



Deposited via The University of York.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/212672/>

Version: Published Version

---

**Article:**

Oakes, Richard, Peschel, Lisa Anne and Barraclough, Nick (2024) Inter-subject correlation of audience facial expressions predicts audience engagement during theatrical performances. *iScience*. 109843. ISSN: 2589-0042

<https://doi.org/10.1016/j.isci.2024.109843>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

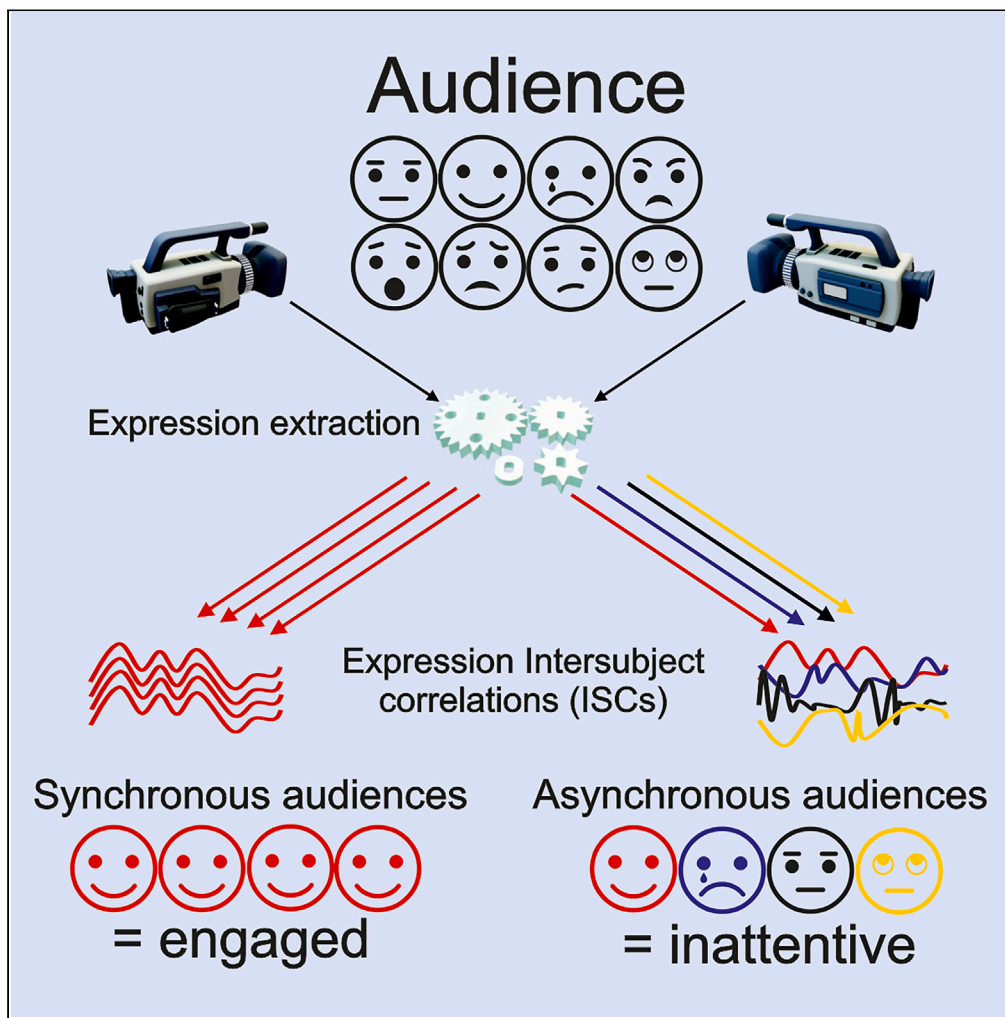
<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.

Article

# Inter-subject correlation of audience facial expressions predicts audience engagement during theatrical performances



Richard A. Oakes,  
Lisa Peschel, Nick  
E. Barraclough

nick.barraclough@york.ac.uk

**Highlights**

During performances audience facial expression synchrony predicts their engagement

Expression synchrony increases with spatial proximity within the auditorium

Younger, female, and more empathetic individuals show greater expression synchrony

Videoring faces of large groups allows real-time non-invasive measures of engagement

Oakes et al., iScience 27, 109843  
June 21, 2024 © 2024 The Authors. Published by Elsevier Inc.  
<https://doi.org/10.1016/j.jisci.2024.109843>



## Article

## Inter-subject correlation of audience facial expressions predicts audience engagement during theatrical performances

Richard A. Oakes,<sup>1,2</sup> Lisa Peschel,<sup>2</sup> and Nick E. Barraclough<sup>1,3,\*</sup>

## SUMMARY

During performances, audiences experience various emotional states, and these are reflected in their ongoing facial expressions. We investigated if audience engagement could be determined by measuring the inter-subject correlation (ISC) of non-invasively recorded audience facial expressions. We filmed the faces of multiple audience members at theatrical performances and determined the intensity of their different facial expressions throughout the performances. Neutral, happy, anger, and disgust expression ISCs accounted for up to 24% of the performance dramaturge's predictions of audience engagement. Expression synchrony was greater between individuals in close proximity, suggesting effects of emotional contagion or cognitive similarities between neighboring individuals, whereas expression synchrony was greatest between individuals who were younger, female, and with greater levels of empathy, showing that individual characteristics impact shared audience experiences. Together, our results show that facial expression synchronization could be used as a real-time non-invasive indicator of engagement in audiences larger than achieved using previous approaches.

## INTER-SUBJECT CORRELATION OF EMOTIONAL EXPRESSIONS IN AUDIENCES DURING THEATRICAL PERFORMANCES

Measuring the facial expressions of members of an audience has recently helped our understanding of their experiences when watching artistic performances. Emotional experiences are one of the key reasons people attend artistic performances and are quantitatively different when experienced live and within an audience group.<sup>1–3</sup> During artistic performances individuals in the audience experience various emotional states, and different emotions are often reflected in specific changes in the facial expressions of the individual.<sup>4–6</sup> Information from facial expressions during the performance can enable an understanding of the emotional experience of the individual audience members complementary to that obtained by self-reports.<sup>7</sup> In this prior study, we filmed the faces of the audiences seated in an auditorium while they listened to a musical concert and subsequently determined their different facial expressions using face expression analysis software. Pieces of music that expressed sadness resulted in more facial expressions of sadness, whereas pieces of music that expressed happiness resulted in more facial expressions of happiness. This close relationship between audience facial expressions during the performance and their subsequent reports of the emotional experiences they experienced permits a non-invasive method of determining how audiences may experience artistic performances.

The evaluation of the ongoing facial expressions of audience members provides information about audience experiences not possible with other methods. Traditionally, audience experiences of artistic performances have been studied using self-reports.<sup>3,8,9</sup> Self-reports, however, are potentially open to bias and only allow a retrospective analysis of emotional experiences. In contrast, measuring the expressions produced by audience members has the potential to provide time-locked moment-by-moment information about emotions experienced during long performances. Although ongoing emotional experiences have previously also been recorded using physiological measures of skin conductance<sup>10</sup> or facial electromyography,<sup>11</sup> these methods have drawbacks. The attachment of electrodes to audience members is time-consuming and invasive and thus impacts the audience experience of the performance. In addition, this becomes unfeasible in larger audiences as the number of individuals that can be studied is restricted by the availability of equipment. In contrast, high temporal resolution measurements of dynamic changes in facial expressions of large audiences are possible using unobtrusive filming of faces, allowing estimates of emotional experiences within ecologically valid environments.<sup>7</sup>

In addition to information about the emotions of audience members, facial expressions may also provide information about audience engagement. Audience engagement is a multifaceted concept that has traditionally been difficult to measure, as discussed in multiple reports.<sup>9,12–14</sup> It is thought to comprise affective responses, attention, cognitive interactions, empathy processes, and social, cultural, and

<sup>1</sup>Department of Psychology, University of York, Heslington, YO10 5DD York, UK

<sup>2</sup>Department of Theatre, Film and Television and Interactive Media, University of York, Heslington, YO10 5DD York, UK

<sup>3</sup>Lead contact

\*Correspondence: [nick.barracough@york.ac.uk](mailto:nick.barracough@york.ac.uk)

<https://doi.org/10.1016/j.isci.2024.109843>



contextual influences. However, during periods of high engagement with the performance, it is thought that audience members enter a form of “flow”<sup>15</sup> where they “get lost” in the performance or are captivated by it. This state of captivation is not dependent upon high degrees of specific emotions, but rather the consistency of how the performance affects the emotional state of the members of the audience. Consequently, the degree of emotional synchronization within the audience as a whole, which reflects how consistently the performance influences the emotional states of all audience members, could potentially be a useful measure of audience engagement. High states of emotional synchrony where there is increased behavioral “alignment”<sup>16</sup> would occur during high audience engagement with the flow of the narrative and emotional content of the performance. In contrast, lower emotional synchrony would occur when the audience members were not attending, having their mind wandering or thinking about topics other than the performance.

Indeed, studies have recently shown that the degree of audience engagement is reflected in the synchrony between a range of different physiological<sup>17–20</sup> and neural<sup>21–23</sup> signals of audience members. These methods were originally developed to analyze the reliability of responses of multiple individuals during the multidimensional dynamics of real-world social contexts.<sup>24–26</sup> Here, the inter-subject correlation (ISC) between individuals provides a measure of shared experiences of complex naturalistic experiences that unfold over many minutes. ISCs can show how differently we synchronize or “tick together,” during emotional experiences,<sup>27</sup> and variance in ISCs reflects differences in experiences between individuals with different psychological characteristics.<sup>28</sup> Consequently, the degree of ISC of a range of different measured signals is related to the degree of audience engagement.<sup>29</sup> For example, synchronization in audience heart rates is related to self-reported narrative engagement when viewing television and film content.<sup>19</sup> Furthermore, during music listening, high ISCs of facial muscles, cardiorespiratory, and skin conductance responses were consistently related to periods of arousing emotions and audience engagement.<sup>18</sup> Whereas peak correlations in neural activity measured using electroencephalography (EEG) correspond also with arousing moments during film viewing and reflect attention- and emotion-modulated processing.<sup>30</sup> And large-scale audience engagement with, and preference for, television adverts can be predicted by peaks in neural synchrony measured in a different and smaller sample of viewers.<sup>22</sup> Thus, ISCs of a range of different signals have all been shown to be good measures of audience engagement.

Given these findings, in this study, we investigated if audience engagement could be determined by measuring the ISC of non-invasively recorded audience facial expressions. Prior research has relied on intrusive methods of measuring audience responses requiring the placement of multiple electrodes on the face, scalp, or torso and have been restricted to relatively small ( $n < 16$ ) audiences. Measuring audience facial expression ISC synchrony has the potential to allow us to assess audience engagement in much larger groups of individuals while also limiting intrusive measures that detract from the audiences’ experience of the artistic performances. Therefore, in this study, we tested if this was possible by videoing the faces of multiple audience members at several different theatrical and cabaret performances. In order to ensure the viability of this method “in the wild,” we ensured the performances were as naturalistic as possible, with actors performing in public auditoriums while audience members were filmed with unobtrusive cameras. We tested our hypothesis that facial expression ISCs would be related to predictions of audience engagement made by the performance “dramaturge” (the expert in the piece of work performed who advised the director and other collaborating artists on the production). Strong evidence of such a relationship would indicate the potential utility for this method to determine audience engagement non-invasively in the future.

