Cognition 208 (2021) 104549

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Inferential affective tracking reveals the remarkable speed of context-based emotion perception

Zhimin Chen [a,*], David Whitney [a,b,c]

[a] Department of Psychology, University of California, Berkeley, CA 94720, United States of America
[b] Vision Science Program, University of California, Berkeley, CA 94720, United States of America
[c] Helen Wills Neuroscience Institute, University of California, Berkeley, CA 94720, United States of America

ARTICLE INFO

Keywords:
Emotion
Context
Tracking
Affect
Latency
Speed

ABSTRACT

Understanding the emotional states of others is important for social functioning. Recent studies show that context plays an essential role in emotion recognition. However, it remains unclear whether emotion inference from visual scene context is as efficient as emotion recognition from faces. Here, we measured the speed of context-based emotion perception, using Inferential Affective Tracking (IAT) with naturalistic and dynamic videos. Using cross-correlation analyses, we found that inferring affect based on visual context alone is just as fast as tracking affect with all available information including face and body. We further demonstrated that this approach has high precision and sensitivity to sub-second lags. Our results suggest that emotion recognition from dynamic contextual information might be automatic and immediate. Seemingly complex context-based emotion perception is far more efficient than previously assumed.

1. Introduction

Rapid inference about the internal emotional states of others is an essential and unique human ability. It is necessary for understanding others, interpersonal communication, and adaptive social functioning. Impaired emotion understanding is associated with a number of disorders, ranging from autism to schizophrenia to depression (Kohler, Turner, Gur, & Gur, 2004). The cognitive foundation of emotion understanding rests on our ability to derive and integrate information from a variety of cues across different modalities, including but not limited to facial expressions (e.g. Calder & Young, 2005; Ekman, 1992;), body postures (e.g. de Gelder, de Borst, & Watson, 2015), background scenes (e.g. Chen & Whitney, 2019; Righart & de Gelder, 2008), and vocal expressions (e.g. Cowen, Laukka, Elfenbein, Liu, & Keltner, 2019).

Previous research on emotion recognition has disproportionately focused on one channel of emotional information: the perception of facial expressions. This might be because faces are often thought to be attention-grabbing, uniquely salient, or evolutionarily significant (Ekman, 1992). The processing of facial expressions has also been considered to be rapid, efficient, and automatic (e.g. Fischer & Whitney, 2011; Poncet, Baudouin, Dzhelyova, Rossion, & Leleu, 2019; Yang & Yeh, 2018). However, faces are usually encountered within situational contexts in everyday life, and humans can seamlessly integrate contextual information in the process of recognizing emotion. For example, recent studies show that human observers can readily learn and utilize associations between emotion context and neutral stimuli (e.g. Ventura-Bort et al., 2016). Sometimes, contextual information can even facilitate and speed up the processing of faces when information is limited or ambiguous (Falagiarda & Collignon, 2019; Liedtke, Kohl, Kret, & Koelkebeck, 2018).

In recent years, an increasing body of research has shown that context can influence and modulate the interpreted emotions of facial expressions (Barrett, Mesquita, & Gendron, 2011; Wieser & Brosch, 2012) and this process has been suggested to be fast and automatic. For example, visual context can strongly influence perceived emotions from facial expressions when the context is irrelevant or subliminally presented (Aviezer, Bentin, Dudarev, & Hassin, 2011; Mumenthaler & Sander, 2015). This context effect remains intact even when observers are cognitively loaded by a concurrent task (Aviezer et al., 2011). The magnitude of this context effect has also been found to correlate with the degree of enhancement of an early electrophysiological component (Meeren, van Heijnsbergen, & de Gelder, 2005; Righart & de Gelder, 2006). Despite these advances, contextual information is still often regarded as secondary to facial information, mainly incorporated to

* Corresponding author at: 2121 Berkeley Way, 3rd floor, University of California, Berkeley, Berkeley, CA 94720, United States of America.
E-mail address: chenzhimin@berkely.edu (Z. Chen).

https://doi.org/10.1016/j.cognition.2020.104549
Received 22 April 2020; Received in revised form 8 December 2020; Accepted 9 December 2020
Available online 17 December 2020

modulate or disambiguate perceived emotion in faces.

