Atypical audiovisual temporal function in autism and schizophrenia: similar phenotype, different cause

Jean-Paul Noel,1,2 Ryan A. Stevenson3,4,5,6 and Mark T. Wallace2,7,8,9,10
1Neuroscience Graduate Program, Vanderbilt University, 7110 MRB III BioSci Bldg, 465, 21st Ave South, Nashville, TN 3721, USA
2Vanderbilt Brain Institute, Vanderbilt University, Nashville, TN, USA
3Department of Psychology, University of Western Ontario, London, ON, Canada
4Brain and Mind Institute, University of Western Ontario, London, ON, Canada
5Department of Psychiatry, Schulich School of Medicine and Dentistry, University of Western Ontario, London, ON, Canada
6Program in Neuroscience, Schulich School of Medicine and Dentistry, University of Western Ontario, London, ON, Canada
7Department of Psychology, Vanderbilt University, Nashville, TN, USA
8Department of Hearing and Speech, Vanderbilt University Medical Center, Nashville, TN, USA
9Department of Psychiatry and Behavioral Sciences, Vanderbilt University Medical Center, Nashville, TN, USA
10Department of Pharmacology, Vanderbilt University, Nashville, TN, USA

Keywords: autism, causal inference, multisensory integration, schizophrenia, speech

Abstract

Binding across sensory modalities yields substantial perceptual benefits, including enhanced speech intelligibility. The coincidence of sensory inputs across time is a fundamental cue for this integration process. Recent work has suggested that individuals with diagnoses of schizophrenia (SZ) and autism spectrum disorder (ASD) will characterize auditory and visual events as synchronous over larger temporal disparities than their neurotypical counterparts. Namely, these clinical populations possess an enlarged temporal binding window (TBW). Although patients with SZ and ASD share aspects of their symptomatology, phenotypic similarities may result from distinct etiologies. To examine similarities and variances in audiovisual temporal function in these two populations, individuals diagnosed with ASD (n = 46; controls n = 40) and SZ (n = 16, controls = 16) completed an audiovisual simultaneity judgment task. In addition to standard psychometric analyses, synchrony judgments were assessed using Bayesian causal inference modeling. This approach permits distinguishing between distinct causes of an enlarged TBW: an a priori bias to bind sensory information and poor fidelity in the sensory representation. Findings indicate that both ASD and SZ populations show deficits in multisensory temporal acuity. Importantly, results suggest that while the wider TBWs in ASD most prominently results from atypical priors, the wider TBWs in SZ results from a trend toward changes in prior and weaknesses in the sensory representations. Results are discussed in light of current ASD and SZ theories and highlight that different perceptual training paradigms focused on improving multisensory integration may be most effective in these two clinical populations and emphasize that similar phenotypes may emanate from distinct mechanistic causes.

Introduction

The integration of information across distinct sensory modalities yields a host of perceptual and behavioral benefits (Calvert et al., 2004; Murray & Wallace, 2012; Stevenson et al., 2014a; Noel & Wallace, 2016), such as enhanced speech intelligibility (Sumby & Pollack, 1954; McGrath & Summerfield, 1985; Rosenblum et al., 1996; Schwartz et al., 2004; Ma et al., 2009). The temporal relationship between auditory and visual events/objects is an important determinant of whether the sensory information from these events/objects should be bound into a unitary, multisensory percept (Meredith et al., 1987; Stein & Meredith, 1993). However, different forms of sensory energy (i.e., light vs. sound) travel at distinct speeds and possess different neural transmission times. Hence, stimuli generated simultaneously from an environmental event may not be synchronous when they arrive at the peripheral organs (for example, due to distance; Van der Stoep et al., 2016; Noel et al., 2016a, in press), and these differences are likely to be propagated and amplified as the sensory information is processed in the brain. To account for these physical and neural differences, as well as the inherent noisiness involved in multisensory temporal coincidence...
detection, sensory information from the different modalities is integrated across a substantial temporal range—the temporal binding window (TBW; Dixon & Spitz, 1980; Lewkowicz, 1996; Munhall et al., 1996; Diederich & Colonius, 2015; van Wassenhove et al., 2005; Stevenson et al., 2012; Wallace & Stevenson, 2014; Noel et al., 2015, 2016b). Reinforcing these temporally based differences, information across the different senses is not necessarily maximally bound when sensory signals arrive at the peripheral nervous system synchronously, but instead reflect the statistical regularities of the environment (Vroomen & Keetels, 2010; the point of subjective simultaneity; PSS). Speech is a classic example of these natural delays, where mouthing naturally precedes audible speech and, consequently, stimuli with visual-leading asynchronies of approximately 50–100 ms are most often reported as synchronous (Dixon & Spitz, 1980; Munhall et al., 1996; van Wassenhove et al., 2005).