There are also other possible sources of behavioral alignment between different individuals during a performance at a theater,<sup>16</sup> including between the actors on stage, between actors and the audience, between individual audience members in the auditorium, or across the audience more broadly. During such shared experiences individuals become “aligned”<sup>31</sup> physiologically, mentally, and in their expression of emotions in their facial movements. There is much research on face-to-face dyadic interactions that have shown that physiological measures<sup>32–34</sup> and facial expressions<sup>35–37</sup> are often synchronized or highly similar, resulting from a continuous mutual adaptation between the two individuals. This type of “dynamic coupling”<sup>31</sup> where one individual is interacting with another individual and each are reacting to the facial expressions of the other is unlikely to contribute significantly to any putative audience expression synchrony. During performances, audience members are not typically interacting directly with each other, instead they will be spending more time facing toward the stage and attending the performance (as confirmed by our recording). Any such interactions will likely be limited to pre- and post-performance discussions between neighboring individuals who are familiar with each other, or at least during performances these interactions will be fleeting. Actor-actor interactions are not the focus of our study, and our methods cannot track their expressions as they move around on stage. Instead, the interaction of interest is that between each audience member and the “performance” by the actor(s) as a collective, which is a function of the degree of audience engagement in the performance.

However, even if audience members do not interact intentionally, they can have a degree of synchrony in their emotional responses due to their shared experience.<sup>38</sup> Pairs of individuals watching a movie together show synchronization of autonomic responses<sup>39,40</sup> and facial expressions,<sup>40–42</sup> which are closely linked.<sup>43,44</sup> Putative mechanisms by which this occur may involve mimicry,<sup>45</sup> a tendency to automatically synchronize affective expressions of another person, or emotional contagion, a related concept where we “catch” other people’s emotion.<sup>46,47</sup> The strongest effects of such emotional synchronization are likely to occur between individuals who are closest to each other in the auditorium for two reasons. First, individuals in closer proximity will be more sensitive to the small changes in the behavior of their immediate neighbors and experience emotional contagion, e.g., see<sup>39,41,46–48</sup>. Second, individuals who attend a performance together will likely be sitting in seats next to, or nearby each other, and may share some similar interests and personal characteristics. Such social similarities may result in increased behavioral and emotional alignment between these individuals. Therefore, we tested the hypothesis that the spatial proximity of individuals within a large audience influences the degree of measured facial expression synchrony.

In addition, it has been shown that the characteristics of individuals can impact ISC of both physiological<sup>49–53</sup> and neural<sup>21,27</sup> signals. For example, skin conductance synchrony between patients and therapists can be related to perceptions of perceived empathy.<sup>50,51</sup> Whereas the ISC of neural activity within the posterior middle temporal gyrus (MTG), a region of cortex associated with reading intentions of other

**Table 1. Average magnitudes and variances in expressions during performances**

Expression	Magnitude	Standard deviation
Neutral	0.68 (0.16)	0.16 (0.05)
Happy	0.10 (0.12)	0.14 (0.09)
Sad	0.08 (0.08)	0.10 (0.07)
Anger	0.08 (0.12)	0.10 (0.10)
Surprise	0.01 (0.02)	0.03 (0.04)
Fear	0.02 (0.04)	0.04 (0.04)
Disgust	0.02 (0.05)	0.03 (0.05)
Contempt	0.065 (0.06)	0.09 (0.07)

Note: Expression magnitude varies between 0 and 1; the value in brackets is the standard deviation in these values across participants. Standard deviation refers to average variance in the expression over the duration of the performance; the value in brackets is the standard deviation in these values across participants.

individuals and more generally with empathy and mentalizing,<sup>54,55</sup> is positively associated with self-reported emotional empathy scores.<sup>27</sup> In addition, individual empathy has also been shown to predict synchrony in small groups of students within the classroom.<sup>21</sup> Given this prior research in both dyads and groups, we predicted that the empathy of those individuals within large audiences at theatrical performances would impact the degree to which their expressions were synchronized. We hypothesized that pairs of individuals in the audience showing higher degrees of empathy would show a greater degree of facial expression ISCs.

Finally, the age and the gender of individuals have been shown to also impact ISC in some cases. Older participants can have reduced ISC when measured using functional magnetic resonance imaging (fMRI)<sup>56,57</sup> or EEG,<sup>58,59</sup> maybe because idiosyncratic brain activity arises with age, leading to increased neural variance and thus a decrease in ISC. We therefore tested whether the age of individuals influenced face expression ISCs and hypothesized that expression ISC would decrease with age.

The impact of gender on ISC is more complex, however. In children, but not adults, greater EEG ISC is seen in males than females.<sup>58</sup> This may reflect that in younger individuals males have comparably reduced brain development than females of the same age,<sup>60</sup> whereas comparatively advanced brain development in the equivalently aged young females results in greater neural variance and thus lower EEG ISCs. Others, however, have observed higher ISC in males compared to females in some brain regions (e.g., posterior and central midline), but lower ISCs in males in other brain regions, e.g., temporal parietal junction and inferior temporal cortex.<sup>56</sup> Given this mixed picture, we could not make strong predictions about the impact of gender of audience members on facial expression ISCs, but instead explored the impact of gender in younger and older individuals within the audience.

In summary, in this paper we report several investigations into the synchrony of facial expressions within large audiences attending theatrical performances. We first describe the extent of the synchrony between a number of different emotional expressions that we could derive from video footage of audience faces. We then test the hypothesis that audience engagement can be determined by measuring the ISC of audience facial expressions. The aim here was to assess the potential utility for this method to determine audience engagement non-invasively. We additionally test whether increased spatial proximity of audience members increases expression synchrony. Finally, we assess the influence of audience member empathy, age, and gender on expression synchrony.

## RESULTS

For the 193 audience members who met our criteria for inclusion, we calculated the mean magnitude of each expression, and the standard deviation of each expression, throughout the entirety of the performance for each individual (see Table 1). These values show the degree to which each expression is on average present during the performances and how much these expressions vary during the performances. The neutral expression was detected most commonly, as indicated by a large mean magnitude, whereas the emotional expressions were detected less. The happy, sad, and anger expressions were the most commonly present emotional expressions. There was considerable variance in expressions during the performances, in particular the neutral, happy, sad, and anger expressions. There were no significant differences in mean or variance of expressions between female or male individuals.

To determine whether FaceReader treated the seven different emotional expressions, and the neutral expression, as mutually exclusive or not, for each individual we correlated these eight signals recorded during the performance. The average *r* values (see Table 2) indicated a substantial degree of independence of these expressions, reflecting FaceReader's method of classification of expressions based upon face action unit activity. Neutral expressions were moderately negatively correlated with the happy expressions, but only weakly negatively correlated with the other expressions. All other expressions were only very weakly correlated with each other.

### Facial expressions are correlated within audiences

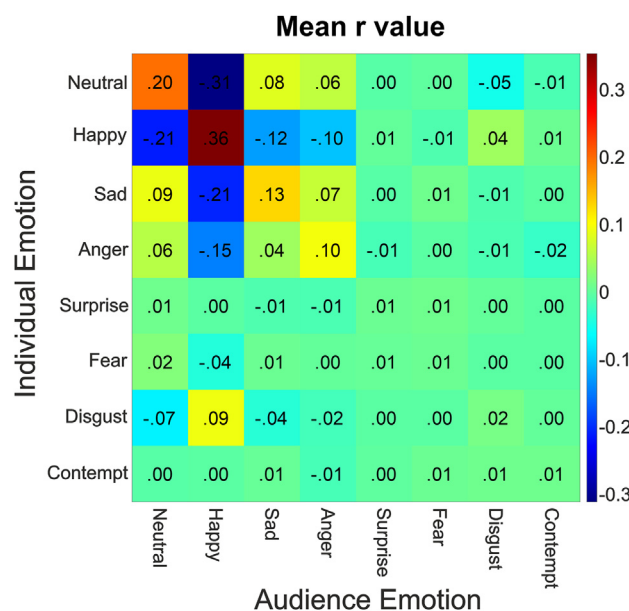
In order to understand how an individual's expressions were related to those produced by the rest of the audience, we examined the correlations between all eight expressions measured. The eight expressions produced during the whole performance for each individual were correlated with the averages of the eight expressions from the rest of the audience (audience *n*-1).

**Table 2. Mean (standard deviation) between-expression correlations for all participants**

	Neutral	Happy	Sad	Anger	Surprise	Fear	Disgust
Happy	-0.56 (0.33)						
Sad	-0.20 (0.31)	-0.17 (0.15)					
Anger	-0.18 (0.32)	-0.12 (0.16)	-0.04 (0.17)				
Surprise	-0.01 (0.18)	0.03 (0.15)	-0.05 (0.14)	-0.05 (0.13)			
Fear	-0.06 (0.18)	0.00 (0.16)	-0.00 (0.15)	-0.00 (0.16)	0.17 (0.18)		
Disgust	-0.14 (0.20)	0.07 (0.18)	0.03 (0.20)	-0.02 (0.17)	0.05 (0.14)	0.12 (0.19)	
Contempt	0.14 (0.15)	-0.03 (0.15)	-0.01 (0.13)	-0.05 (0.11)	-0.01 (0.11)	-0.02 (0.11)	-0.03 (0.12)

To understand how audience expressions are correlated with each other across all 16 performances, we calculated the mean *r* values for expression correlations from all 193 similarity matrixes and present this as a single similarity matrix in Figure 1. Here, a larger positive mean *r* value indicates that for a particular expression in individuals there is a positive correlation with a particular expression in the rest of the audience as a whole (i.e., for those expressions, audience expressions are in synchrony). Negative mean *r* values, in contrast, indicate that for a particular expression in individuals there is negative correlation with a particular expression in the rest of the audience as a whole (i.e., for those expressions, the audience expressions are asynchronous).

As might be expected given that each member of an audience experiences the same performance, expressions are largely correlated within-expression. For example, the level of happiness expressed by one individual appears associated with the level of happiness expressed in the rest of the audience as a whole. This suggests a degree of shared emotional experience across audience members. Of all the expressions measured, happy expressions appear most correlated with each other, and these appear negatively correlated with non-happy expressions, in particular neutral, sad, and angry expressions. To assess the degree to which synchrony is within-expression, we compared mean within-expression (on-diagonal) and mean between-expression (off-diagonal) correlation values with a Bayesian paired-samples *t* test. Our null hypothesis was that there was no difference between within-expression correlations and between-expression correlations. Our alternative hypothesis was that within-expression correlations would be greater than between-expression correlations. The Bayes factor ( $BF_{10} = 1.33 \times 10^{52}$ ) indicated extreme evidence for our alternative hypothesis that within-expression correlations were greater than between-expression correlations. The error % could not be calculated as the  $BF_{10}$  value was so large; however, the median of the resulting median posterior distribution was 1.60, with a central 95% credible interval that ranged from 1.39 to 1.81. Thus, as audience expressions are more correlated within-expression, it suggests the emotions of individuals within an audience tend to be in synchrony with each other.



**Figure 1. Relationship between expressions of each audience member and the average expressions calculated across the rest of the audience**

Expressions of the individuals are on the y axis, expressions of the rest of the audience are on the x axis. Redder colors indicate larger positive *r* values, and bluer colors indicate larger negative *r* values; mean transformed *r* values are indicated within each cell.

**Table 3. Within-expression correlations vs. between-expression correlations**

Expression	BF <sub>10</sub>	Error %	Median posterior distribution	Central 95% credible interval
Neutral	5.72 × 10 <sup>27</sup>	NaN	0.987	[0.815, 1.161]
Happy	3.02 × 10 <sup>54</sup>	NaN	1.662	[1.441, 1.882]
Sad	2.71 × 10 <sup>26</sup>	NaN	0.956	[0.785, 1.127]
Angry	1.35 × 10 <sup>21</sup>	NaN	0.798	[0.827, 0.992]
Surprise	2.46	0.0004	0.169	[0.037, 0.309]
Fear	0.45	<0.0001	0.112	[0.011, 0.249]
Disgust	199	<0.0001	274	[0.132, 0.417]
Contempt	1.36	0.0004	0.150	[0.026, 0.029]

We then assessed how well each of the different facial expressions were correlated with themselves compared with their correlations with the other expressions. We did this by first averaging the seven between-expression (off-diagonal) correlations for each expression separately and then compared these values to the respective within-expression (on-diagonal) correlation with separate Bayesian paired-samples t tests. The Bayes factors (Table 3) indicated extreme evidence for our alternative hypothesis that within-expression correlations were greater than between-expression correlations for the majority of expressions including neutral, happy, sad, angry, and disgust expressions and anecdotal evidence for a difference for surprise and contempt expressions. Whereas Bayes factors indicated anecdotal evidence that fear within-expression correlations were not different from fear between-expression correlations.