Although there has been extensive research on perceived emotion from faces in the presence of contextual information, and the interaction between these sources of information, the inference of emotion from context alone in the absence of facial expression has remained largely unknown. However, a recent study shows that observers make remarkably good predictions of other peoples' affect (valence and arousal) when only the contextual information is available, while the face and body are blurred out (Chen & Whitney, 2019). Contextual information alone is therefore sufficient for an accurate interpretation of affective state, and the influence of the context on perceived emotion can be as substantial as the facial expression itself (Chen & Whitney, 2019). Context may therefore be a primary cue to emotion, not simply a secondary or modulatory cue.

The primacy and usefulness of context-based emotion perception depends on the speed of the available information. If context is as efficient (fast) as when the face and body information are present, it would suggest that context plays a pivotal and primary role. Previous literature has typically regarded the use of context alone, in the absent of overt expressions, to be indirect and deliberate because we have to rely on abstract causal principles rather than direct perceptual cues (Ekman, 1992; Skerry & Saxe, 2014). Further, conceptual models of perception tend to assimilate context or scene effects only at a relatively late and high-level processing stage (Bar, 2004). The implication of these previous studies is that contextual effects on emotion recognition might be relatively slow, or at least slower than the recognition of a facial expression. This may be, in part, because previous studies used unnatural or static stimuli and did not dynamically measure emotion.

In this paper, we adopted the inferential affective tracking method previously introduced in Chen and Whitney (2019) to characterize the continuous process of affect inference using dynamic and naturalistic stimuli. Here we developed a new analysis approach to measure the speed of recognizing emotion from contextual information alone. We further provide additional experiments to support and validate the precision of this approach. Our method is ideally suited to measure the speed of context-based emotion perception in the millisecond (msec) range, and it will be useful for quantitative models of emotion perception.

2. Experiment 1

The goal of experiment 1 was to quantify the speed of inferring emotion from contextual information alone, relative to the speed of recognizing emotion with all available information including facial expression. There are competing hypotheses in this experiment. One hypothesis suggests that contextual effects on emotion recognition might be relatively slow, in which case one might expect that inferring emotion from only contextual information would lag behind recognizing emotion with all available information. On the other hand, it is possible that contextual information could be processed with a short latency or in parallel with facial expression information, in which case the latency of inferred emotion perception could be very brief.

2.1. Method

2.1.1. Participants

We tested 90 healthy participants in total (16 male; age range 18–34, Mean = 21.0, SD = 4.22). They were students from the University of California, Berkeley participating for course credits. All participants had normal or corrected-to-normal vision. Informed consent was obtained from all participants and the study was approved by Institutional Review Board at the University of California, Berkeley.

2.1.2. Stimuli

We made use of a publicly available stimulus set from Chen and Whitney (2019). The stimulus set was originally made for experiments that assessed the unique contribution of context versus face and body information in emotion recognition. The stimuli consisted of 34 silent (no audio) video clips derived from Hollywood movies, home videos and documentaries. The lengths of the videos ranged from 36 s to 160 s, totaling 2749 s for all videos. The resolution of the videos was 1280 × 720 and the frame rate was 30 frames per second. The original silent videos with all visual information visible were defined as the "fully informed" condition (see Fig. 1A). For each video, we selectively masked the face and body of a target character, frame-by-frame, with a Gaussian blurred mask, to generate stimuli for the

"context-only" condition (see Fig. 1B). In the context-only condition, the target character was never visible. The blurred mask appeared in the videos an average of 77.1% (SD = 19.1%) of the time. During the period when the blurred mask was present, the mask area covered an average of 31.5% (SD = 11.6%) of the entire frame.

To effectively evaluate tracking performance, we used video clips that contained variations in emotion. This means that many videos in our dataset contained more than 1 emotion and some transitions from positive to negative emotions or from negative to positive. In most of the clips, the affect was quite heterogeneous (see Fig. S1, supplemental materials).