In concert with the adaptive value ascribed to possessing extended windows of time over which sensory information is bound, findings suggest that the TBW is highly plastic (Powers et al., 2009; Stevenson et al., 2013, 2017a; Schlesinger et al., 2014; De Nier et al., 2017), dependent upon stimulus structure and complexity (Stevenson et al., 2013), and perhaps most importantly, TBWs are anomalous in psychopathology. Indeed, while the general characterization of multisensory processes—and their temporal profile—in psychopathological conditions such as ASD and SZ populations has yield conflicting results, in the case of SZ for instance, stronger (Stone et al., 2011), similar (Wynn et al., 2014; Zyvagintsev et al., 2017), and weaker (Williams et al., 2010) multisensory facilitation vs. controls has been reported, the reports regarding multisensory TBWs in psychopathology are largely congruent—inclusively across the distinct pathologies (see Zhou et al., 2018; for a recent review and meta-analysis of multisensory temporal function in ASD and SZ). More precisely, recent work has suggested that individuals with diagnoses of autism spectrum disorder (ASD; Foss-Feig et al., 2010; Kwakye et al., 2011; Foxe et al., 2013; Stevenson et al., 2014b; Noel et al., 2016c, 2017a,b) and schizophrenia (SZ; Foucher et al., 2007; Martin et al., 2013; Su et al., 2015; Balz et al., 2016; Stevenson et al., 2017b) possess atypically large TBWs, particularly for speech stimuli. Given that (multi)sensory integration is a fundamental building block in the construction of perceptual and cognitive representations, it has been postulated that alterations in multisensory temporal function may partially scaffold the higher-order deficits present in the conditions (Russo et al., 2010; Brandwein et al., 2013; Woynaroski et al., 2013; Baun et al., 2015a; Stevenson et al., 2017c; Noel et al., 2017b). While ASD and SZ share clinical features (Volkmar & Cohen, 1991), as conceptualized in the Research Domain Criteria framework (RDOC; Insel et al., 2010), they are considered diagnostically distinct conditions. Thus, it should not be assumed that these two groups share similar etiologies of enlarged TBWs. However, to our knowledge, no previous study of temporal perception has examined both ASD and SZ individuals with an emphasis on better elucidating the mechanistic bases of their altered multisensory temporal function (see Zhou et al., 2018, for a recent review and meta-analysis). Furthermore, most reports of temporal perception in ASD and SZ rely on descriptive as opposed to principled psychophysical analyses. Here, we aim to address these two gaps in the literature.

In this study, individuals with diagnoses of ASD and SZ performed an audiovisual speech simultaneity judgment task (Baum et al., 2015b). In addition to the standard analysis entailing fitting these data with a Gaussian function and deriving the TBW (Van der Burg et al., 2013; Noel et al., 2017b; Simon et al., 2017), we applied a Bayesian causal inference model to the data (Kording et al., 2007; Shams & Beierholm, 2010). More specifically, we adopt a simplified version of the Magnotti et al., 2013 model, through which we can derive ‘Bayes-optimal synchrony windows.’ This model fitting further allows us to estimate each participant’s (i) prior probability of ascribing common cause (i.e., $P(C = 1)$), (ii) a measure of inherent sensory noisy (i.e., $\sigma_{\text{sensory noise}}^2$), and (iii) expectation of the asynchrony between audio and visual speech-related events when these do not belong to a single cause (i.e., $\mu$ ($C = 2$)), mean of the likelihood when $C = 2$; Magnotti et al., 2013). A heightened tendency to bind sensory information across time within these populations should be reflected in an increased prior probability of common cause $P(C = 1)$, while larger TBW values may also be a result of an increase in sensory noise (i.e., wider distributions of the likelihood of the source of sensory information in the world). In the latter case, larger TBWs would result from anomalous sensory processing, while in the former case, larger TBWs are more a result of a more generalized (and perhaps mal-adaptive) tendency to bind sensory information together.

Methods

Participants

This study included 118 participants: 86 in Experiment 1 and 32 in Experiment 2. In Experiment 1, 46 subjects were autistic (35 males, age = 14.1 ± 4.12 years old), and 40 were an age-matched and sex-matched typically developing (TD) cohort (29 males, age = 13.9 ± 3.99 years old, independent samples t-test for age, $t_{84} = 2.22, P = 0.082$; Chi-square for sex; $\chi^2 (2, N = 86) = 0.14, P = 0.70$). Individuals in the TD group did not have a diagnosis of ASD, SZ, or any other psychiatric disorder. Participants in the ASD group had been previously diagnosed with ASD from a clinician practitioner according to the diagnostic criteria of the DSM-5 and research-reliable clinicians using the autism Diagnosis Observation Schedule (ADOS, Lord et al., 2000) and/or Autism Diagnostic Interview-Revised (ADI-R; Lord et al., 1994) confirmed diagnosis prior to this study. Participants in the ASD and TD group were matched for IQ [Wechsler Abbreviated Scale of Intelligence, Second Edition (WASI-2, Wechsler, 1999), TD = 109.52, SD = 12.43; ASD, M = 107.94, SD = 14.14, independent samples $t$-test, $t_{84} = 0.55, P = 0.58$]. A subset of the data from individuals with ASD presented here (26 of 46 individuals) has been published in a prior report (Noel et al., 2016a,b,c,d). Data from eight (17% of total) ASD and seven (17%) TD participants were excluded from analyses due to excessively poor causal inference fits (see below).

In Experiment 2, 16 subjects were individuals with SZ (eight males, age = 42.3 ± 8.9 years old), and another 16 were healthy age-matched and sex-matched controls (TD, six males, age = 41.9 ± 9.3 years old, independent samples $t$-test for age, $t_{30} = 0.12, P = 0.90$; Chi-square for sex; $\chi^2 (2, N = 32) = 0.97, P = 0.32$). Individuals in the TD group did not have a diagnosis of SZ, ASD, or any other psychiatric disorder. SZ symptoms were rated using the Brief Psychiatric Rating Scale (Overall & Gorham, 1962), the Scale for Assessment of Positive Symptoms (Andreasen, 1983b), and the Scale for Assessment of Negative Symptoms (Andreasen, 1983a). The sample of individuals with schizophrenia had an IQ within the normal range (SZ: $M = 104.0, SD = 8.0, 95\% CI = [88.3, 119.6]$; normative: $M = 100, SD = 15$). The data reported in Experiment 2 have been described in a previous report (Stevenson et al., 2016). Data from two (12% of total) SZ and one (6%) TD participants were excluded from analyses due to excessively poor causal inference fits (see below). All participants had
self-reported normal visual and auditory acuity. Vanderbilt University Medical Center’s Institutional Review Board approved all experimental protocols, and written informed consent was obtained from all participants and/or their guardians.

