

Jason S. Chan* & Marcus J. Naumer
Institute of Medical Psychology
Goethe-University, Frankfurt

KEYWORDS: Autism Spectrum Disorder, Multisensory Integration, Temporal Binding Window

Acknowledgements: This was funded by the Hessian initiative for the development of scientific and economic excellence (LOEWE) Neuronal Coordination Research Focus Frankfurt (NeFF).

*Corresponding author:
Jason Chan
chan@med.uni-frankfurt.de
Abstract

In this article, we review the paper *Multisensory Temporal Integration in Autism Spectrum Disorders*, by Stevenson et al. (2014). This paper illustrates the need to present different forms of stimuli in order to characterize the perceptual abilities of people with autism spectrum disorder (ASD). Furthermore, we will discuss their behavioral results and offer an opposing viewpoint to the suggested neuronal drivers of ASD.
The difference in propagation time between an auditory and a visual stimulus can be substantial, depending on the distance between the perceiver and stimulus location. For this reason there is a ‘window’ of time within which stimulus information received from the different sensory systems can be integrated to form a single percept. The type of stimuli presented can affect these temporal binding windows (TBW). Additionally, studies have demonstrated that the TBW is also affected by neurodevelopmental disorders (e.g., autism; Foss-Feig et al., 2010). Previous studies investigating the TBW have used one type of stimuli only. However, a recent paper by Stevenson et al. (2014) illustrates the need to present different forms of stimuli in order to characterize the perceptual abilities of people with ASD.

In three experiments, Stevenson et al. (2014) explored the changing TBW in children with autism spectrum disorder (ASD) and typically developing (TD) children (6-18 years old). Each experiment possessed different levels of stimulus complexity and social context. In the flash-beep task, a simple visual annulus was flashed along with an auditory beep at different stimulus-onset asynchronies (SOAs) between them. In the tool-task, participants were presented with a video of a hammer striking a table. The sound of the strike was either presented simultaneously to the hammer hitting the table or offset using a variable SOA. In the third temporal task, the visual stimulus was a video of a face verbalizing a syllable (e.g., “ba”) and the auditory stimulus was the vocalization of the same syllable. Once again, the auditory stimulus was either presented simultaneously to the vocalization or offset using a variable SOA. In all three tasks, participants determined whether the auditory and visual stimuli were presented simultaneously or temporally misaligned. The authors found no significant difference between the TD and ASD cohorts in the flash-beep and tool tasks. Interestingly, in the face-voice experiment, ASD participants compared to the TD participants required longer SOAs between the visual mouth movement and the auditory vocalization to detect the temporal asynchrony between the two modalities. These results suggest that for the ASD participants, increasing the stimulus complexity and social context widened their TBW.

In a separate task, all participants were presented with the McGurk illusion. The classic McGurk illusion is a powerful demonstration of audio-visual integration whereby an auditory “ba” is accompanied by the visual mouth movements of “ga”. This combination is perceived by the listener as a completely different syllable: “da” (McGurk and MacDonald 1976). In this task, participants indicated the phoneme they ‘heard’ while maintaining fixation on the mouth of a person vocalizing another syllable. Stevenson et al. found the participants with ASD perceived fewer McGurk illusions compared to TD. ASD participants were also more likely to indicate the auditory phoneme, suggesting an attentional or response bias to the auditory stimulus. The number of perceived McGurk illusions was negatively correlated with the size with each of the TBWs in ASD participants. The strongest correlation was between the McGurk illusion and the face-voice TBW. In other words, the larger the face-voice TBW the less likely they were to perceive the McGurk illusion. They did not
find any correlation between any of the TBWs and the amount of perceived McGurk illusions in the TD participants.

One puzzling issue is that the authors did not find significant TBW differences between the ASD and TD participants in the flash-beep task (although, there is a trend). However, previous data from the same laboratory have illustrated that people with ASD perceive the sound-induced flash illusion (an illusion that requires the temporal integration of auditory and visual stimuli) at larger SOAs compared to TD (Foss-Feig et al. 2010). It is possible that the conflicting results are due to stimulus- and task-related differences between the two experiments. However, Donohue, Darling, and Mitroff (2012) also found a significant correlation between the audio-visual point-of-subjective-simultaneity (PSS; another task related to the TBW) and the ‘level’ of autism, as measured by the Autism Self-Questionnaire (ASQ). The visual stimulus was a simple checkerboard pattern and the auditory stimulus was a 1,200 Hz pure tone. It is important to point out that the participants in Donohue et al.’s study did not receive a formal clinical diagnosis of autism.