2.1.3. Procedure

Participants completed the experiment on a custom-made website online. The videos were presented to participants in a random order. Half of the videos that each participant viewed were from the fully informed condition and the other half from the context-only condition. Although it is slightly more complex, this mixed within-subject design has several advantages. First, every participant viewed a given video in either the fully informed or the context-only condition, but not both

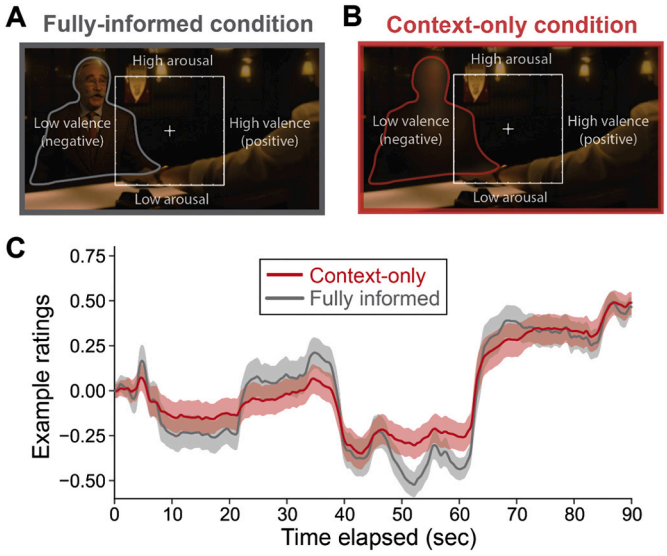


Fig. 1. Experimental conditions and data from a single example video in Experiment 1. (A) Participants viewed a silent movie clip while continuously reporting the valence and arousal of a specified character in the video. In the fully informed condition, participants were asked to track the affect of the target character (outlined in gray) when everything was visible. Due to copyright restrictions, the example video frame here is for visualization purposes only; the full set of videos is available here: https://osf.io/f9rxn/. (B) In the context-only condition, participants tracked the blurred target (outlined in red) while the context remained visible. (C) Example raw context-only valence ratings of the invisible target (red curve, $n = 41$ participants) appear to follow a similar time-course as the fully informed valence ratings of the visible target (gray curve, $n = 42$ independent participants). The data here are for one single video; a total of 34 videos were tested. Shaded regions represent bootstrapped 95% confidence intervals. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

conditions, which avoided any memory or interference effects between conditions. Second, this design controlled for subject-specific sources of noise such as network latency or monitor settings. As a result of the random assignment of videos to different conditions for each subject, different videos in the same condition may have had a slightly different number of participants assigned (within +/− 1 participant). On average, for every video in either the fully informed condition or the context-only condition, we collected affect ratings from 45 participants. To record real-time affective judgments, a 2D valence-arousal affect rating grid was superimposed on top of the video (Fig. 1A & 1B). Participants were required to position the mouse at the center of the affect rating grid before the video presentation. As participants watched each video, and in real-time, they were instructed to move a mouse pointer within the affect rating grid to continuously report the valence and arousal of the (blurred or visible) target character in the video. The mouse position was recorded every 20 milliseconds (50 Hz). After a video ended, participants were asked whether they had seen the video prior to the experiment, and they rated their level of familiarity with the video clip on a scale from 1 (Not at all familiar) to 5 (Extremely familiar). Participants were also asked whether the video played smoothly.

To estimate the noise ceiling in our data, 52 participants were asked to rate a random subset of video clips twice. The second rating was collected approximately 1 h after the first rating of the clip. The mean correlation coefficients between the initial ratings and the repeated ratings was used as the noise ceiling to normalize cross-correlation coefficients.

2.1.4. Data analysis

We confirmed that participants reported smooth video playback in 98.4% of the trials and the remaining 1.6% of trials were removed from the analysis. We also confirmed the exclusion of these 1.6% of trials did not change the results. We obtained mean affect ratings for every video under each condition by averaging across responses from all participants. This helps reduce noise in the data that is caused by idiosyncrasies from individual participants. We then used a cross-correlation analysis to detect the time lag between the mean fully informed affect ratings and the mean context-only affect ratings. Many studies on perception, performance, psychophysiology, and neuroscience use time-lagged cross-correlation analysis to assess the similarity and synchrony relationship between pairs of time series or signals (Dean & Dunsmuir, 2016). Cross-correlation is measured by incrementally shifting one signal in time and repeatedly calculating the correlation between two signals (see Fig. S2, supplemental materials). We used this technique to compare the emotion inferred when all information is available with the emotion inferred when only context information is available. The temporal offset at which the two signals are most synchronized (correlated) indicates the difference in perceptual latency between the context-only and fully informed conditions.

Time series analysis requires the series to be stationary (Shumway & Stoffer, 2014), which is defined as having constant statistical (e.g mean and variance) properties that do not change over time. We transformed our time series data to be stationary by applying differencing (Sims, 1988), which involves subtracting every value x_t from x_{t+1} to obtain successive differences between adjacent values in time. The Dickey-Fuller test confirmed that all transformed/differenced time series were stationary.