**Materials and apparatus**

Audiovisual stimuli were single-syllable utterances, which were selected from a stimulus set that has been previously used successfully in studies of multisensory integration (Stevenson et al., 2014c, 2016; Noel et al., 2016a,b,c,d). These stimuli consisted of two audiovisual clips of a female speaker uttering single instances of the syllables ‘ga’ and ‘ba’. Visual stimuli were grayscale (18.25 × 18.25 cm, 400 × 400 pixels, and subtended 17.38° of visual angle) and 2 s in duration. Each presentation contained the entire articulation of the syllable, including pre-articulatory gestures. The stimuli were presented at parametrically varied SOAs: 0 ms to ±300 ms in 50 ms intervals and ±400 ms, for a total of 15 levels of asynchrony. Negative SOAs indicate an auditory lead and positive SOAs indicate a visual lead.

**Procedure**

Participants sat in a dimly lit, sound-attenuated room (SE 2000 Series, Whisper Room Inc.) and instructed to perform a binary (synchronous or asynchronous) simultaneity judgment via button press. They were asked to continuously fixate toward a centrally presented white fixation cross (1 × 1 cm, black background) presented at a distance of 57 cm on a computer monitor (Samsung syncmaster 22-inch 2233 RZ LCD, refresh-rate = 120 Hz). Visual stimuli were presented on the computer screen, and auditory stimuli were presented through centrally aligned speakers at 65 dB (no interaural time or level difference). Each SOA was repeated 14 times for a total of 210 trials. The inter-trial interval was randomly jittered from 500 to 1500 ms (uniform distribution). MATLAB (The MathWorks, Inc.) with the PsychToolBox extension (Brainard, 1997; Pelli, 1997) was used for stimulus presentation and data collection. Trial orders were randomized. Finally, a trained experimenter actively monitored participants for compliance, and timing of stimuli was confirmed via oscilloscope (Hameg, 507).

**Analyses**

Two analysis approaches were undertaken. First, reports of synchrony were fit with a Gaussian curve to allow for comparison with previous reports and confirm that individuals in the ASD and SZ groups have larger TBWs than their respective control groups. Next, and novelty within the context of multisensory temporal acuity in psychopathology, reports of synchrony were analyzed via Bayesian causal inference (Magnotti et al., 2013; see Fig. 1). Non-parametric statistics are used throughout, as the assumption of normality was violated in most statistical contrasts. Single-subject fits of audiovisual speech synchrony reports using Bayesian causal inference model are illustrated in supplementary materials for completeness (see Figs S2 and S3).

**Psychometric fitting**

Reports of synchrony were averaged as a function of SOA, and these distributions were subsequently fitted with a scaled Gaussian distribution whose amplitude (AMP), mean, and standard deviation were free to vary (see Eqn 1, where y is the amplitude-scaled Gaussian distribution). The amplitude values were allowed to exceed a proportion of 1 (i.e., 100%). The mean of the best-fitting distribution was taken as the PSS and the distribution’s standard deviation as a measure of the TBW (Noel et al., 2017a,b,c; Van der Burg et al., 2013)

\[
y(x) = \text{amp} \times \exp\left(-\frac{(x-PSS)^2}{2\text{SD}^2}\right).
\]

**Causal inference model**

A detailed exposition and derivation of the causal inference model are beyond the scope of the current report and have been previously published (Magnotti et al., 2013; the analyses codes are all available online; http://openwetware.org/wiki/Beauchamp:CIMS); however, a brief explanation is provided here for completeness.

The causal inference model is a first-principles analysis of how the (in this case, temporal) relationship between multiple cues can be used to estimate the likelihood that auditory and visual events/objects emanate from a single talker vs. multiple talkers. Ecologically valid speech originating from a single talker (C = 1; Left column in Fig 1) naturally contains a small and variable (i.e., word-dependent), yet constrained, delay whereby the visual onset of speech (i.e., mouthing) precedes the audible component of speech (red distribution in left column in Fig. 1). This delay results in a distribution of asynchronies, which can artifically be placed at μ_c=1 = 0 ms (red distribution in left column in Fig. 1; X_v−X_a is centered around a delay equal to 0), and contains a certain (relatively small) variance (σ_c=1). If we do so, when there are two talkers (C = 2; Fig. 1, rightward column), there should be a priori no relationship between the visual and auditory onsets, resulting in a broad distribution of physical asynchronies (σ_c=2; black distribution in the right column in Fig. 1), which relative to the previously placed C = 1 distribution (i.e., μ_c=1 = 0 ms) ought to have a negative mean, μ_c=2 (see Magnotti et al., 2013; for a similar approach).

