties that predict anticipatory sounds early in life. These processes are used to make powerful inferences about environments, and could play a large role in implementing more general predictions.

Our sensory environment contains dynamic objects that behave in accordance with their physical properties (Torralba & Oliva, 2003), constraints (Oliva & Torralba, 2007; Torralba et al., 2006), and statistical dependencies (Kersten & Yuille, 2003; Penev & Atick, 1996). Infants too represent objects and their properties in natural environments (Bremner et al., 2015; Spelke

& Kinzler, 2007), suggesting that these skills are foundational for normative perception, affording the infant opportunities to reason about the physical properties of objects and how they interact in natural environments (Baillargeon, 2008; Feigenson & Carey, 2003; Rosenberg & Carey, 2006; Spelke & Kinzler, 2007). Early perceptual sensitivities toward the predictable properties of objects affords the infant opportunities to learn statistical regularities embedded in environments (Bulf et al., 2011; Kirkham et al., 2002), which can build into increasingly more complex physical abstractions such as perceptual causality (Kominsky et al., 2017; Leslie & Keeble, 1987; Newman et al., 2008; Oakes & Cohen, 1990; Saxe & Carey, 2006; Scholl & Tremoulet, 2000; Shultz, 1982). Such reasoning abilities inform the infant's expectations about future sensory events, which helps to resolve perceptual ambiguities as they arise (Kouider et al., 2015; Schlottmann, 2001; Trainor, 2012). In the first year of life, early perceptual sensitivities toward dynamic objects play a vital role in explaining how infants make sense of increasingly complex sensory environments. Because infants are sensitive to auditory predictions, and the processes used to represent predictions appear to be different in autism, further examining the development of these mechanisms may offer an explanatory framework for understanding the emergence of autism prior to reliable diagnoses. As such, early disruptions underlying the processes related to perceiving and interpreting basic sensory predictions about objects are likely to have a cascading influence on the development of higher-order predictive models.

Considering that sensory systems employ mechanisms to predict sound (chapter one) and are impacted in autism (chapter two), it is logical to investigate how neural variations in infancy influence the capacity to predict sound. This, in turn, could contribute to the early emergence of autistic symptoms seen later in childhood. Multiple lines of evidence suggest that alterations in neural connectivity underlie the core ASD phenotype (Belmonte, 2004; Bourgeron, 2015;

Geschwind & Levitt, 2007; Parikshak et al., 2015; Port et al., 2014). Infants later diagnosed with autism also experience accelerated rates of cortical surface area hyper-expansion between 6 to 12 months, which give rise to altered sensorimotor/attentional experiences (for review, Piven et al., 2017). Aberrant perceptual experiences in infants later diagnosed with ASD leads to altered experience-dependent neural development, resulting in a decreased elimination of neural processes (for review see Akshoomoff et al., 2002) and brain volume overgrowth during the second and third years of life (Hazlett et al., 2011). Thus, disrupted neural connectivity in the first year of life can be detected in sensory systems, which could affect perceptual experiences well before autistic symptoms arise. Interestingly, wide-spread neural alterations also co-occur within similar developmental time frames in which infants are beginning to learn how objects interact in natural environments. Such developmental overlap suggests that processes involved in simple audio-visual object relations may be ideal perceptual readouts of disrupted neural circuitry designed to process expected sounds early in life.

In sum, quantifying the biological and developmental mechanisms that govern the processing of expected sensory information in infants at risk for autism is a promising target to understand the perceptual consequences of diffuse structural alterations underlying integrative sensory systems in the brain. Altered dynamic sensory representations may impede the infant's ability to learn from their sensory world, consequently influencing how the infant reasons about the nature of objects. Disruptions to these early formative perceptual mechanisms may, in turn, have downstream consequences for later development, giving rise to cognitive and social communication differences that characterize neurodevelopmental disorders like autism. Thus, the auditory signals measured in this dissertation may offer tractable targets of predictive differences prior to the emergence of autistic symptoms.

References

Akshoomoff, N., Pierce, K., & Courchesne, E. (2002). The neurobiological basis of autism from a developmental perspective. *Development and Psychopathology*, *14*(3), 613–634. https://doi.org/10.1017/S0954579402003115

Baillargeon, R. (2008). Innate ideas revisited: For a principle of persistence in infants' physical reasoning. *Perspectives on Psychological Science*, *3*(1), 2–13. https://doi.org/10.1111/j.1745-6916.2008.00056.x

Belmonte, M. K. (2004). Autism and abnormal development of brain connectivity. *Journal of Neuroscience*, *24*(42), 9228–9231. https://doi.org/10.1523/JNEUROSCI.3340-04.2004

Bourgeron, T. (2015). From the genetic architecture to synaptic plasticity in autism spectrum disorder. *Nature Reviews Neuroscience, 16*, 551-563. https://doi.org/10.1038/nrn3992

Bremner, J. G., Slater, A. M., & Johnson, S. P. (2015). Perception of object persistence: The origins of object permanence in infancy. *Child Development Perspectives*, *9*(1), 7–13. https://doi.org/10.1111/cdep.12098