For every video, we computed the cross-correlation function (CCF) between the context-only and the fully informed condition using the differenced affect ratings. We applied Fisher z transformation on the CCFs and averaged the transformed z values to obtain the mean Fisher z. transformed CCF. We then estimated the noise ceiling by computing the CCFs between initial and repeated differenced ratings made by the same subject. These CCFs were also transformed to Fisher z values and averaged to obtain the mean Fisher z transformed CCF across videos. The peak Fisher z value of the mean transformed CCF between initial and repeated ratings was identified as the noise ceiling. We then divided the mean Fisher z transformed CCF between the context-only and the fully informed condition by this noise ceiling, and inverse transformed the normalized Fisher z back to Pearson r values, in order to obtain the normalized mean CCF across all videos. We then fit a skew-Cauchy distribution (Bahrami, Rangin, & Rangin, 2010) to the mean normalized CCF in order to capture the shape of the CCF. We confirmed that all skew-Cauchy curve fitting reached successful convergence to the optimal parameter values to ensure goodness of fit. We measured the time lag of the context-only condition by identifying the lag that has the highest correlation value along the skew-Cauchy curve fitted on the mean normalized CCF. The measured time lag was based on the mean CCF across videos because the focus of our study is on the temporal characteristics of context that are general to all video stimuli but are not specific to a single stimulus. The normalizing procedure by using the noise ceiling did not affect the result of time lag detection because the same noise ceiling was uniformly applied to cross correlation values of all time lags.

There is inevitable temporal dependency in the time series data because our method involved continuous affect ratings collected while viewing dynamic videos. Therefore, we did not use parametric tests (e.g. ANOVA and *t*-test) to evaluate statistical significance, because they make assumptions about the data and its distribution, which would often not hold true for our continuous data. Instead, we used non-parametric resampling (e.g. bootstrapping) and Monte Carlo permutation methods to generate null distributions and confidence intervals. To generate null distributions for CCFs, trial labels (whole continuous sequences of ratings) were shuffled (time points were not shuffled) and the CCF was calculated between affect ratings from randomly paired videos with different target characters. This permutation method preserved the exact temporal structure and autocorrelations inherent to the continuous ratings but not any clip-specific information. We estimated the bootstrapped confidence intervals by randomly sampling the CCFs of individual videos with replacement, recalculating the mean CCF across sampled videos and reidentifying the peak lag (that has the highest correlation value) from the bootstrapped mean CCF. This process was repeated 10,000 times to generate bootstrap distributions for the mean CCF and its peak lag; 95% confidence intervals were calculated based on the bootstrapped distributions.

As autocorrelation in time series could lead to spurious correlations, prewhitening has been proposed as a preprocessing step before calculating cross-correlation functions to further remove autocorrelation in time series (Shumway & Stoffer, 2014). Prewhitening is typically performed by fitting autoregressive integrated moving average (ARIMA) models to original time series and separating out the time series of residuals from the original series as the prewhitened series. Prewhitening has been shown to be effective when applied to some cases (Dean & Dunsmuir, 2016; Probst, Stelzenmuller, & Fock, 2012), while it is not always informative and can be detrimental in other cases (Bayazit & Önöz, 2007; Razavi & Vogel, 2018). In a separate analysis, we applied prewhitening to the differenced time series and confirmed that our basic results remained consistent whether or not it was applied.

2.2. Results

We calculated the skew-Cauchy fitted mean CCF between the context-only affect ratings of the invisible target character and the fully informed affect ratings of the visible target character (e.g., the cross correlation between the red and gray data in Fig. 1C). This reveals the relative delay between the use of contextual information alone and the use of all available information including the face and body. We observed a significant peak normalized cross-correlation between context-only and fully informed ratings (red line in Fig. 2A; mean $r = 0.52$, bootstrapped 95% CI: 0.38–0.65; $p < 0.001$, permutation tests), which confirmed that context information alone is indeed informative when inferring the affect of invisible characters. It is noteworthy that the fully-informed and context-only ratings were made by different groups