Further, importantly, observers do not have perfect knowledge of the physical asynchrony present in the environment (distribution of X_v−X_a), but instead must measure this asynchrony via their sensory systems, a process that is subject to individual-specific and trial-specific noise (σ_sensory noise; green distribution in Fig. 1). Present both in the case that one talker or two exist, and centered around the true asynchrony—green dot in Fig. 1). Overlaying the posterior C = 1 and C = 2 distributions (respectively red and black in the rightmost panel in Fig. 1), which result from Bayes theorem (see Eqn 2), shows that there is a window of measured asynchronies for which C = 1 is more likely than C = 2; this region (shaded in Fig 1) is the Bayes’ optimal synchrony window and is used by the observer to make the synchronous/asynchronous decision.

There are a total of six parameters: two subject parameters, a prior expectation that sensory events emanate from the same cause (P(C = 1)) and sensory noise (σ_sensory noise), and four stimulus parameters, the mean and standard deviations of the Gaussians that are taken to represented the likelihoods of sensory stimuli when there is a single cause [N(mean, standard deviation); N1(μ_c=1, σ_c=1)] and when there are two causes [N2(μ_c=2, σ_c=2)]. In this case, a single audiovisual speech condition is tested, however in the general form, when multiple conditions are tested, the subject parameters are shared across conditions, while the stimulus parameters are not (see Magnotti et al., 2013). Particular instances of auditory (X_a) and visual (X_v) events are drawn from the N1 and N2, and all parameters of the model are governed and jointly linked via Bayes’ rule (Eqn 2; see Shams & Beierholm, 2010, for more detail):
\[ P(C = 1|X_A, X_V) = \frac{P(X_A, X_V|C = 1)P(C = 1)}{P(X_A, X_V|C = 1)P(C = 1) + P(X_A, X_V|C = 2)(1 - P(C = 1))}. \] \tag{2}

Here, to simplify and increase robustness of the previous model (Magnotti et al., 2013) insight of utilizing it with clinical populations and with a single experimental condition—thus yielding an under-determined system if attempting to fit all possible free parameters—we set \( \mu_{C=1} = 0 \), \( \sigma_{C=1} = 65 \), and \( \sigma_{C=2} = 126 \) (the values estimated for the general population in the original paper, Magnotti et al., 2013; see Experiment 2 in Magnotti et al., 2013, for a similar approach). The rest of parameters are estimated via maximization of the binomial log-likelihood function on the observed data. \( P(C = 1) \) is initialized within a possible range between 0.01 and 0.99, \( \sigma_{\text{sensory noise}} \) is constrained within the range between 1 and 400, and finally, \( \mu_{C=2} \) is constrained within the \(-400\) to \(400\) ms range. Participants were excluded from statistical analysis when optimization of the parameters led to estimates within 5% of the boundary value for a parameter value. That is, we excluded participants hitting or approaching the boundary of permissible parameter values. Interestingly, fitting the full set of parameters [as opposed to solely \( P(C = 1) \), \( \sigma_{\text{sensory noise}} \), and \( \mu_{C=2} \)] results in an increased number of participants hitting the boundary of permissible parameter values (17 ASDs, 11 TD-ASDs, 1 SZs, and 2 TD-SZs) and an overall decrease in goodness of fit of the causal inference model to reports of synchrony (overall \( R^2 = 0.72 \) across all groups with fixed parameters, and \( R^2 = 0.69 \) across all groups with free parameters). A bootstrapping approach (see Supporting Information and Fig S1 for detail) estimated the noise-to-signal ratio in the parameter estimates resulting from the causal inference model fit to be at or under 10% for all parameters and subjects.

Results

**Experiment 1: autism spectrum disorder**

The Gaussian distributions describing the reports of synchrony as a function of SOA fit the data well across both the ASD and control groups (ASD, \( R^2 = 0.78 \); TD, \( R^2 = 0.82 \)). Consistent with prior work, psychometric fitting of synchrony judgments illustrates that those with ASD have significant larger (median = 496.87 ms, \( M = 514.95 \) ms,

© 2018 Federation of European Neuroscience Societies and John Wiley & Sons Ltd

*European Journal of Neuroscience, 47, 1230–1241*
SEM = 26.9 ms) TBWs than their TD (median = 482.79 ms, M = 404.38 ms, SEM = 24.28 ms) counterparts [Mann–Whitney U = 436, z = 2.20, P = 0.03, r = 0.23 (Pallant, 2007); see top panel in Fig. 2 for averaged data]. There was no significant difference between the groups among the other variables dictating the shape of the Gaussian distribution indexing synchrony judgments: AMP (Mann–Whitney U = 524, z = 1.18, P = 0.24, r = 0.13) and PSS (Mann–Whitney U = 767, z = 1.61, P = 0.11, r = 0.17).

The fitting procedure resulting from the causal inference model proved to accurately account for the reports of synchrony as a function of SOA (ASD, $R^2 = 0.74$, range = 0.47–0.97; TD, $R^2 = 0.84$, range = 0.48–0.96). When indexing the Bayes’ optimal window of synchrony judgments, calculated as the distance between the intersects (i.e., functions crossing) of the posterior likelihood that $C = 1$ (a single cause) and $C = 2$ (two causes), those with ASD have a larger window (median = 723.25 ms, $M = 788.60$ ms, SEM = 38.04 ms) than do those with TD (median = 681.29 ms, $M = 639.37$ ms, SEM = 40.21 ms; Mann–Whitney $U = 441$, $z = 2.14$, $P = 0.03$, $r = 0.23$; see Fig. 2, middle panel depicts the entire optimal window of synchrony judgments for ASD and TD participants. Top panel: averaged proportion of synchronous responses (y-axis) as a function of stimulus onset asynchrony (SOA; negative values indicate audio-leading SOAs and positive values indicate visual-leading conditions) and fit to the average (for visualization only). Temporal binding windows for audiovisual speech stimuli are larger for ASD (orange) than TD (blue) participants. Error bars represent ±1 SEM. Bottom panel: Top, $C = 1$ (continuous) and $C = 2$ (dashed) distributions as a function of SOA for a representative ASD (orange) and TD (blue) participant. Bottom, individual data for all ASD (orange) and TD (blue) participants. Each participant is represented by two dots indicating the location in time where their $C = 1$ and $C = 2$ curves intersect, one at position $x < 0$ and one at position $x > 0$ along the x-axis. The median for each group is represented via a box, the length of which illustrates the average Bayes’ optimal window of synchrony judgment for each group. [Colour figure can be viewed at wileyonlinelibrary.com].