#### Facial expression inter-subject correlations are related to predictions of audience engagement

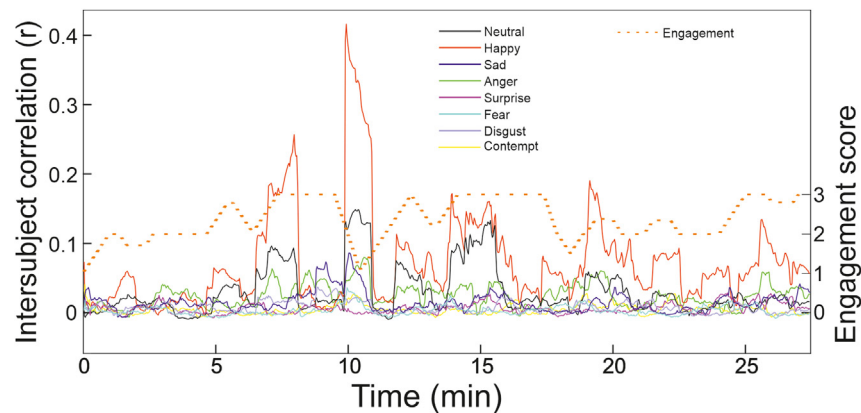
Expression ISCs varied considerably over the duration of a performance and were independent of the absolute magnitude of the expressions themselves (see Figure S1; Table S1). Audience expression synchronization could be considerable at points, for example with peaks in correlation of happy expressions greater than  $r = 0.4$  during a performance in York in 2017. On average, happy expressions were the most synchronized (mean  $r$  across performances = 0.070, SD = 0.004, max = 0.416). Indeed, happy expressions were typically over twice as synchronized as the neutral expressions and over four times as synchronized as sad expressions (see Table 4). What is most evident is that there is considerable variation in the ISCs over the duration of the performances, indicating that while audiences freely viewed complex multimodal theatrical performances there are notable periods when the emotional state of audiences appear to “tick together.”<sup>61</sup>

A comparison between the expression ISCs and the dramaturge’s predictions of audience engagement across three different performances (Figure 2) showed that the neutral, happy, anger, and disgust expression ISCs were consistently and positively correlated with estimated audience engagement across the three performances (see Table 5). Bayes factors indicated extreme evidence for a positive relationship between engagement and neutral, happy and disgust expression ISCs in all three performances, and for angry expression in two performances, with moderate evidence for angry expressions in one performance. Bayes factors indicated very strong evidence for sadness, fear, and contempt expression ISCs being positively correlated with audience engagement in one performance. Finally, the Bayes factor indicated extreme evidence for surprise expression ISCs being negatively correlated with audience engagement in one performance.

We then calculated three separate (one for each performance) Bayesian linear regression models to examine the predictive value of expression ISCs on the estimate of audience engagement. For all three performances, following observation of the data, the odds in favor of models where expression ISCs predicted audience engagement increased. Optimal models are described in Table 6; however, the odds in favor of many other models where different combinations of expression ISCs predicted audience engagement also increased, but for brevity

**Table 4. ISCs for three performances**

Expression	Maximum $r$	Mean $r$
Neutral	0.17	0.03
Happy	0.42	0.07
Sad	0.09	0.02
Anger	0.09	0.02
Surprise	0.05	0.01
Fear	0.05	0.00
Disgust	0.05	0.01
Contempt	0.04	0.01



**Figure 2. Inter-audience expression synchrony is related to predictions of audience engagement**

Example from a single performance at York in 2017. Expression synchrony is illustrated as raw waveforms representing fluctuations in ISC  $r$  values for each expression over the duration of the performance (solid lines). Prediction of audience engagement on a 1–3 scale at each point in the performance is also illustrated as a waveform (dotted orange line).

are not reported here. For the optimal models, Bayes factors indicated very strong evidence that expression ISCs predicted audience engagement, and they accounted for 24%, 19%, and 22% of the variance in the three performances.

### ISCs are influenced by audience member proximity

In line with our prediction, there were small negative correlations between the distances between audience members and expression ISCs, indicating that the further apart audience members were from each other, the less their expressions were synchronized (see Figure 3). Bayes factors indicated extreme evidence for a negative relationship between audience member proximity and neutral ( $BF_{10} = 11088$ ), happy ( $BF_{10} = 2317$ ), sad ( $BF_{10} = 916$ ), angry ( $BF_{10} = 498$ ), surprise ( $BF_{10} = 276$ ), and fear ( $BF_{10} = 6665$ ) expressions and anecdotal evidence for this relationship with disgust ( $BF_{10} = 2.68$ ) and contempt expression ISCs ( $BF_{10} = 1.18$ ). A permutation test (see Table S2) indicated that all  $r$  values shown in Figure 3 were robust.

To test whether audience member spatial proximity predicted the different expression ISCs, we calculated eight separate Bayesian linear regression models. Following observation of the data, the odds in favor of models where audience member proximity predicted facial expression ISCs ( $BF_M$ ) increased for all expressions (see Table 7).

### ISCs are dependent upon audience member age, gender, and level of empathy

We tested the impact of audience member age and level of empathy on facial expression ISCs. We predicted that ISCs would decline with the age of audience pairs, but increase with audience pair empathy. For each audience pair we calculated the combined (total) age as well as the combined (total) empathy score. Bayesian correlations between expression ISCs and total age and total empathy are described in Table 8. For all expressions age was negatively correlated with ISCs. Bayes factors indicated extreme evidence for a negative relationship between age and neutral, happy and angry ISCs, and anecdotal evidence for a negative relationship with fear and disgust ISCs. Whereas there was

**Table 5. Bayes factors for correlations between facial expression ISCs and engagement for performances P1–P3**

Expression	P1		P2		P3	
	$r^2$	$BF_{10}$	$r^2$	$BF_{10}$	$r^2$	P3 $BF_{10}$
Neutral	0.35	$1.05 \times 10^{25}$	0.22	$2.39 \times 10^7$	0.40	$3.84 \times 10^{30}$
Happy	0.36	$7.70 \times 10^{26}$	0.29	$3.10 \times 10^{14}$	0.30	$2.88 \times 10^{16}$
Sad	0.07	0.41	0.26	$5.06 \times 10^{11}$	0.10	4.51
Angry	0.25	$1.85 \times 10^{11}$	0.36	$5.91 \times 10^{24}$	0.11	8.99
Surprise	–0.14	466.62	–0.05	0.14	–0.05	0.12
Fear	0.07	0.36	0.23	$3.51 \times 10^8$	–0.04	0.07
Disgust	0.15	1667.67	0.28	$1.24 \times 10^{13}$	0.24	$4.37 \times 10^9$
Contempt	–0.06	0.24	0.10	2.23	0.13	41.49

**Table 6. Optimal Bayesian regression models for expression ISCs predicting audience engagement**

Performance	Optimal model	Optimal model $BF_M$	Optimal model $R^2$
P1	Neutral, Happy, Anger, Surprise, Fear, Disgust	140.82	0.24
P2	Anger, Fear, Disgust	57.64	0.19
P3	Neutral, Surprise, Fear, Disgust, Contempt	129.05	0.22

anecdotal evidence for no relationship between age and both sad and contempt ISCs and strong evidence for no relationship with surprise ISCs.

In contrast to age, audience empathy appeared positively correlated with most expressions ISCs, apart from sad and fear ISCs. Bayes factors indicated extreme evidence for a positive relationship between empathy and neutral and happy ISCs and very strong evidence for a positive relationship with angry ISCs. However, there was moderate evidence for no relationship between empathy and disgust ISCs and very strong evidence for no relationship with sad, surprise, fear, and contempt ISCs.

We then calculated eight separate Bayesian linear regression models to examine the predictive value of the age and empathy scores on the magnitude of each of the expression ISCs. Following observation of the data, the odds in favor of models where both audience age and empathy predicted expression ISCs increased for neutral, happy, and angry expressions (see Table 9). The inclusion Bayes factors indicated that for these expressions age is an extremely strong predictor of decreasing ISCs, but empathy also has a slightly smaller, but very strong predictive role of increasing ISCs. In addition, following observation of the data, the odds in favor of only age predicting ISCs increased for fear and disgust expressions. The inclusion Bayes factors indicated that for these expressions there was anecdotal evidence of age predicting decreasing ISCs, whereas strong or anecdotal evidence that empathy did not predict fear or disgust ISCs, respectively. Finally, following observation of the data, the odds in favor of neither age nor empathy predicting sad, surprise, and contempt expressions (the null model) increased.

To explore the impact of gender on expression ISCs, for each audience pair we calculated whether the pair were of the same gender where both were male, or both were female, or whether the audience pair were of different genders. Given that some prior research has indicated that gender differences in ISCs can be dependent upon age,<sup>58</sup> we also calculated whether the audience pairs were young (both <25 years old) or old (both >25 years old) or whether the audience pair were of different ages. We then analyzed how the magnitudes of different expression ISCs were dependent upon gender and age with eight separate Bayesian ANOVA models. We compared both a model with gender as a main effect and a model including both the main effect of gender and the interaction between gender and age, against a null model that also included the main effect of age (as the effect of age had already been explored earlier).