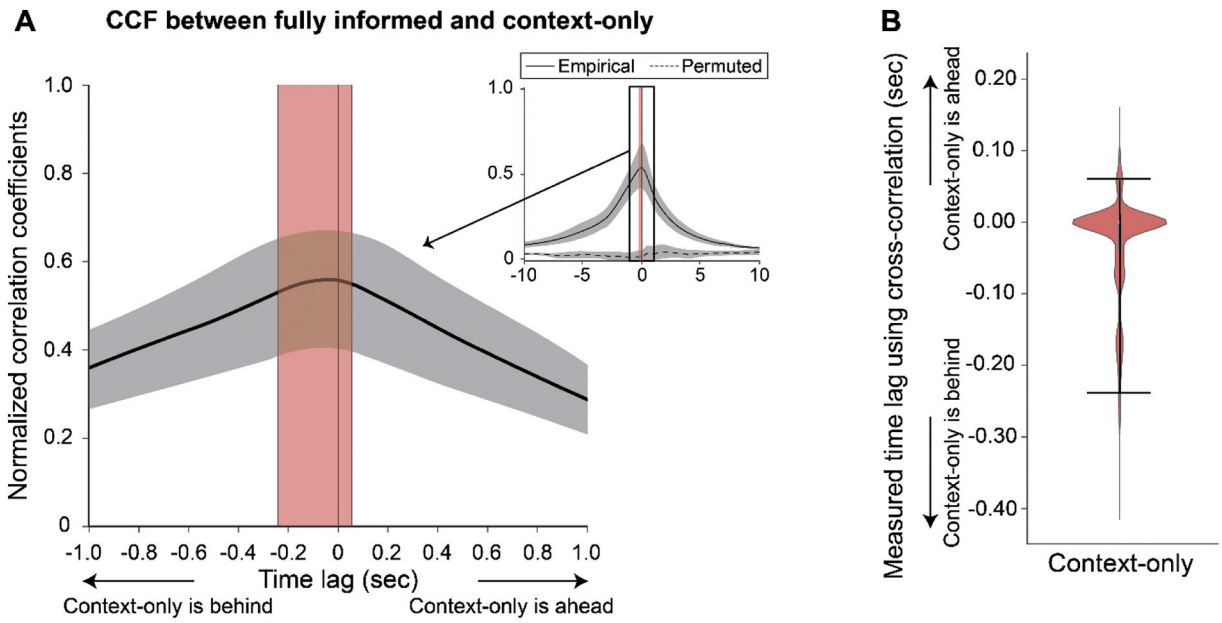


Fig. 2. Results of Experiment 1. Skew-Cauchy fitted cross correlation functions (CCF) for detecting a time lag between conditions. (A) Skew-Cauchy fitted cross correlation functions (CCF), between context-only affect ratings of the invisible target and fully informed affect ratings of the visible target, as a function of the time displacement/lag between them (solid black line). The shaded gray region around the black lines represents bootstrapped 95% confidence intervals of mean cross correlation coefficients averaged across clips and targets. The red shaded area represents bootstrapped 95% confidence intervals of measured peak lags identified from the skew-Cauchy fitted CCF. The dashed line near (near zero, in the inset plot) represents the permuted null cross correlation functions generated by shuffling the video clip labels of continuous ratings. (B) A violin plot of the measured time lags in the context-only condition estimated by bootstrapping the peak of the CCF in panel A. The peak delay is narrowly tuned and clustered around zero lag. Error bars represent bootstrapped 95% confidence intervals of the mean measured time lag (same as the red region in panel A). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of independent participants, so the cross correlation is not confounded by within-subject dependence or memory.

More importantly, we found that the peak of the skew-Cauchy fitted mean CCF had effectively zero time lag (Fig. 2A). We estimated the variability of the measured time lag using bootstrapping, which revealed no substantial lag (Fig. 2B and red shaded area in Fig. 2A, mode lag: 0 msec, mean lag: −33 msec, bootstrapped 95% CI: −240 to 60 msec). These results suggest that, on balance, context information alone is used nearly as fast as having additional facial expression information for emotion perception.

3. Experiment 2

The temporal cross-correlation analyses in Experiment 1 revealed that using context information alone added no substantial processing delay compared to using information that includes the face. Contextual information is therefore sufficient to perceive emotion with very little if any delay. However, one might wonder whether our inferential emotion tracking method or the cross-correlation approach have enough precision to detect a time lag if it is indeed present. We designed experiment 2 to test the precision of our method for detecting the temporal lag in tracking emotion continuously. We inserted a small lag (100 msec in Experiment 2a and 200 msec in Experiment 2b) in the video stimuli to create physically lagged conditions. We expected to find a significant time lag when cross-correlating between the affect ratings of the lagged condition and those of the no-lag condition.