Given the potential co-linearity between parameters of the causal inference model, initially a 2 (ASD vs. control) × 3 ($P(C = 1)$, sensory noise, and $\mu_{C>2}$) MANOVA was conducted. This analysis suggested that while simultaneously accounting for the variance across dependent variables, the ASD and control groups had significantly different causal inference model parameters ($F_{3,67} = 6.58, P < 0.001$, partial $\eta^2 = 0.228$, Wilks’ $\lambda = 0.772$). Subsequently, thus, separate Mann–Whitney $U$-tests were conducted for each of the three free parameters in the Bayesian model utilized to fit synchrony reports. The three parameters are as follows: (i) the prior that sensory events emanate from the same cause ($P(C = 1)$), (ii) sensory noise ($\sigma_{sensory noise}$), and (iii) the mean of the likelihood distribution from where stimuli originate when they have different causes $\mu_{C>2}$. As illustrated in Fig. 3, these analyses suggested that ASD participants (median = 0.56, $M = 0.59$, SEM = 0.014) had a significantly higher prior for common cause (Mann–Whitney $U = 873$, $z = 2.83$, $P = 0.004$, $r = 0.30$)
relative to TD participants (median = 0.52, M = 0.54, SEM = 0.014). On the other hand, these groups did not significantly differ regarding $\sigma_{\text{sensory noise}}$ (ASD, median = 183.43 ms, M = 198.53 ms, SEM = 11.02 ms; TD, median = 181.23 ms, M = 174.49 ms, SEM = 10.43 ms; Mann–Whitney U = 737, z = 1.27, P = 0.20, r = 0.13, Fig. 3B) or $\mu_{C>2}$, although for this last parameter, there was a trend toward a significant difference (ASD, median = –0.61 ms, M = –9.74 ms, SEM = 4.37 ms; TD, median = –19.58 ms, M = –28.85 ms, SEM = 7.54 ms; Mann–Whitney U = 774, z = 1.69, P = 0.09, r = 0.18, Fig. 3C). Hence, in ASD, the differences in multisensory temporal perception as indexed via synchrony judgments seem largely driven by differences in the likelihood to ascribe a common cause.

**Experiment 2: schizophrenia**

The Gaussian distributions describing the reports of synchrony as a function of SOA fit the data well across both the SZ and control groups (SZ, $R^2 = 0.79$; TD, $R^2 = 0.86$). Psychometric fitting of synchrony judgments suggested that individuals with SZ have significant larger (median = 544.33 ms, $M = 547.29$ ms $SEM = 22.16$ ms) TBWs relative to their TD (median = 333.04 ms, $M = 374.07$ ms, $SEM = 49.60$ ms) counterparts (Mann–Whitney U = 28, $z = 3.36$, $P = 4.21 \times 10^{-4}$, $r = 0.59$; top panel in Fig. 4 depicts averaged data). Further, there was no significant difference between the groups in the AMP (Mann–Whitney U = 63, $z = 1.83$, $P = 0.70$, $r = 0.32$) or PSS (Mann–Whitney U = 91, $z = 0.61$, $P = 0.56$, $r = 0.10$) parameters delineating the shape of the Gaussian best describing the reports of speech audiovisual synchrony.

As in the case of Experiment 1, the fitting procedure resulting from the causal inference model proved to accurately account for the reports of synchrony as a function of SOA (SZ, $R^2 = 0.82$, range = 0.59–0.95; TD, $R^2 = 0.89$, range = 0.82–0.96). Indexing of Bayes’ optimal windows of synchrony judgments via causal inference modeling also revealed that windows were larger in SZ (median = 848.46 ms, $M = 877.79$ ms, $SEM = 68.96$ ms) than in TD individuals (median = 566.158 ms, $M = 531.79$ ms, $SEM = 40.07$ ms; Mann–Whitney U = 24, $z = 3.53$, $P = 1.77 \times 10^{-4}$, $r = 0.62$; see Fig. 4, middle and bottom panels respectively for representative single subject data and all/averaged data).