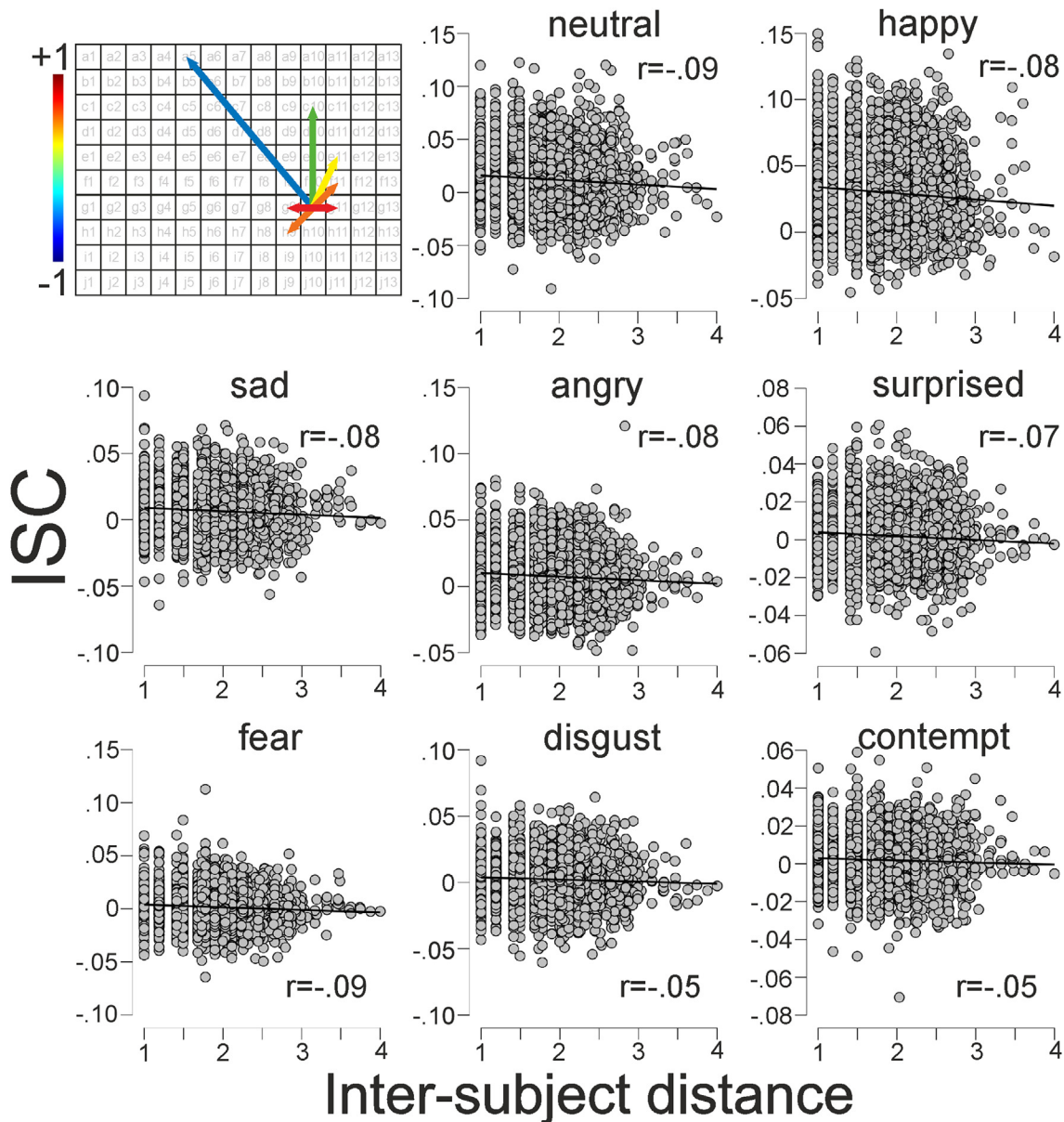
Following observation of the data, there was strong or greater evidence that ISCs were different across genders for neutral ( $BF_M = 63.00$ ), happy ( $BF_M = 455.01$ ), and angry ( $BF_M = 104.20$ ) expressions (see Figure 4). For these expressions, ISCs between females were greater than ISCs between males. In contrast, for fear ( $BF_M = 0.03$ ) and contempt ( $BF_M = 0.03$ ) expressions there was strong evidence that ISCs were similar across genders. For the other expressions, there was only moderate or anecdotal evidence that gender mattered for sad expressions ( $BF_M = 8.46$ ) or did not matter for surprise ( $BF_M = 0.12$ ) or disgust ( $BF_M = 0.13$ ) expressions.

The model that included the interaction between gender and age was compared against all other models, allowing us to examine whether gender differences in ISCs were dependent upon age. Following observation of the data, there was strong or greater evidence that there was no interaction between gender and age for neutral ( $BF_M = 0.01$ ), happy ( $BF_M = 0.01$ ), sad ( $BF_M = 0.02$ ), angry ( $BF_M = 0.04$ ), surprise ( $BF_M = 4.45 \times 10^{-4}$ ), fear ( $BF_M = 9.40 \times 10^{-5}$ ), disgust ( $BF_M = 3.65 \times 10^{-4}$ ), and contempt ( $BF_M = 1.34 \times 10^{-4}$ ) expressions. In summary, there was no evidence that gender differences in facial expressions ISCs were different between individuals of different ages.

## DISCUSSION

Across a range of different theatrical and cabaret performances we filmed the expressions conveyed by large numbers of audience members as the performances unfolded. These measurements allowed us to assess the intensity of neutral, happy, sad, angry, surprise, fear, disgust, and contempt expressions every 20 min in each face throughout the duration of the performances. We evaluated the degree of expression synchrony by measuring the correlations between expression intensity functions. We found that the facial expressions of audience members showed varying degrees of synchrony between the different expressions and varying synchrony over time. The greatest degree of audience synchrony was seen within-expression particularly for neutral, happy, sad, angry, disgust, and contempt expressions; there was little between-expression synchrony, indeed these could often be asynchronous.

Intersubject correlations (ISCs) provided different information than measures of facial expression magnitude<sup>7,11</sup> and avoid some issues relating to their analysis. Facial expression magnitude can tell us about how much each audience member might experience particular emotions during performances, as facial expressions are correlated with subjective emotional experiences.<sup>62</sup> This type of measure could also in principle provide information on the degree of emotions experienced during discrete periods of performances. However, measuring expression magnitudes is dependent upon obtaining a baseline assessment of the degree of emotions expressed by each individual face when not experiencing the performance. Expressions would then need to be compared to this baseline, as faces can vary in terms of structure that resemble different emotional expressions.<sup>63,64</sup> This approach would require all audience members to present an unexpressive or "neutral" face to the camera at a defined moment, not always possible when testing audiences "in the wild." Furthermore, there can never be a



**Figure 3. ISCs are related to audience proximity**

Upper left panel illustrates a grid representing the arrangement of seats in a hypothetical auditorium. For a given audience member (e.g., sitting in row g seat 10), the distance between them and each of the other audience members was calculated and then correlated with the degree of synchrony between each pair of individuals (other panels). The color of the arrows representing these relationships illustrate a hypothetical situation where audience proximity is related to emotional synchrony (redder colors = higher correlations, bluer colors = little to no correlations). Other panels: correlations (including  $r$  values) between transformed audience ISCs and the square root of the spatial proximity in the auditoriums.

guarantee that during any such baseline period audience members would not be experiencing any emotions that might impact their expressions. In contrast, expression ISC measures are not affected by this problem.

We tested how expression ISCs were related to predictions, made by the performance dramaturge, of the level of audience engagement during specific scenes of the performances. Performances were complex, involving multiple actors, scene changes, music, singing, and speech, and dealt with topics involving considerable emotional range. Predictions of how intensely these different elements of the performances might engage the audience consequently varied over the duration of the performances. However, predictions regarding periods of high engagement were not linked to a specific emotion or type of experience and could occur, for example, from periods of humor, jollity, or tragedy. Instead, the dramaturge's scores were holistic assessments of the degree to which they might command the attention of audience

**Table 7. Bayesian regression models testing if proximity predicts ISCs**

Emotion	BF <sub>M</sub>	Model R <sup>2</sup>	Mean	SD	95% credible interval
Neutral	18,985	0.01	−0.004	<0.001	[−0.006, −0.003]
Happy	3,988	0.01	−0.005	0.001	[−0.007, −0.003]
Sad	1,582	0.01	−0.003	<0.001	[−0.005, −0.001]
Angry	862	0.01	−0.003	<0.001	[−0.003, −0.002]
Surprise	479	0.01	−0.002	<0.001	[−0.003, −0.001]
Fear	11,431	0.01	−0.002	<0.001	[−0.003, −0.001]
Disgust	4.73	0.00	−0.001	0.001	[−0.002, −0.000]
Contempt	2.08	0.00	−0.000	<0.001	[−0.002, −0.000]

members. In line with our hypothesis, we found that ISCs were positively correlated with predictions of audience engagement made by the performance dramaturge, in particular neutral, happy, anger, and disgust expression ISCs. The maximal correlation between expression ISCs and engagement in our study was seen with neutral expressions, reaching  $r = 0.40$  for one performance, and average correlations between engagement and both neutral and happy ISCs were  $r = 0.32$ . Bayesian linear regression analyses showed that expression ISCs accounted for up to 24% of the variance in predicted audience engagement.

This relationship between facial expressions ISCs and predicted audience engagement is somewhat lower than a potentially comparable study examining neural synchrony and post-class ratings of engagement by students.<sup>21</sup> Here, Dikker et al. measured EEG signals from 12 students during classes and calculated student-to-group synchrony while students rated the classes after the end of each class and at the end of the semester. They found that student-to-group synchrony was substantially positively correlated ( $r = 0.61$ ) to the post-semester ratings. In a different study, Dmochowski et al.<sup>22</sup> measured the relationship between neural synchrony and audience engagement and preference for television adverts. Neural synchrony was also measured with EEG from 16 individuals and was correlated with ( $r = 0.40$ ) and explained only 16% of the variance in the engagement of a much larger and different group of individuals as measured by audience tweet frequency.

However, there are substantial differences between the methods we employ here compared with these other studies that may explain some of the differences in the magnitude of the relationship between synchrony and engagement. First, the signals we measure (facial expressions) are different from the neural signals recorded by Dikker et al.<sup>21</sup> and Dmochowski et al.<sup>22</sup> Second, the cognitive constructs underlying audience synchrony in our study are likely to be somewhat different. Dmochowski et al.<sup>22</sup> localized sources of neural synchrony during advert viewing to large areas of cortex in both hemispheres stretching from sensory association areas in occipital cortex, along the superior temporal gyrus to the temporal poles, and within areas of parietal cortex. These areas include some regions involved in the experience of emotions.<sup>65,66</sup> As such, our method captures in particular the affective responses to artistic performances, which are often a predominant feature of audience engagement,<sup>3,12,13,67</sup> whereas these other techniques may capture additional cognitive interactions. Finally, our method shows that there appear to be independent synchrony profiles for the different emotional expressions, and these will be due to synchrony during different psychological processes. We saw greater synchrony of happy expressions in comparison with the other expressions measured. Although which expression was most synchronized varied through the performances, and this may reflect that different emotional processes synchronize differently during different cognitive appraisal of the performance.<sup>42</sup>

Any comparison with previous research measuring the relationship between the synchrony of different autonomic signals and audience engagement<sup>17,18</sup> is somewhat harder due to the complex relationship between the production of expressions and the autonomic nervous system.<sup>29,68</sup> However, most of these previously recorded signals (heart rate, respiration rate, skin conductance, etc.) are broadly dependent upon arousal level. Thus, there is perhaps some overlap between these measures and happy and anger expressions that tend to reflect high

**Table 8. Correlations between facial expression ISCs and audience age and empathy**

Emotion	Age		Empathy	
	r <sup>2</sup>	BF <sub>10</sub>	r <sup>2</sup>	BF <sub>10</sub>
Neutral	−0.12	1.02 × 10 <sup>9</sup>	0.08	605.64
Happy	−0.19	1.18 × 10 <sup>27</sup>	0.08	1012.55
Sad	−0.04	0.40	−0.01	0.03
Angry	−0.13	9.51 × 10 <sup>10</sup>	0.07	52.18
Surprise	−0.01	0.03	0.01	0.03
Fear	−0.05	1.18	−0.00	0.02
Disgust	−0.05	1.32	0.03	0.10
Contempt	−0.04	0.20	0.01	0.02

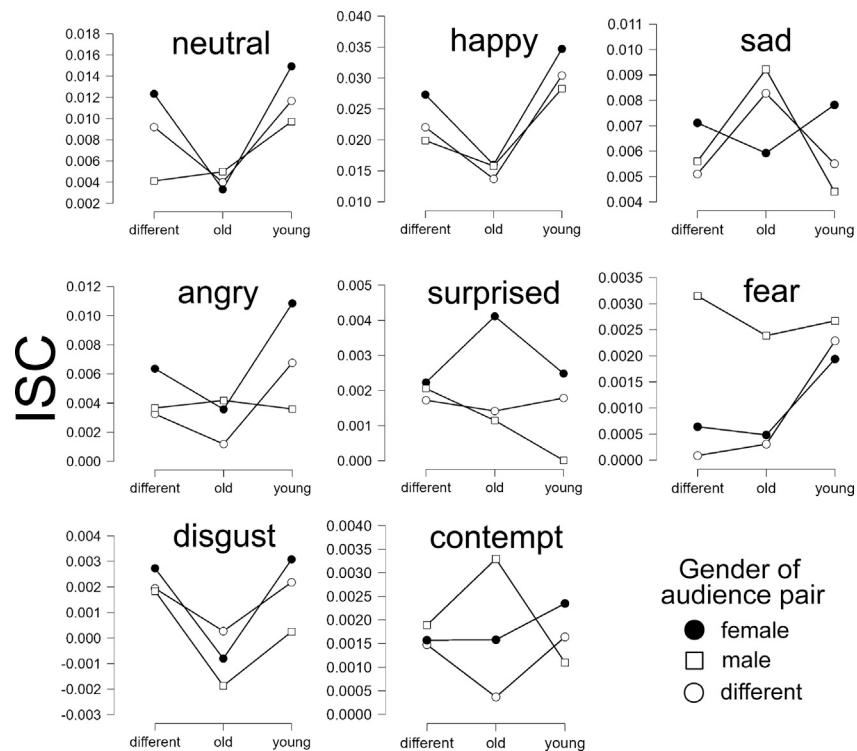
**Table 9. Bayesian regression models testing if age or empathy predicts ISCs**