3.1. Method

3.1.1. Participants

In total, we tested 80 healthy participants in Experiment 2a (18 male; age range 18–26, Mean = 20.2, SD = 1.48) and 76 healthy participants in Experiment 2b (21 male; age range 18–23, Mean = 20.1, SD = 2.43). Participants in Experiment 1, 2a, and 2b did not overlap. They were students from the University of California, Berkeley participating for course credits. All participants had normal or corrected-to-normal vision. Informed consent was obtained from all participants and the study was approved by Institutional Review Board at the University of California, Berkeley. With this sample size and the statistical effect we observed in this study, we can reach a power of over 0.9 with an alpha value of 0.05.

3.1.2. Stimuli

The stimuli in the “no-lag” condition were identical to the fully informed condition in Experiment 1 (Fig. 3A). In a separate condition, we edited these video stimuli by inserting a lag (100 msec in Experiment 2a and 200 msec in Experiment 2b), close to the start of each video; these were the “lagged” conditions (Fig. 3B). To insert the lag, we selected a random video frame between the first 5 to 10 s of each video, and we repeated the same frame for 100 (or 200) msec. The remaining frames in each video were therefore lagged by 100 (or 200) msec compared to the no-lag condition. Any fluctuations in timing could only add noise and reduce the measured precision but could not introduce a systematic lag between conditions.

3.1.3. Procedure

The procedure was identical to Experiment 1. The videos were presented to participants in a random order, with half of them from the lagged condition and the other half from the no-lag condition. Experiment 2 also used a mixed within-subject design because every participant viewed the same video in either the lagged or the no-lag condition, but not both conditions. On average, for every video in either the lagged condition or the no-lag condition, we collected affect ratings from 40 participants in Experiment 2a and 38 participants in Experiment 2b. The mouse position was recorded every 20 msec (50 Hz) in Experiment 2a and every 100 msec (10 Hz) in Experiment 2b.

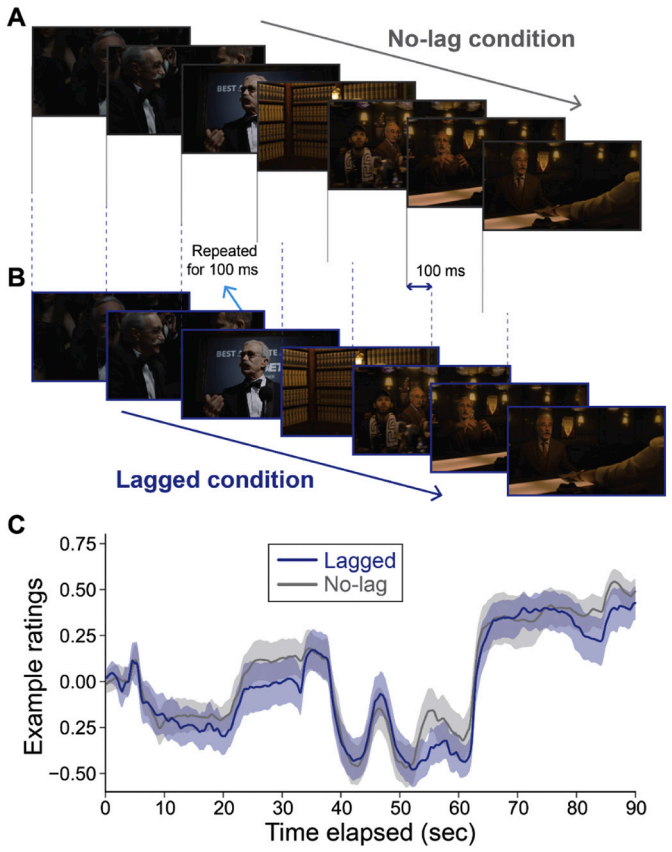


Fig. 3. Experiment 2a design and approach for an example video. (A) In the no-lag condition, participants viewed the original fully informed videos while tracking the valence and arousal of the target character. (B) In the lagged condition, participants viewed fully informed videos with a 100-msec time lag inserted at a random time after viewing the first 5–10 s of the video. To insert the time lag, we repeated the same video frame for 100 msec. As a result, all the video frames after the repeated frame lagged behind the no-lag condition by 100 msec. Experiment 2b had the same experimental design as Experiment 2a except that the lag inserted was 200 msec. (C) Example raw lagged valence ratings of the target (blue curve) relative to the no-lag valence ratings of the target (gray curve), for one example video. The example ratings for the no-lag condition have data from 35 participants and the example ratings for the lagged condition have data from 36 participants. Shaded regions represent bootstrapped 95% confidence intervals. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.1.4. Data analysis