Bulf, H., Johnson, S. P., & Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition*, *121*(1), 127–132. https://doi.org/10.1016/j.cognition.2011.06.010

Feigenson, L., & Carey, S. (2003). Tracking individuals via object-files: Evidence from infants' manual search. *Developmental Science*, *6*(5), 568–584. https://doi.org/10.1111/1467-7687.00313

Geschwind, D. H., & Levitt, P. (2007). Autism spectrum disorders: Developmental disconnection syndromes. *Development*, *17*(1), 103–111. https://doi.org/10.1016/j.conb.2007.01.009

Hazlett, H. C., Poe, M. D., Gerig, G., Styner, M., Chappell, C., Smith, R. G., Vachet, C., & Piven, J. (2011). Early brain overgrowth in autism associated with an increase in cortical surface area before age 2 years. *Arch Gen Psychiatry, 68*(5), 467-476. https://doi.org/10.1001/archgenpsychiatry.2011.39

Kersten, D., & Yuille, A. (2003). Bayesian models of object perception. *Current Opinion in Neurobiology*, *13*(2), 150–158. https://doi.org/10.1016/S0959-4388(03)00042-4

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35–B42. https://doi.org/10.1016/S0010-0277(02)00004-5

Kominsky, J. F., Strickland, B., Wertz, A. E., Elsner, C., Wynn, K., & Keil, F. C. (2017). Categories and constraints in causal perception. *Psychological Science*, *28*(11), 1649–1662. https://doi.org/10.1177/0956797617719930

Kouider, S., Long, B., Le Stanc, L., Charron, S., Fievet, A.-C., Barbosa, L. S., & Gelskov, S. V. (2015). Neural dynamics of prediction and surprise in infants. *Nature Communications*, *6*(1). https://doi.org/10.1038/ncomms9537

Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition*, *25*(3), 265–288. https://doi.org/10.1016/S0010-0277(87)80006-9

Newman, G. E., Choi, H., Wynn, K., & Scholl, B. J. (2008). The origins of causal perception: Evidence from postdictive processing in infancy. *Cognitive Psychology*, *57*(3), 262–291. https://doi.org/10.1016/j.cogpsych.2008.02.003

Oakes, L. M., & Cohen, L. B. (1990). Infant perception of a causal event. *Cognitive Development*, *5*(2), 193–207. https://doi.org/10.1016/0885-2014(90)90026-P

Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences*, *11*(12), 520–527. https://doi.org/10.1016/j.tics.2007.09.009

Parikshak, N. N., Gandal, M. J., & Geschwind, D. H. (2015). Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. *Nature Reviews Genetics*, *16*(8), 441–458. https://doi.org/10.1038/nrg3934

Penev, P. S., & Atick, J. J. (1996). Local feature analysis: A general statistical theory for object representation. *Network: Computation in Neural Systems*, *7*(3), 477–500. https://doi.org/10.1088/0954-898X_7_3_002

Piven, J., Elison, J. T., & Zylka, M. J. (2017). Toward a conceptual framework for early brain and behavior development in autism. *Molecular Psychiatry*, *22*(10), 1385–1394. https://doi.org/10.1038/mp.2017.131

Port, R. G., Gandal, M. J., Roberts, T. P. L., Siegel, S. J., & Carlson, G. C. (2014). Convergence of circuit dysfunction in ASD: A common bridge between diverse genetic and environmental risk factors and common clinical electrophysiology. *Frontiers in Cellular Neuroscience*, *8*. https://doi.org/10.3389/fncel.2014.00414

Saxe, R., & Carey, S. (2006). The perception of causality in infancy. *Acta Psychologica*, *123*(1–2), 144–165. https://doi.org/10.1016/j.actpsy.2006.05.005

Schlottmann, A. (2001). Perception versus knowledge of cause and effect in children: When seeing is believing. *Current Directions in Psychological Science*, *10*(4), 111–115. https://doi.org/10.1111/1467-8721.00128

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, *4*(8), 299–309. https://doi.org/10.1016/S1364-6613(00)01506-0

Shultz, T. R. (1982). Rules of causal attribution. *Monographs of the Society for Research in Child Development*, *47*(1), 1. https://doi.org/10.2307/1165893

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*(1), 89–96. https://doi.org/10.1111/j.1467-7687.2007.00569.x

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network: Computation in Neural Systems*, *14*(3), 391–412. https://doi.org/10.1088/0954-898X_14_3_302

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, *113*(4), 766–786. https://doi.org/10.1037/0033-295X.113.4.766

Trainor, L. J. (2012). Predictive information processing is a fundamental learning mechanism present in early development: Evidence from infants. *International Journal of Psychophysiology*, *83*(2), 256–258. https://doi.org/10.1016/j.ijpsycho.2011.12.008

Rosenberg, R. D., & Carey, S. (2006). Infants' indexing of objects vs. non-cohesive substances. *Journal of Vision*, *6*(6), 611–611. https://doi.org/10.1167/6.6.611