Emotion	Optimal model	Optimal BF <sub>M</sub>
Neutral	Age & Empathy	2,275
Happy	Age & Empathy	2,565
Sad	Null model	3.70
Angry	Age & Empathy	203
Surprise	Null model	28.69
Fear	Age	5.37
Disgust	Age	4.10
Contempt	Null model	7.00

levels of arousal.<sup>69</sup> Indeed, measures of heart rate and respiration rate synchrony appear related to concurrent measures of zygomatic muscle synchrony in audiences.<sup>18</sup>

Furthermore, the measures of audience engagement in our study and prior studies<sup>18,21,22</sup> are quite different. In contrast to these other studies, we used a prediction of audience engagement rather than a direct measure from the audience themselves. These predictions were based upon the dramaturge’s knowledge of the theatrical performances, including the historical context, purpose, and meaning behind the performances. As such, they represent a holistic estimate of many of the elements of the multifaceted concept of engagement<sup>9</sup> including emotional responses, captivation, attention, and intellectual stimulation. Although these were predictions, and not direct measurements of engagement from the audiences themselves, what our method allowed was an estimate of audience engagement at different time points throughout the duration of the performances. This provides us with information to understand the relationship between the moment-to-moment fluctuations in audience expression synchrony, with our predictions of engagement at the same time points.

However, given that the measurement of audience engagement has proved difficult,<sup>9</sup> measuring facial expression synchrony may provide a valuable method to non-invasively determine audience engagement during artistic performances as well as other group activities in the



**Figure 4. The influence of gender on ISCs**

Separate panels indicate the mean transformed ISCs for the eight expressions. Mean ISCs are calculated between pairs of individuals grouped according to gender: female (both female), male (both male), different (male and female), and according to age: young (both <25 years old), old (both >25 years old), different (one individual <25 and one individual >25).

future. This technique has considerable benefits over other methods of measuring group synchrony. First, unlike other forms of measuring physiological (e.g., heart rate, respiration rate, skin conductance etc.) or neural (e.g., fMRI, EEG) activity, filming faces has little impact on the audience experience. There is no need to attach electrodes to the participants or interfere with the way they might experience the performance. Although our cameras were necessarily in view of the audience, they were often surreptitiously placed, and were likely quickly forgotten, so the audience experienced the performances in a naturalistic fashion and as intended by the directors. Currently, this is not achievable with any other technique.

Although not determined within this study, the synchrony of the different facial expressions within an audience may provide information about distinct aspects of audience engagement in the future (e.g., recall, change in attitudes, future attendance, etc.). The role of the neutral expression is currently uncertain but appears to be strongly linked to predictions of audience engagement and so clearly has potential utility for evaluating audiences non-invasively in the future. FaceReader classifies the intensity of the neutral and the other seven emotional expressions independently from each other. As such they provide different signals related to the emotional state of the audience members. The neutral expression does not represent the mere absence of all other emotional expressions, as it is only moderately negatively correlated with the other expressions within each individual. These findings might reflect that people can experience more than one emotion at the same time. This may be particularly the case when attending complex and varied artistic performances like ours where comic and tragic narratives were entwined. In contrast, more simple emotional experiences may result in more separable expressions.

Our methods allow the measurement of the simultaneous experience of the audiences as a group. This is not possible with fMRI, where ISCs are typically measured asynchronously.<sup>25,61</sup> With other physiological signals (e.g., heart rate, skin conductance, etc.) simultaneous recording across the group is possible with multiple measurement systems but will be restricted by the availability of the equipment and to date has been restricted to low numbers of individuals.<sup>10,21</sup> Furthermore, these methods are time-consuming to set up and invasive. In this study, we analyzed the simultaneous facial expressions of up to ~50 individuals, which along with providing a perhaps more representative sample, enabled us to examine the dynamics of inter-individual synchrony. In principle, it would be possible to examine facial expression synchrony with even great numbers of individuals than those we measured here.

In alignment with our predictions, we found very strong evidence of a small effect of proximity on ISCs, where individuals closer to each other within the auditorium were more likely to show synchrony of neutral, happy, sad, angry, surprise, and fear expressions. One reason for the greater expression synchrony between individuals near to each other may be due to emotional contagion, as individuals in proximity have a tendency to "catch" other persons' emotions, including their emotional appraisals, subjective feelings, patterned physiological response, and tendencies.<sup>39,41,46</sup> Alternatively, this effect may reflect enhanced synchrony between friends who know each other sitting together in the auditorium. Typically individuals tend to forge friendships with others who share similar age, gender, and other demographic characteristics.<sup>70</sup> Such similarities in personal characteristics can result in increased synchrony between individuals who are friends compared to individuals outside of their social network.<sup>71</sup> This enhanced synchrony between nearby individuals may reflect either similar responses to the performance or enhanced emotional contagion due to proximity. Ultimately, these questions could be directly tested by comparing expression synchrony of individuals when alone to when they are sitting together as a group, either with strangers or with neighbors they know intimately.

The analysis of ISCs in large groups of individuals additionally allowed us to determine the impact of audience member age, empathy, and gender on ISCs. In line with our hypothesis, several (neutral, happy and angry) expressions were negatively correlated with age, whereas there was little evidence for a relationship for the other expressions. These results parallel those of others that have shown similar effects when measuring neural synchrony.<sup>56-59</sup> Our results suggest that older audience members' emotional experiences during the performances were more idiosyncratic, possibly resulting from this group's poorer ability to recognize the emotions expressed by the performers.<sup>72-74</sup>

Empathy had a less substantial impact on ISCs, but as per our predictions, was positively correlated with neutral, happy, and angry expression ISCs.<sup>21</sup> These results suggest that individuals showing higher empathy may experience these emotions alike during theatrical performances. However, our measure of audience member empathy was limited to completing a self-report questionnaire, for reasons of practicality. This method may not be optimal as individuals may not have the metacognitive insight to be able to accurately evaluate their own ability in this area.<sup>75,76</sup> Therefore, it will take future testing with behaviorally verified measures of empathy to accurately measure the relationship between expression ISCs and participant empathy.

Due to the paucity of research on the role of gender in interindividual synchrony, we did not have strong predictions about the impact of gender on expression ISCs. Female audience members showed greater neutral, happy, and angry expression ISCs than males, indicating greater similarity in the emotions of female audience members. However, we had strong evidence that fear and contempt ISCs were the same across genders. These results stand in contrast to Petroni et al.,<sup>58</sup> who saw greater ISC in males than females, but only in children and not in adults. We also saw no real evidence of an interaction between gender and age and so consider that these differences in our findings and Petroni et al.'s are likely due to the substantive differences in measures of facial emotional expressions and EEG signals.

### Limitations of the study

Measuring facial expressions in naturalistic environments is heavily dependent upon the lighting conditions that could be achieved and the constraints imposed by the physical space and the artistic demands of the performance directors. In earlier testing (e.g., in Madison in 2015 and 2016), our ability to record faces with video cameras at a high-enough quality to allow for face detection and expression extraction by our software was significantly compromised, and consequently we obtained very few faces on which to base our analyses. In contrast, when we have control over the positioning of the cameras and audience and the level of lighting (e.g., in the three performances in the Holbeck cinema, York 2017), we could ensure that faces were clearly visible to the cameras and were well illuminated. Here, we were

able to obtain very good measures of facial expressions from the vast majority of the audience and through the entirety of the performances. As such, the use of this technique is limited to those situations where a predominantly full-frontal view of a well-lit face is possible.

Audience engagement was only positively correlated with neutral, happy, anger, and disgust expressions. This may perhaps be a product of the ability of FaceReader to classify these emotions well from our particular videos where faces were not filmed in optimal lighting conditions and videos were “up-sampled” by the digital zooming process. For example, classification of happy expressions relies heavily on detecting contraction of the zygomaticus muscle, which might be reliably detected with lower resolution information. In contrast, for example, classification of fear expressions relies on the detection of a combination of more subtle action units (e.g., inner brow raiser, upper lid raiser, lips part etc.) that might be harder to detect under the same conditions. Furthermore, FaceReader can confuse fear and surprise expressions.<sup>77</sup> Consequently, the lack of relationship between engagement and some expressions may reflect the nature of filming multiple faces with each camera and the limitations of the version of FaceReader we used during our analysis. We will ultimately be able to test this possibility in the future with the use of many more cameras than we used here and as face expression classification technology improves.

In conclusion, we show that the synchrony of the facial expressions of audience members is related to predictions of audience engagement, as well as some of the characteristics of the audience members themselves. This technique could be used to measure the engagement of individuals at artistic performances and during other group activities. In comparison to previous measures of inter-individual neural and physiological synchrony, it is non-invasive and does not detract from the audience experience itself.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- [KEY RESOURCES TABLE](#)
- [RESOURCE AVAILABILITY](#)
  - Lead contact
  - Materials availability
  - Data and code availability
- [EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS](#)
  - Participants
- [METHOD DETAILS](#)
  - Theatrical and cabaret performances
  - Filming audience facial expressions
  - Procedure
- [QUANTIFICATION AND STATISTICAL ANALYSIS](#)
  - Analysis of facial expressions
  - Intersubject correlation (ISC) analyses
  - Audience engagement predictions
  - Statistical analyses

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2024.109843>.

## ACKNOWLEDGMENTS

This research was supported by a grant from the Arts and Humanities Research Council (AH/M004457/1).

## AUTHOR CONTRIBUTIONS

Conceptualization, N.E.B. and L.P.; methodology, N.E.B.; investigation, R.A.O., N.E.B., and L.P.; writing—original draft, N.E.B.; writing—review & editing, R.A.O., N.E.B., and L.P.; funding acquisition, N.E.B. and L.P.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 7, 2023