Regarding the parameters governing the causal inference model, the initial 2 (SZ vs. control) × 3 ($P(C = 1)$, sensory noise, and $\mu_{C>2}$) MANOVA suggested global differences between the clinical group and its control ($F_{25} = 4.43$, $P = 0.012$, partial $\eta^2 = 0.347$, Wilks’ $\lambda = 0.347$). Subsequently, as in Experiment 1, separate Mann–Whitney U-tests were conducted for each of the three free parameters in the Bayesian causal model utilized to fit synchrony reports of SZ and TD participants. As illustrated in Fig. 5A, these analyses show that in contrast to the comparison between ASD and TD, the comparison between SZ (median = 0.62, $M = 0.67$, $SEM = 0.03$) and TD (median = 0.53, $M = 0.59$, $SEM = 0.03$) individuals did not demonstrate a significant difference (Mann–Whitney U = 148, $z = 1.87$, $P = 0.06$, $r = 0.33$) for probability of common cause, although a strong trend was apparent. The comparison between SZ (median = 155.52 ms, $M = 169.07$ ms, $SEM = 17.44$ ms) and TD (median = 122.75 ms, $M = 133.21$ ms, $SEM = 8.24$) individuals regarding the $\sigma_{\text{sensory noise}}$ parameter showed a significance difference (Mann–Whitney U = 150, $z = 1.96$, $P = 0.05$, $r = 0.34$, see Fig. 5B), although this difference just crossed the threshold for significance. Lastly, the $\mu_{C>2}$ parameters were not significantly different across the two groups (Mann–Whitney U = 109, $z = 0.175$, $P = 0.88$, $r = 0.03$, see Fig. 5C). Overall, in SZ, the differences in multisensory temporal perception as indexed via synchrony judgments do not appear largely driven by a single cause, but both by anomalies in the ascription of common cause and poor sensory fidelity.

**ASD vs. SZ comparison**

Lastly, we employed a bootstrapping procedure to directly compare $P(C = 1)$, $\sigma_{\text{sensory noise}}$, and $\mu_{C>2}$ values between the ASD and SZ groups. Importantly, it must be acknowledged that the explicit aim of the modeling work here was not to contrast ASD and SZ groups directly, but to ascribe their reported impaired multisensory temporal acuity relative to matched controls (Stevenson et al., 2014, 2017c; Noel et al., 2016a,b,c,d; Zhou et al., 2018) to particular principled variables. Indeed, the age and gender distribution of patients with ASD and SZ differ in the general population, and hence, matching these groups directly for age and gender would have supposed a deviation from naturalistic conditions. Similarly, contrasting absolute of $P(C = 1)$, $\sigma_{\text{sensory noise}}$, or $\mu_{C>2}$ values across groups would confound with age and gender differences across the psychopathological groups (however, see Supporting Information for this analysis). Thus, here we first normalize these parameters by their respective

---

© 2018 Federation of European Neuroscience Societies and John Wiley & Sons Ltd

European Journal of Neuroscience, 47, 1230–1241
Fig. 4. Temporal binding window and Bayes’ optimal window of synchrony judgments for SZ and TD participants. Top panel: averaged proportion of synchronous responses (y-axis) as a function of stimuli onset asynchrony (SOA, negative values indicate audio-leading SOAs and positive values indicate visual-leading conditions) and fit to the average (for visualization only). Temporal binding windows—indexed via Gaussian fitting—for audiovisual speech stimuli are larger for SZ (red) than TD (green) participants. Error bars represent ±1 SEM. Bottom panel: Top, $C = 1$ (continuous) and $C = 2$ (dashed) distributions as a function of SOA for a representative SZ (red) and TD (green) participant. Bottom panel: population data for SZ (red) and TD (green) participants. Each participant is represented by two dots, one at position $x < 0$ and one at position $x > 0$ along the x-axis. The median for each group is represented via a box, the length of which illustrates the Bayes’ optimal window of synchrony judgment for each group. [Colour figure can be viewed at wileyonlinelibrary.com].

Fig. 5. Comparison of Bayesian causal inference parameters inferred from fitting reports of audiovisual synchrony in SZ and TD participants. (A) Probability of common cause; there was a strong statistical trend suggesting that participants with SZ may have on average a higher general tendency to bind auditory and visual information together across time. (B): Sensory noise: SZ participants exhibited a greater degree of sensory noise than their TD counterparts. (C) Mean of the source likelihood when two causes are ascribed: no difference in $\mu_C$ between SZ and TD groups. Individuals dots are single participants (SZ = red; TD = green). Black horizontal lines represent the group’s mean. [Colour figure can be viewed at wileyonlinelibrary.com].
control groups. More precisely, for each parameter separately (e.g., $P(C=1)$) and for every patient separately (in the case of ASD, 46 patients), we subtract the patients estimate from the estimate of every subject in the matched control pool. Hence, in the case of patients with ASD, we built a distribution of 1840 $P(C=1)$ values (46 ASD patients $\times$ 40 ASD controls) which is centered on the true difference between ASD and their controls, and which has a variance that is representative of the joint (ASD and control) distribution. A similar procedure is undertaken for the SZ group, and finally, the two control-normalized distributions are contrasted via an independent samples $t$-test. As illustrated in Fig. 6, results suggested a significant difference between the ASD and SZ group regarding the $\sigma_{\text{sensory noise}}$ parameter ($t = 2.25, P = 0.02, 95\% CI = [-40.7, -2.85]$; Fig. 6 middle panel), but no significant difference for the other two parameters (all $Ps > 0.09$).