The differences between conditions could also be due to the different ocular scanning patterns of the face. Dalton et al. (2005) demonstrated using eye-tracking that people with ASD tend to avoid fixating on the eyes, compared to healthy controls. However, they found no significant difference in the number and duration of fixations to the mouth and face in general. In other words, the ratio of fixations between the mouth and eyes was greater in the people with ASD. Stevenson and colleagues did not track the eye movements of their participants in each task. If the ASD participants made fewer fixations on the mouth, this would explain why they perceived fewer McGurk illusions. They simply did not integrate the visual and auditory stimuli. In the TBW task, ASD participants integrated more audio-visual stimuli but integrated less in the McGurk illusion task. These results could be due to task or instructional differences. In the TBW task, participants were required to attend to both modalities to complete the task; while in the McGurk illusion, the task could be completed by attending to the auditory modality, alone. They attempted to control for these biases by incorporating unimodal-vision and auditory trials.

Stevenson et al. suggest that their behavioral results can be explained by the *weak-coherence theory* (Brock et al. 2002; Happé and Frith 2006). This theory suggests that people with ASD have decreased coherence between brain areas. It attempts to explain why people with autism do not integrate multiple features into a single coherent percept (e.g., when presented with the Kanizsa triangle illusion). However, the fact that Stevenson and colleagues did not find significant behavioral group differences in their ‘flash-beep’ and ‘tools’ experiments suggests that, at least, coherence between areas processing less complex audio-visual stimuli is not ‘weak’. There is plenty of research suggesting that there is something special about face-speech stimuli (see Vatakis and Spence 2007). It is possible that the coherence between object recognition regions, such as the lateral occipital complex
and auditory cortex is stronger than the coherence between face processing regions (e.g., fusiform face area) and the auditory cortex.

Recently, weak-coherence theory has come under scrutiny (Pellicano and Burr 2012a). Originally, the weak-coherence theory was described as simply a ‘cognitive style’ (Happé and Frith 2006), offering an explanation as to why people with ASD are less likely to combine the local and global features of objects (e.g., Navon stimuli). Navon stimuli are objects where the global shape of an object is comprised of smaller features. However, recent evidence has demonstrated that people with ASD are able to perform at similar levels compared to typically developed adults when explicitly asked to integrate the local-global features of Navon stimuli as well as perform match-to-sample tasks where the local and global features are manipulated (Bernardino et al. 2012; Plaisted et al. 1999). Differences between the ASD and TD groups emerged when comparing TD to ASD with intellectual disabilities (Bernardino et al. 2012). Bernardino et al. suggest that the ASD participants may not have problems with coherence between cortical regions, but that there are differences in their dorsal attention network.

An alternative to the weak-coherence theory is the theory of predictive coding. Predictive coding has been discussed as another possible model to explain the differences in sensory integration in people with ASD (Pellicano and Burr 2012b). In general, predictive coding suggests that higher-order brain areas are constantly sending predictions to early-sensory areas to explain perception. Then, lower-brain areas feed-forward the prediction errors back to the higher-level areas. This cycle continues until the prediction error is minimized. Stevenson et al., suggest that the TBW can be seen as an indication of sensory integration reliability. Thus, predictive coding would expect those ASD participants with a small TBW to have ‘stronger’ Bayesian priors to integrate audio-visual information more reliably. The ‘wider’ temporal binding window can be seen as a poor reliability (i.e., weaker priors) between the senses, and thus these ASD participants perceived fewer McGurk illusions.

Neuroimaging results based on Stevenson et al.’s. behavioral data will help to elucidate which theory is correct and further advance the important field of ASD research. According to sensorimotor theory, there should be an overall increase in the structural connectivity between the brain areas responsible for the processing of the flash-beep and tool-use areas (e.g., primary sensory areas, superior temporal gyrus/sulcus, and the lateral occipital complex); compared to areas responsible for face-voice processing (e.g., primary sensory areas, fusiform face area, superior temporal gyrus/sulcus, planum temporale, and middle temporal gyrus). Predictive coding would rely on the information transfer between these brain areas and thus predict reduced beta-band activity for the ASD participants when they are presented with the face-voice stimuli. This is because top-down activity is related to beta-band activity (see Bastos et al. 2012 for a review). There should be no significant difference in beta-band activity between the two groups when they view the flash-beep and the audiovisual tool stimuli. Furthermore, more recent techniques such as transfer entropy could elucidate the direction of
EXPLAINING AUTISM SPECTRUM DISORDERS

information transfer between higher- and lower-brain areas by investigating the frequencies associated
with top-down and bottom-up processing.
References