Revised: January 10, 2024

Accepted: April 25, 2024

Published: April 29, 2024

## REFERENCES

1. Manstead, T. (2005). The social dimension of emotion. *Psychol.* 18, 484–487.
2. North, A.C., Hargreaves, D.J., and Hargreaves, J.J. (2004). Uses of music in everyday life. *Music Percept.* 22, 41–77. <https://doi.org/10.1525/mp.2004.22.1.41>.
3. Lamont, A. (2011). University students' strong experiences of music: Pleasure, engagement, and meaning. *Music. Sci.* 15, 229–249. <https://doi.org/10.1177/1029864911403368>.
4. Ekman, P., and Friesen, W.V. (1971). Constants across cultures in the face and emotion. *J. Pers. Soc. Psychol.* 17, 124–129. <https://doi.org/10.1037/h0030377>.
5. Rosenberg, E.L., and Ekman, P. (1997). *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)* (Oxford University Press).
6. Ekman, P., Friesen, W.V., and Ancoli, S. (1980). Facial signs of emotional experience. *J. Pers. Soc. Psychol.* 39, 1125–1134. <https://doi.org/10.1037/h0077722>.
7. Kayser, D., Egermann, H., and Barraclough, N.E. (2022). Audience facial expressions detected by automated face analysis software reflect emotions in music. *Behav. Res. Methods* 54, 1493–1507. <https://doi.org/10.3758/s13428-021-01678-3>.
8. Zentner, M., Grandjean, D., and Scherer, K.R. (2008). Emotions evoked by the sound of music: characterization, classification, and measurement. *Emotion* 8, 494–521. <https://doi.org/10.1037/1528-3542.8.4.494>.
9. Brown, A.S., and Novak, J.L. (2008). Assessing the Intrinsic Impacts of a Live Performance.
10. Egermann, H., Sutherland, M.E., Grewe, O., Nagel, F., Kopiez, R., and Altenmüller, E. (2011). Does music listening in a social context alter experience? A physiological and psychological perspective on emotion. *Music. Sci.* 15, 307–323. <https://doi.org/10.1177/1029864911399497>.
11. Egermann, H., Pearce, M.T., Wiggins, G.A., and McAdams, S. (2013). Probabilistic models of expectation violation predict psychophysiological emotional responses to live concert music. *Cogn. Affect. Behav. Neurosci.* 13, 533–553. <https://doi.org/10.3758/s13415-013-0161-y>.
12. Brown, A.S., and Novak-Leonard, J.L. (2013). Measuring the intrinsic impacts of arts attendance. *Cult. Trends* 22, 223–233. <https://doi.org/10.1080/09548963.2013.817654>.
13. Radbourne, J., Glow, H., and Johanson, K. (2010). Measuring the intrinsic benefits of arts attendance. *Cult. Trends* 19, 307–324. <https://doi.org/10.1080/09548963.2010.515005>.
14. McConachie, B. (2008). *Engaging Audiences: A Cognitive Approach to Spectating in the Theatre* (Springer).
15. Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience* (Harper & Row).
16. Konvalinka, I., Xygalatas, D., Bulbulia, J., Schjødt, U., Jegindø, E.-M., Wallot, S., Van Orden, G., and Roepstorff, A. (2011). Synchronized arousal between performers and related spectators in a fire-walking ritual. *Proc. Natl. Acad. Sci. USA* 108, 8514–8519. <https://doi.org/10.1073/pnas.1016955108>.
17. Marci, C.D. (2006). A Biologically Based Measure of Emotional Engagement: Context Matters. *J. Advert. Res.* 46, 381–387. <https://doi.org/10.2501/S002184990600466>.
18. Czepiel, A., Fink, L.K., Fink, L.T., Wald-Fuhrmann, M., Tröndle, M., and Merrill, J. (2021). Synchrony in the periphery: inter-subject correlation of physiological responses during live music concerts. *Sci. Rep.* 11, 22457–22516. <https://doi.org/10.1038/s41598-021-00492-3>.
19. Hammond, H., Armstrong, M., Thomas, G.A., and Gilchrist, I.D. (2023). Audience immersion: validating attentional and physiological measures against self-report. *Cogn. Res. Princ. Implic.* 8, 22. <https://doi.org/10.1186/s41235-023-00475-0>.
20. Tschacher, W., Greenwood, S., Egermann, H., Wald-Fuhrmann, M., Czepiel, A., Tröndle, M., and Meier, D. (2023). Physiological synchrony in audiences of live concerts. *Psychol. Aesthet. Creativ. Arts* 17, 152–162. <https://doi.org/10.1037/aca0000431>.
21. Dikker, S., Wan, L., Davidesco, I., Kaggen, L., Oostrik, M., McClintock, J., Rowland, J., Michalareas, G., Van Bavel, J.J., Ding, M., and Poeppel, D. (2017). Brain-to-brain synchrony tracks real-world dynamic group interactions in the classroom. *Curr. Biol.* 27, 1375–1380. <https://doi.org/10.1016/j.cub.2017.04.002>.
22. Dmochowski, J.P., Bezdek, M.A., Abelson, B.P., Johnson, J.S., Schumacher, E.H., and Parra, L.C. (2014). Audience preferences are predicted by temporal reliability of neural processing. *Nat. Commun.* 5, 4567–4569. <https://doi.org/10.1038/ncomms5567>.
23. Poulsen, A.T., Kamronn, S., Dmochowski, J., Parra, L.C., and Hansen, L.K. (2016). Measuring Engagement in a Classroom: Synchronised Neural Recordings during a Video Presentation. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1604.03019>.
24. Hasson, U., Ghazanfar, A.A., Galantucci, B., Garrod, S., and Keysers, C. (2012). Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends Cogn. Sci.* 16, 114–121. <https://doi.org/10.1016/j.tics.2011.12.007>.
25. Hasson, U., Malach, R., and Heeger, D.J. (2010). Reliability of cortical activity during natural stimulation. *Trends Cogn. Sci.* 14, 40–48. <https://doi.org/10.1016/j.tics.2009.10.011>.
26. Nastase, S.A., Gazzola, V., Hasson, U., and Keysers, C. (2019). Measuring shared responses across subjects using intersubject correlation. *Soc. Cogn. Affect. Neurosci.* 14, 667–685. <https://doi.org/10.1093/scan/nsz037>.
27. Nummenmaa, L., Glerean, E., Viinikainen, M., Jääskeläinen, I.P., Hari, R., and Sams, M. (2012). Emotions promote social interaction by synchronizing brain activity across individuals. *Proc. Natl. Acad. Sci. USA* 109, 9599–9604. <https://doi.org/10.1073/pnas.1206095109>.
28. Hasson, U., Avidan, G., Gelbard, H., Vallines, I., Harel, M., Minshew, N., and Behrmann, M. (2009). Shared and idiosyncratic cortical activation patterns in autism revealed under continuous real-life viewing conditions. *Autism Res.* 2, 220–231. <https://doi.org/10.1002/aur.89>.
29. Madsen, J., and Parra, L.C. (2022). Cognitive processing of a common stimulus synchronizes brains, hearts, and eyes. *PNAS nexus* 1, pgac020. <https://doi.org/10.1093/pnasnexus/pgac020>.
30. Dmochowski, J.P., Sajda, P., Dias, J., and Parra, L.C. (2012). Correlated components of ongoing EEG point to emotionally laden attention—a possible marker of engagement? *Front. Hum. Neurosci.* 6, 112. <https://doi.org/10.3389/fnhum.2012.00112>.
31. Hasson, U., and Frith, C.D. (2016). Mirroring and beyond: coupled dynamics as a generalized framework for modelling social interactions. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 371, 20150366. <https://doi.org/10.1098/rstb.2015.0366>.
32. Behrens, F., Snijdwint, J.A., Moulder, R.G., Prochazkova, E., Sjak-Shie, E.E., Boker, S.M., and Kret, M.E. (2020). Physiological synchrony is associated with cooperative success in real-life interactions. *Sci. Rep.* 10, 19609. <https://doi.org/10.1038/s41598-020-76539-8>.
33. Mitkidis, P., McGraw, J.J., Roepstorff, A., and Wallot, S. (2015). Building trust: Heart rate synchrony and arousal during joint action increased by public goods game. *Physiol. Behav.* 149, 101–106. <https://doi.org/10.1016/j.physbeh.2015.05.033>.
34. Ferrer, E., and Helm, J.L. (2013). Dynamical systems modeling of physiological coregulation in dyadic interactions. *Int. J. Psychophysiol.* 88, 296–308. <https://doi.org/10.1016/j.ijpsycho.2012.10.013>.
35. Wood, A., Lipson, J., Zhao, O., and Niedenthal, P. (2021). Forms and functions of affective synchrony. In *Handbook of embodied psychology: Thinking, feeling, and acting*, pp. 381–402. [https://doi.org/10.1007/978-3-030-78471-3\\_17](https://doi.org/10.1007/978-3-030-78471-3_17).
36. Riehle, M., Kempkensteffen, J., and Lincoln, T.M. (2017). Quantifying facial expression synchrony in face-to-face dyadic interactions: Temporal dynamics of simultaneously recorded facial EMG signals. *J. Nonverbal Behav.* 41, 85–102. <https://doi.org/10.1007/s10919-016-0246-8>.
37. Hess, U., and Bourgeois, P. (2010). You smile—I smile: Emotion expression in social interaction. *Biol. Psychol.* 84, 514–520. <https://doi.org/10.1016/j.biopsycho.2009.11.001>.
38. Palumbo, R.V., Marraccini, M.E., Weyandt, L.L., Wilder-Smith, O., McGee, H.A., Liu, S., and Goodwin, M.S. (2017). Interpersonal autonomic physiology: A systematic review of the literature. *Pers. Soc. Psychol. Rev.* 21, 99–141. <https://doi.org/10.1177/1088868316628405>.
39. Golland, Y., Arzouan, Y., and Levit-Binnun, N. (2015). The mere co-presence: Synchronization of autonomic signals and emotional responses across co-present individuals not engaged in direct interaction. *PLoS One* 10, e0125804. <https://doi.org/10.1371/journal.pone.0125804>.
40. Cheong, J.H., Molani, Z., Sadhukha, S., and Chang, L.J. (2023). Synchronized affect in shared experiences strengthens social connection. *Commun. Biol.* 6, 1099. <https://doi.org/10.1038/s42003-023-05461-2>.
41. Golland, Y., Mevorach, D., and Levit-Binnun, N. (2019). Affiliative zygomatic synchrony in co-present strangers. *Sci. Rep.* 9, 3120. <https://doi.org/10.1038/s41598-019-40060-4>.
42. Bruder, M., Dosmukhambetova, D., Nerb, J., and Manstead, A.S.R. (2012). Emotional signals in nonverbal interaction: Dyadic facilitation and convergence in expressions, appraisals, and feelings. *Cogn. Emot.* 26, 480–502. <https://doi.org/10.1080/02699931.2011.645280>.
43. Kreibitz, S.D. (2010). Autonomic nervous system activity in emotion: A review. *Biol.*