**Discussion**

A host of recent studies focused on multisensory temporal function have highlighted the important role that this process may have in scaffolding aspects of perception and cognition (Foss-Feig et al., 2010; Kwakye et al., 2011; Foxx et al., 2013; Stevenson et al., 2014a,b,c,d, 2017c; Noel et al., 2016c, 2017a,b,c; Su et al., 2015; Balz et al., 2016) with an emphasis on the domains of speech and communication (Woynaroski et al., 2013; Noel et al., 2017b). Recent work has also emphasized difference in multisensory temporal function in ASD (Stevenson et al., 2014a,b,c,d, 2017c) and SZ (Stevenson et al., 2016). Here, we extend this prior work by suggesting that while from a descriptive and phenomenological standpoint, ASD and SZ populations appear to judge multisensory synchrony similarly across a range of temporal disparities (Zhou et al., 2018), the cause behind these differences may be distinct. These potentially interesting mechanistic differences were revealed through the use of a causal inference model (Magnotti et al., 2013). Overall, this analysis suggested that while individuals with ASD may exhibit larger windows of temporal integration with respect to an age-matched and gender-matched control group due to anomalous priors (i.e., probability of ascribing common cause), SZ individuals’ differences in multisensory temporal function (with regard their matched control group) are likely a result from a combination of changes in priors as well as weaknesses in the nature of their sensory representations. Indeed, the direct comparison between the estimated parameter values governing the causal inference model for ASD and SZ groups suggested that these clinical groups differed solely in their degree of sensory noise (larger in SZ than ASD).

The current results highlight that whether examining temporal binding via psychometric fitting (i.e., TBWs) or Bayesian causal inference, those with ASD and SZ, categorize auditory and visual events as co-occurring in time or not. Thus, the fact that both patients with ASD and SZ exhibit larger Bayes’ optimal windows of synchrony is a novel finding distinct from the numerous previous reports of enlarged TBWs in these two populations.

The current results, in particular those regarding the prior probability of attributing a common cause to sensory events (i.e., $P(C=1)$), are timely within the framework of two recent and influential neurobiological theories regarding ASD. The predictive coding hypothesis of autism (Pellicano & Burr, 2012) proposes that many of the sensory abnormalities present in the ASD population may be a result of an attenuation of Bayesian priors. Stated differently, individuals with ASD are posited to possess ‘hypo-priors’, meaning that their expectations based on prior sensory experience influences their current percept less than it does in TD individuals (see Brock, 2012 and Pell et al., 2016; for a distinct interpretation from that of Pellicano & Burr, 2012). Although the interpretation of the current findings must remain agnostic regarding the directionality of the effect (i.e., the current results

---

**Fig. 6.** Direct comparison of Bayesian causal inference model parameter values across ASD and SZ groups. The distributions depicted (ASD = orange; SZ = red) are bootstrapped differences between the individuals composing the psychopathological group and their respective control groups. That is, the distributions are centered on the true difference between the clinical group and their control, and have a variance that corresponds to subject-to-subject contrasts. Top panel: distributions of $P(C=1)$ differences between patients and controls. Solid vertical line indicates the median of the distribution and ‘ns’, non-significant. Middle panel: distributions of $\sigma_{\text{sensory noise}}$ differences between patients and controls. In this case, the distributions of $\sigma_{\text{sensory noise}}$ differences were statistically different from one another (*indicates $P < 0.05$). Bottom panel: distributions of $\mu_{C=2}$ differences between patients and controls. [Colour figure can be viewed at wileyonlinelibrary.com].
could be interpreted both as indicating a stronger prior for common cause—integration—and a weakened prior for separate causes—segregation), they do indeed point toward alterations in the prior applied during the examination of the sensory world. Somewhat similarly, it has recently been postulated (Dinstein et al., 2015) that increased variance of neural responses in ASD may be at the root of their sensory disturbances (Simmons et al., 2009; Dinstein et al., 2012; Uhlhaas & Singer, 2012 Davis & Plaisted-Grant, 2015; Simmons & Milne, 2015). In this case, the interpretation of the current results is more straightforward, in that examination of the sensory noise parameter suggested no difference between the ASD population and the tested control group (see Butler et al., 2016, for more recent evidence arguing against the case that autistic individuals have more ‘unreliable’ brains).

In terms of SZ, dysfunctional information binding has been described since the first descriptions of the condition (Bleuler, 1911) and it is a feature that persists in modern theories of SZ, specifically for patients experiencing hallucinations (Behrendt & Young, 2004). Interestingly, it has recently been suggested that there is a strong relationship between multisensory temporal acuity and proneness to hallucinations (Stevenson et al., 2016,; see Noel et al., 2017a,b,c, under revision; for a similar argument regarding multisensory processing in the spatial dimension). More generally, aberrant sensory integration may result in perceptual incoherence that manifests in many of the clinical symptoms of SZ (Postmes et al., 2014). Interestingly, the neural substrates underpinning multisensory temporal function and the production of hallucinations are overlapping (namely the superior temporal sulcus and inferior frontal gyrus; Kim et al., 2013), and the association between the latter (i.e., proneness to hallucinations) and unreliable sensory representations is commonplace (Shergill et al., 2000). Here, we add to current conceptions of SZ by highlighting that impaired multisensory temporal function may be rooted in both impaired quality of sensory representation and differences in causal ascription and may consequently play an important role in perceptual incoherence.

An interesting area of investigation the current results apply to is that of perceptual learning within psychopathology and the potential clinical utility of perceptual learning as a tool in remediation. In fact, multisensory temporal function has been shown to be highly plastic and work has shown that multisensory temporal acuity can be improved via perceptual training (Powers et al., 2009, 2012; Stevenson et al., 2013, 2017a; Schlesinger et al., 2014; De Nieg et al., 2017). Clinically, this approach may hold promise in strengthening sensory and perceptual representations, which may cascade into higher-order benefits, such as enhanced speech comprehension. Within such a therapeutic context, the current results might suggest two different approaches toward the remediation of multisensory temporal dysfunction in ASD and SZ. That is, it appears that in SZ, there would be benefit from training protocols that make sensory representation more reliable, whereas in ASD, this approach would seem likely to have less impact. Conversely, in both ASD and SZ, approaches that focus upon strengthening priors and ascription of common cause would be predicted to yield benefit. Future work founded in this evidence should seek to design training protocols that selectively benefit these different aspects of multisensory temporal perception and that may ultimately be ushered together to create individualized training protocols.