- Psychol. 84, 394–421. <https://doi.org/10.1016/j.biopsycho.2010.03.010>.
44. Levenson, R.W. (2003). Blood, sweat, and fears: The autonomic architecture of emotion. *Ann. N. Y. Acad. Sci.* 1000, 348–366. <https://doi.org/10.1196/annals.1280.016>.
  45. Chartrand, T.L., and Bargh, J.A. (1999). The chameleon effect: The perception–behavior link and social interaction. *J. Pers. Soc. Psychol.* 76, 893–910. <https://doi.org/10.1037/0022-3514.76.6.893>.
  46. Hatfield, E., Hatfield, C., Cacioppo, J.T., and Rapson, R.L. (1994). *Emotional Contagion* (Cambridge University Press).
  47. Hatfield, E., Cacioppo, J.T., and Rapson, R.L. (1993). Emotional contagion. *Curr. Dir. Psychol. Sci.* 2, 96–100. <https://doi.org/10.1111/1467-8721.ep10770953>.
  48. Dezechache, G., Conty, L., Chadwick, M., Philip, L., Soussignan, R., Sperber, D., and Grèzes, J. (2013). Evidence for unintentional emotional contagion beyond dyads. *PLoS One* 8, e67371. <https://doi.org/10.1371/journal.pone.0067371>.
  49. Messina, I., Palmieri, A., Sambin, M., Kleinbub, J.R., Voci, A., and Calvo, V. (2013). Somatic underpinnings of perceived empathy: The importance of psychotherapy training. *Psychother. Res.* 23, 169–177. <https://doi.org/10.1080/10503307.2012.748940>.
  50. Marci, C.D., and Orr, S.P. (2006). The effect of emotional distance on psychophysiological concordance and perceived empathy between patient and interviewer. *Appl. Psychophysiol. Biofeedback* 31, 115–128. <https://doi.org/10.1007/s10484-006-9008-4>.
  51. Marci, C.D., Ham, J., Moran, E., and Orr, S.P. (2007). Physiologic correlates of perceived therapist empathy and social-emotional process during psychotherapy. *J. Nerv. Ment. Dis.* 195, 103–111. <https://doi.org/10.1097/01.nmd.0000253731.71025.fc>.
  52. Robinson, J.W., Herman, A., and Kaplan, B.J. (1982). Autonomic responses correlate with counselor–client empathy. *J. Counsel. Psychol.* 29, 195–198. <https://doi.org/10.1037/0022-0167.29.2.195>.
  53. Chatel-Goldman, J., Congedo, M., Jutten, C., and Schwartz, J.-L. (2014). Touch increases autonomic coupling between romantic partners. *Front. Behav. Neurosci.* 8, 95. <https://doi.org/10.3389/fnbeh.2014.00095>.
  54. Pelphrey, K.A., Morris, J.P., and McCarthy, G. (2004). Grasping the intentions of others: The perceived intentionality of an action influences activity in the superior temporal sulcus during social perception. *J. Cogn. Neurosci.* 16, 1706–1716. <https://doi.org/10.1162/0899929042947900>.
  55. Preston, S.D., and de Waal, F.B.M. (2002). Empathy: Its ultimate and proximate bases. *Behav. Brain Sci.* 25, 1–71. <https://doi.org/10.1017/S0140525X02000018>.
  56. Chen, G., Taylor, P.A., Qu, X., Molfese, P.J., Bandettini, P.A., Cox, R.W., and Finn, E.S. (2020). Untangling the relatedness among correlations, part III: inter-subject correlation analysis through Bayesian multilevel modeling for naturalistic scanning. *Neuroimage* 216, 116474. <https://doi.org/10.1016/j.neuroimage.2019.116474>.
  57. Campbell, K.L., Shafto, M.A., Wright, P., Tsvetanov, K.A., Geerligns, L., Cusack, R., Cam-CAN, Tyler, L.K., Bullmore, E., and Calder, A. (2015). Idiosyncratic responding during movie-watching predicted by age differences in attentional control. *Neurobiol. Aging* 36, 3045–3055. <https://doi.org/10.1016/j.neurobiolaging.2015.07.028>.
  58. Petroni, A., Cohen, S., Langer, N., Henin, S., Vanderwal, T., Milham, M.P., and Parra, L.C. (2018). The Variability of Neural Responses to Naturalistic Videos Change with Age and Sex. *eNeuro* 5, ENEURO.0244-17.2017. <https://doi.org/10.1523/ENEURO.0244-17.2017>.
  59. McIntosh, A.R., Kovacevic, N., and Itier, R.J. (2008). Increased brain signal variability accompanies lower behavioral variability in development. *PLoS Comput. Biol.* 4, e1000106. <https://doi.org/10.1371/journal.pcbi.1000106>.
  60. Lenroot, R.K., Gogtay, N., Greenstein, D.K., Wells, E.M., Wallace, G.L., Clasen, L.S., Blumenthal, J.D., Lerch, J., Zijdenbos, A.P., Evans, A.C., et al. (2007). Sexual dimorphism of brain developmental trajectories during childhood and adolescence. *Neuroimage* 36, 1065–1073. <https://doi.org/10.1016/j.neuroimage.2007.03.053>.
  61. Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., and Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science* 303, 1634–1640. <https://doi.org/10.1126/science.1089506>.
  62. McIntosh, D.N. (1996). Facial feedback hypotheses: Evidence, implications, and directions. *Motiv. Emot.* 20, 121–147. <https://doi.org/10.1007/BF02253868>.
  63. Said, C.P., Sebe, N., and Todorov, A. (2009). Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion* 9, 260–264. <https://doi.org/10.1037/a0014681>.
  64. Hester, N. (2019). Perceived negative emotion in neutral faces: Gender-dependent effects on attractiveness and threat. *Emotion* 19, 1490–1494. <https://doi.org/10.1037/emo0000525>.
  65. Adolphs, R. (1999). Social cognition and the human brain. *Trends Cogn. Sci.* 3, 469–479. [https://doi.org/10.1016/S1364-6613\(99\)01399-6](https://doi.org/10.1016/S1364-6613(99)01399-6).
  66. Adolphs, R. (2001). The neurobiology of social cognition. *Curr. Opin. Neurobiol.* 11, 231–239. [https://doi.org/10.1016/S0959-4388\(00\)00202-6](https://doi.org/10.1016/S0959-4388(00)00202-6).
  67. Lamont, A., and Webb, R. (2010). Short-and long-term musical preferences: What makes a favourite piece of music? *Psychol. Music* 38, 222–241. <https://doi.org/10.1177/0305735609339471>.
  68. Cacioppo, J.T., Berntson, G.G., Larsen, J.T., Poehlmann, K.M., and Ito, T.A. (2000). The psychophysiology of emotion. In *Handbook of Emotions, 2nd Edition* (Handbook of Emotions, pp. 173–191).
  69. Russell, J.A. (1980). A circumplex model of affect. *J. Person. Social Psychol.* 39, 1161–1178. <https://doi.org/10.1037/h0077714>.
  70. McPherson, M., Smith-Lovin, L., and Cook, J.M. (2001). Birds of a feather: Homophily in social networks. *Annu. Rev. Sociol.* 27, 415–444. <https://doi.org/10.1146/annurev.soc.27.1.415>.
  71. Parkinson, C., Kleinbaum, A.M., and Wheatley, T. (2018). Similar neural responses predict friendship. *Nat. Commun.* 9, 332. <https://doi.org/10.1038/s41467-017-02722-7>.
  72. Demenescu, L.R., Mathiak, K.A., and Mathiak, K. (2014). Age- and gender-related variations of emotion recognition in pseudowords and faces. *Exp. Aging Res.* 40, 187–207. <https://doi.org/10.1080/0361073X.2014.882210>.
  73. Horning, S.M., Cornwell, R.E., and Davis, H.P. (2012). The recognition of facial expressions: An investigation of the influence of age and cognition. *Aging Neuropsychol. Cognit.* 19, 657–676. <https://doi.org/10.1080/13825585.2011.645011>.
  74. Olderbak, S., Wilhelm, O., Hildebrandt, A., and Quoidbach, J. (2019). Sex differences in facial emotion perception ability across the lifespan. *Cogn. Emot.* 33, 579–588. <https://doi.org/10.1080/02699931.2018.1454403>.
  75. Murphy, B.A., and Lilienfeld, S.O. (2019). Are self-report cognitive empathy ratings valid proxies for cognitive empathy ability? Negligible meta-analytic relations with behavioral task performance. *Psychol. Assess.* 31, 1062–1072. <https://doi.org/10.1037/pas0000732>.
  76. Borsboom, D., Mellenbergh, G.J., and Van Heerden, J. (2004). The concept of validity. *Psychol. Rev.* 111, 1061–1071. <https://doi.org/10.1037/0033-295X.111.4.1061>.
  77. Skiendziel, T., Rösch, A.G., and Schultheiss, O.C. (2019). Assessing the convergent validity between the automated emotion recognition software Noldus FaceReader 7 and Facial Action Coding System Scoring. *PLoS One* 14, e0223905. <https://doi.org/10.1371/journal.pone.0223905>.
  78. Abbasi, A.R., Dailey, M.N., Afzulpurkar, N.V., and Uno, T. (2013). Issues in automated facial expression analysis: comparing self, observer, and FaceReader interpretations. *BUJICT* 6, 23–31. [https://www.researchgate.net/publication/233994185\\_Issues\\_in\\_Automated\\_Facial\\_Expression\\_Analysis\\_Comparing\\_Self\\_Observer\\_and\\_FaceReader\\_Interpretations](https://www.researchgate.net/publication/233994185_Issues_in_Automated_Facial_Expression_Analysis_Comparing_Self_Observer_and_FaceReader_Interpretations).
  79. Wakabayashi, A., Baron-Cohen, S., Wheelwright, S., Goldenfeld, N., Delaney, J., Fine, D., Smith, R., and Weil, L. (2006). Development of short forms of the Empathy Quotient (EQ-Short) and the Systemizing Quotient (SQ-Short). *Pers. Individ. Differ.* 41, 929–940. <https://doi.org/10.1016/j.paid.2006.03.017>.
  80. Baron-Cohen, S., and Wheelwright, S. (2004). The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *J. Autism Dev. Disord.* 34, 163–175. <https://doi.org/10.1023/B:JADD.0000022607.19833.00>.
  81. Reniers, R.L.E.P., Corcoran, R., Drake, R., Shryane, N.M., and Völlm, B.A. (2011). The QCAE: a Questionnaire of Cognitive and Affective Empathy. *J. Pers. Assess.* 93, 84–95. <https://doi.org/10.1080/00223891.2010.528484>.
  82. Noldus (2016). *FaceReaderTM 7: Tool for Automatic Analysis of Facial Expressions* (Noldus Information Technology).
  83. Jeffreys, H. (1961). *The Theory of Probability*, 3rd Ed (Oxford University Press).
  84. JASP-Team (2021). *JASP (Version 0.16) (computer software)*.

## STAR★METHODS

## KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Raw data and statistical analyses	This study	<a href="https://osf.io/uzrhf/">https://osf.io/uzrhf/</a>
Software and algorithms		
JASP	<a href="https://jasp-stats.org/">https://jasp-stats.org/</a>	version 0.16.2
MATLAB	<a href="https://www.mathworks.com/">https://www.mathworks.com/</a>	version 2020a

## RESOURCE AVAILABILITY

## Lead contact

Further information and requests for resources will be fulfilled by the lead contact, Nick Barraclough ([nick.barraclough@york.ac.uk](mailto:nick.barraclough@york.ac.uk)).

## Materials availability

This study did not generate new unique reagents or other materials.

## Data and code availability

- Raw data, secondary data and JASP statistical analyses have been deposited at the OSF and are publicly available as of the date of publication. The DOI is reported in the [key resources table](#).
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

## EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

## Participants

Participants were audience members that attended one (or more) of 16 separate theatre and cabaret performances at different venues in the UK, USA and Australia, and consented to both have their face filmed during the performance and complete related questionnaires. A number of individuals attended these performances but did not consent to either filming or completing questionnaires. These additional individuals sat in seats in the auditorium out of the field of view of cameras. In total 455 participants were tested (130 male, 254 female, 71 other/prefer not to say; mean age = 39.9 years, SD = 20.8 years, range 15-86). Sample size calculations for the number of individuals required for analyses could not be carried out in advance as no suitable previous effect sizes were relevant. In addition, sample sizes for each performance could not be pre-specified as this was determined by the number of individuals prepared to fill in questionnaires and be filmed on the day of the performance. However, the samples used in our analyses are greater than any used previously that have measured neural or physiological ISCs. To ensure our measures of facial expressions were meaningful to the performance as a whole, rather than only a short segment of the performance, we only analyzed the expressions of those audience members whose facial expressions were possible to detect and model for a high degree (> 90%) of the time during the performance, and who had completed all questionnaires, and there was more than one individual that met these criteria in the audience (see [Figure S2](#)). These individuals attended one of 7 performances of "A comedy of us Jews" held in York, UK, 2017, or one of 3 performances of "Prince Bettliend" held in Sydney, Australia, 2017. In total, our sample was 193 audience members (135 females, 58 males; mean age = 29.75, S.D. = 17.24, range 18 - 80). The study was approved by the ethics committees of the Departments of Psychology and Theatre, Film and Television at the University of York and local ethics committees where required, and were performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki. Participants provided written informed consent for all testing.