Before fully implementing these individualized training protocols, however, a few words of caution merit to be exposed. Indeed, the current results suggest that partially different causes (e.g., the addition to sensory noise in SZ) may drive the phenotypically similar multisensory temporal function impairments in ASD and SZ; however, a number of limitations must be raised. First, although we do directly contrast the ASD and SZ groups for their estimated prior probability of ascribing a common cause, their sensory noise, and their expected asynchrony between audio and visual speech-related events when these do not belong to a single cause, these contrasts are executed solely after normalizing by their respective control groups. We consider this correction important within the current dataset, as the clinical groups differed in a number of factors beyond their clinical diagnosis, for example, age and gender distribution. However, in the future, it may be interesting to directly compare age-matched and gender-matched ASD, SZ, and control groups. This putative protocol would depart from what is observed in the general populations (i.e., patients with ASD and SZ typically differ in age) but would equally allow for contrasting multisensory temporal function before model fitting. In this same vein, it must be emphasized here that while we contrast ASD and SZ groups for the parameters of the Bayesian causal inference model (Magnotti et al., 2013), the reliability of these contrasts is limited by the explanatory power of the model itself. Secondly, we must highlight that the size of the SZ sample tested was about a third the size of the ASD sample (16 Szs vs. 46 ASDs), and the fact within the former group, statistical contrasts were merely marginally significant. Thus, in the future, it will be important to extend the current results to a larger and more diverse population.

In conclusion, in the current study, individuals with ASD and SZ judged the synchrony between speech audiovisual stimuli, and their reports were analyzed both via a common psychometric fitting procedure and a via a Bayesian causal inference model. While from a gross phenotypic standpoint the results corroborated prior work showing similar multisensory temporal deficits in the two clinical populations, the modeling approach suggested that the mechanistic basis for these temporal differences may be partially distinct. While ASD subjects appear to form appropriate sensory representations (relative to their age-matched and gender-matched controls), they did not appear to have learned through experience (or at least not as strongly as TD individuals) that events occurring in close temporal proximal are likely to emanate from the same cause. In comparison, patients with SZ exhibit difference in multisensory temporal function (vs. age-matched and gender-matched controls) that is seemingly the result of both differences in ascription of common cause and unreliability in sensory representations.

Supporting Information

Additional supporting information can be found in the online version of this article:

Fig. S1. Uncertainty associated with causal inference model parameter estimates. The difference (y-axis) between the ‘real’ estimates (based on synchrony reports) and simulated estimates (based on a bootstrapping approach detailed in text) is depicted for every subject (x-axis) and all parameters (leftmost: P(C = 1); center: sensory noise; rightmost: μ(C = 2)). Zero (on y-axis) indicates no difference between real and mean bootstrapped estimates, and error bars depict ±1 standard deviation.

Fig. S2. Individual-subjects reports of audiovisual speech synchrony for ASD (left panel) and TD (right panel) participants fitted via the causal inference model. X-axis is stimulus onset asynchrony (SOA) and the Y-axis is the report of synchrony. In each subpanel, each black dot is the mean report of synchrony for a given SOA. Participants that were kept for the group analyses reported in the main text are fitted with a blue continuous line, while those that were rejected are fitted with a red continuous line.
Fig. S3. Individual-subjects reports of audiovisual speech synchrony for SZ (left panel) and TD (right panel) participants fitted via the causal inference model. X-axis is stimulus onset asynchrony (SOA) and the Y-axis is the report of synchrony. In each subpanel, each black dot is the mean report of synchrony for a given SOA. Participants that were kept for the group analyses reported in the main text are fitted with a blue continuous line, while those that were rejected are fitted with a red continuous line.

Acknowledgements

The authors thank Dr. John Magnotti and Dr. Michael Beauchamp for their thorough help in implementing and interpreting their Bayesian causal inference model.

Funding

JPN was supported by the National Science Foundation Graduate Research Fellowship (NSF GRF) under Grant No 1445197. RAS is funded by NSERC Discovery Grant (RGPIN-2017-04656), SSHRC Insight Grant (435-2017-0936), and the University of Western Ontario Faculty Development Research Fund. MTW was supported by NIH grants MH109225, CA183492, and HD083211 and the Wallace Foundation.

Conflict of interest

The authors declare that they have no competing interests.

Data accessibility

The datasets analyzed and analyses tools during the current study are available from the corresponding author on request. These are not posted on an online repository, as this was not indicated on original consent forms.

Author contributions

JPN, RAS, and MTW designed the experiment. JPN and RAS collected the data, which was analyzed by JPN. JPN drafted the manuscript, which was edited by RAS and MTW. All authors approved the final version of the manuscript.

References

Andreasen, N. (1983a) The Scale for the Assessment of Negative Symptoms (SANS). University of Iowa, Iowa City, IA.

Andreasen, N. (1983b) The Scale for the Assessment of Positive Symptoms (SAPS). University of Iowa, Iowa City, IA.


© 2018 Federation of European Neuroscience Societies and John Wiley & Sons Ltd
European Journal of Neuroscience, 47, 1230–1241

Causal inference in schizophrenia and ASD 1239