## METHOD DETAILS

## Theatrical and cabaret performances

Performances were as follows: "Laugh with us" by Lisa Peschel, based on a scripts by Felix Porges, Vítězslav "Pidla" Horpatzky, Pavel Stránský et al., performed in Madison, Wisconsin, USA, August 2015; "Harlequin in the ghetto" developed by a student ensemble based at the University of Louisiana Baton Rouge based on a script by Zdeněk Jelínek, led by Lisa Peschel and Alan Sikes, performed in Madison, Wisconsin, USA, May 2016; two different versions of "Harlequin in the ghetto" performed in succession on 2 separate nights in York, UK, June 2016 (on any

one night audience members attended both performances in succession with an interval, however, audiences saw the performances in a different order on each day, and each performance is treated separately in analyses below); seven performances of “A comedy of us Jews” by Jac Weinstein, performed in 2 different venues in York, UK, June 2017; three performances of “Prince Bettliegend” developed by a professional cast at the University of Sydney, based on a script and lyrics by Josef Lustig and František Kovanic, led by Lisa Peschel and Ian Maxwell, performed in Sydney, New South Wales, Australia, August 2017 (during 2 of these performances each half of the audience were given differing introductory material for the purposes of research not reported here, each audience group in these two performances are treated as separate ‘performances’ in analyses below). All performances were part of a larger project, “Performing the Jewish Archive” where the intention was to perform previously neglected works by Jewish artists who had been displaced by various upheavals during the long twentieth century (ptja.leeds.ac.uk). The performances varied considerably and included monologues, dialogues and conversations, singing, comedy, pathos, and instrumental passages. Performances ranged from about 30 minutes to 1 hour in duration.

### Filming audience facial expressions

Participant faces were filmed (1920 x 1080 pixels at 50 frames per second) with up to 4 HD cameras (3 x Panasonic HDC-TM900, 1 x Panasonic HDC-SD90) at each performance. Each camera was allowed to vary its white balance and focus automatically during the performance to help optimize images of the faces whilst the stage lighting varied during performances. Typically, cameras were situated unobtrusively on tripods or fixed to gantries at a height behind the actors so that approximately 15-20 faces were present in each camera’s field of view. As face expression analysis software relies on movies made under good light conditions,<sup>78</sup> where possible, house lights were kept as bright as possible to help illuminate participant faces, although occasionally luminance was limited by the aesthetic concerns of the performance dramaturge. An additional camera, placed at the back of the auditorium, filmed the stage on which performances occurred.

### Procedure

Testing was carried out in a range of auditoriums (see below table); typically, these were public or University venues used regularly for theatrical performances, and included a stage area and tiered seating. Auditoriums allowed for up to 100 seats, although not all seats were always taken at each performance.

#### List of auditoriums where data were collected

Auditorium name	Address	Performance	Number of performances	Date
Promenade Hall, Overture Center for the Arts	University of Madison, Madison, WI, USA	Laugh with us	1	2015
Fredric March Play Circle, Memorial Union	University of Madison, Madison, WI, USA	Harlequin in the Ghetto	1	2016
BlackBox Theater	University of York, York, UK	Harlequin in the Ghetto	4	2016
Quaker Meeting House	Festival of Ideas, York, UK	A comedy of us Jews	4	2017
Holbeck Cinema	University of York, York, UK	A comedy of us Jews	3	2017
Seymour Center	University of Sydney, Sydney, NSW, Australia	Prince Bettliegend	3	2017

Participants were all paying members of the public (except for 3 performances of “A comedy of us Jews” performed in the Holbeck cinema in the Department of Theatre, Film and Television at the University of York UK in 2017 where all participants were paid £20 to attend). For paying members of the public, prior to purchase of tickets, participants were informed that the performance was for research purposes and that they could opt-in to this study by choosing to sit in specific seats on the day of the performance. When they arrived at the venue, information was provided to all audience members to indicate which seats were to be filmed and which were not so they could choose whether to take part in the study. Those audience members that wished to be participants in the study could choose to sit on those seats that were to be filmed, on which was also provided an envelope containing a pen, information sheet, consent form, and questionnaires. Each pack was labeled with a seat code, so that the participant self-reports could be linked to facial expressions in subsequent analyses.

Once all audience members had taken their seats, and prior to the start of the performance, a brief presentation explaining the study was given by one of the authors. Participants were asked to read the information sheet and sign the consent form. If they did not want to take part in the study they were informed that they could move to sit in an alternate seat at the venue. Before the performances, participants were given 10 minutes to complete a questionnaire to measure their degree of empathy (EQ-short: Madison 2015, 2016 & York 2016; QCAE: York 2017 & Sydney 2017) and an additional questionnaire to measure their current emotional state (data for an additional study not reported here). The EQ-short<sup>79</sup> is a short form of the original Empathy Quotient<sup>80</sup> developed to provide a fast self-report measure of empathy using only 22 questions. The EQ-short provides a measure of total empathy between 0 = no empathy and 44 = very high empathy. The QCAE is the Questionnaire of Cognitive and Affective Empathy<sup>81</sup> and is a 31-item measure of empathy, but additionally can differentiate between cognitive empathy (e.g. perspective taking) and affective empathy (including emotional contagion and emotional and peripheral responsivity). The QCAE also provides a measure of total empathy between 4 = no empathy and 124 = very high empathy.

At the end of this period the cameras were turned on and the performance started. After the performance, cameras were turned off, and participants completed a questionnaire to measure their emotional state during the performance (data for an additional study not reported here), and to provide simple demographic information (age, gender). On completion of the final forms, participants placed all forms inside the envelope, sealed it, and placed the envelope within a box as they left the venue; following which they were offered a debriefing sheet.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Analysis of facial expressions

To prepare the film for subsequent face expression analyses the film from each performance was edited (Adobe Premiere Pro CC; Adobe, San Jose, CA, USA). First, for each performance, films from all cameras were synchronized with each other by identifying the onset of a light marker flashed before the onset of the performance. We then generated separate movies of each participant face visible in each camera's field of view. Where a face appeared in more than one camera field of view, we kept the movie that provided the most data (see below). The durations of face movies from an individual performance were identical, and lasted from the start of the performance (onset of acting) until the end of the performance (onset of audience clapping). Light intensity for the individual movies of faces from the 2015 and 2016 performances were increased using Adobe Premiere in order to help maximize subsequent detection of facial expressions. Lighting quality during 2017 performances was good enough so that movie light intensities were not adjusted. In addition, we 'digitally zoomed' into each face to upscale the video to 1920 x 1080 pixels for subsequent analysis.

Facial expressions were determined using automated face analysis software FaceReader 7.<sup>82</sup> FaceReader works by fitting a model to the face present in the frame. However, it is not always possible to correctly fit this model, for example if the face is temporarily obscured by the participant's hand or other object, the participant looks away from the front of the auditorium, etc. During frames when this occurred, it was not possible to derive information about any of that participant's facial expressions. All faces were analysed using the software's 'general' face model option unless data was available to indicate that the participant was 60 years old or over, where the 'elderly' face model option was applied.

Once a face is detected, FaceReader then applies two methods to determine what expressions are present. The first method synthesizes an artificial model of the participant's face based upon approximately 500 key points on the surface of the face along with the facial texture of the area within these points. From a database of annotated images, the model calculates the main sources of variation in the face and then describes the deviations found in new images of the participant's face from an average version of their face as a vector. Expression classification is done by inputting the vector into an artificial neural network. This network was trained on over 10,000 annotated images to classify seven emotional expressions and the neutral expression.

The second method of classification uses a deep artificial neural network to recognize patterns in the face. FaceReader directly classifies the facial expressions from image pixels, without generating any models; this helps FaceReader classify faces even when partially hidden. Finally, FaceReader combines these two methods to provide a classification of the 8 different expressions on a frame-by-frame basis. This classification is a measure of the intensity of each of 8 different expressions where 0 = expression not detected and 1 = expression fully present (see [Figure S3](#) for an example output during a performance).

Faces of some individuals have a bias towards appearing to display certain emotions (e.g. anger) when in their resting (neutral) state. FaceReader 7 has the ability to be calibrated to adjust for this bias based on the participant's neutral facial expression provided in a separate video of the participant's face. For our earlier performances (2015, 2016) we used no calibration method. For our later performances (2017), we asked the audience before the performance to provide a neutral expression for a period of 15 seconds whilst looking directly at the cameras in front of them. From this neutral expression video FaceReader 7 was calibrated by taking the mid-point of the 15 seconds of the neutral expressions and running the calibration process on each participant's face to control for these individual variations in physiognomy. The FaceReader option for automatic continuous calibration was not used.

### Intersubject correlation (ISC) analyses

Intersubject correlations<sup>26,61</sup> were calculated by correlating the expression intensity functions produced by FaceReader using MATLAB (The MathWorks, Natick, MA). Each expression of each individual audience member was correlated with the same expression in every other individual in the audience separately (i.e. happy-happy, sad-sad etc.). Series of correlations were calculated within a 'sliding window': a discrete sample of facial expressions over a period of 1 minute (3000 face images at 50 faces/s). Initially the window was positioned at the start of the performance, and measures of expression intensity from the first 3000 images of every face were correlated with measures of expression intensity from the first 3000 images from each of the other faces. Similar calculations were performed as the window was moved forward in time in 2 s jumps (100 face images) until the end of the performance was reached. Finally, *r* values calculated in each window from all face pairs were averaged separately for each expression, to provide measures of how synchronized the different audience expressions were at different points during the performance.

### Audience engagement predictions

We restricted the analyses of how ISCs are related to predictions of audience engagement to 3 performances with large audiences (*n* = 44 to 50) that could contribute facial expression data for >90% of the whole performance in order to ensure a substantial number of ISCs (*n* = 946 to 1225) on which to base our analyses. In order to provide a prediction of the relative likelihood of audience engagement over the course of the

performance, the dramaturge (author LP) provided a score for each scene of each performance by viewing the videos of the performance, without knowledge of the facial expressions of the audience members or the results of the ISC analyses. Scenes were defined to the nearest second and were interpreted as a change in action in the performance (e.g. character entering/exiting/doing something new). Scores were on a simple 1-3 Likert scale; 3 = high engagement, 2 = moderate engagement, 1 = low engagement. These estimates (for an example see [Table S3](#)) were specific predictions of how strongly the different scenes will have commanded the audience member's attention, based upon a range of different features of the performance including: the level of physical and verbal action on the stage, the level of emotion displayed by the actors, and how interesting or surprising certain moments of dialogue or action should be. High engagement was defined as the audience being 'riveted' during that section of the performance, following the performance both intellectually and emotionally very closely. Low engagement was defined as the audience likely not attending, having their mind wandering, or thinking about topics other than the performance. Whilst moderate engagement was defined as between these other definitions.

To allow a comparison with the expression ISCs, the dramaturge's estimates of audience engagement during the performance were re-sampled by using the same sliding window method as for the ISC calculations. This had the effect of slightly smoothing the transitions between the 3 estimated levels of audience engagement. We then calculated the correlation between different expression ISCs and estimated audience engagement for each performance separately.

### Statistical analyses

Before any statistical analyses were conducted, we examined the data for normality by plotting data distribution functions and Q-Q plots. We transformed all  $r$  and ISC values using Fisher's  $r$  to  $z$  transform prior to statistical analysis.

In order to examine the influence of audience member spatial proximity on the synchrony of their facial expressions, we first calculated the Euclidian distance between all pairs of audience members. For simplicity we assumed that 1. the distance between adjacent seat positions on a row were equivalent to 1 unit, and 2. that the distance between similar numbered seats on adjacent rows was also equivalent to 1 unit. For statistical comparisons with ISCs, we used the square root of audience member distance to ensure that the data was normally distributed. We analyse the relationship between ISCs and proximity via Bayesian correlations and Bayesian regression. Although this analysis was conducted on transformed  $r$  to  $z$  values, and the square of the Euclidean distance between audience members, we additionally performed a permutation test to evaluate this analysis without assuming normality and homoscedasticity, and allow a more flexible assessment of the coefficients. We randomly permuted each of the expression ISC values and calculated the correlation coefficients between ISCs and distances; repeating this process 10,000 times to generate a null distribution. P-values for each expression were then calculated as the proportion of permuted correlations greater than or equal to the observed correlation.

Statistical analyses on ISCs were performed using Bayesian methods,<sup>83</sup> using JASP.<sup>84</sup> In all cases, a flat prior was specified as we had no prior published data that could inform our analyses and so we assigned a Cauchy prior distribution of  $r = 1/\sqrt{2}$  truncated to allow only positive effect size values. This study was not pre-registered, however, all analyses conducted are reported here.