The affect ratings collected for the first few seconds before the 100 (or 200) msec lag was introduced were removed from the cross-correlation data analyses. The truncated affect ratings were processed in the same way as Experiment 1. To examine the reliability of our lag detection method, we split the data in the no-lag condition into two halves and computed the cross-correlation between the mean ratings obtained from the two halves of data. We expected to find a narrowly tuned zero time lag in this (0 msec) condition. To further demonstrate the precision of our method, we quantified the relationship between measured time lag and physical time lag by fitting a linear regression function on every bootstrap iteration using the data from 0 msec, 100 msec and 200 msec. We expected the fitted linear regression function to have a slope close to 1 if the peak lag measured using our method matched the inserted physical lag.

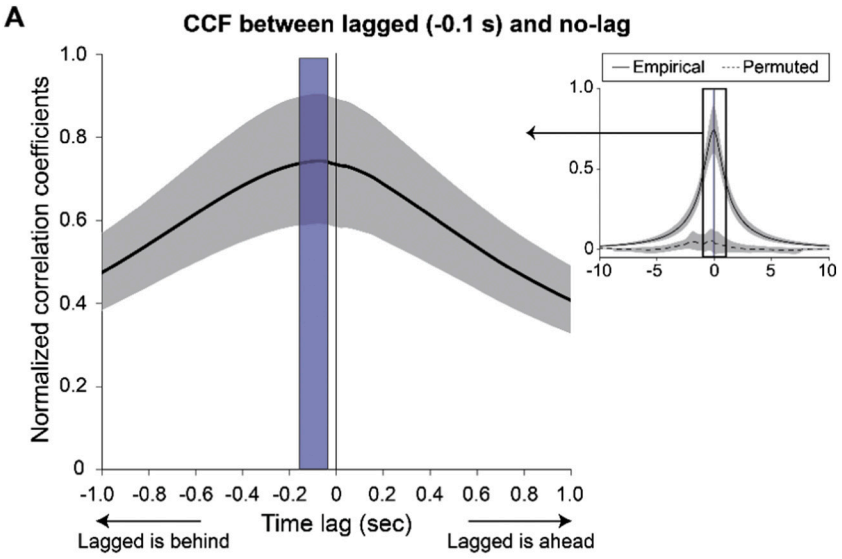
3.2. Results

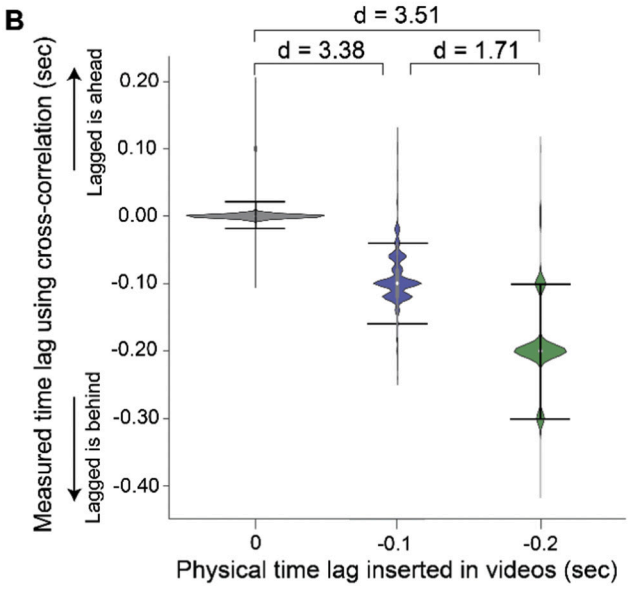
We calculated the skew-Cauchy fitted mean CCF between the no-lag affect ratings of the visible target character and the lagged affect ratings of the same character. We confirmed that the peak normalized cross-correlations between no-lag and lagged ratings were high for both Experiment 2a (peak of the black line in Fig. 4A; mean: 0.75; bootstrapped 95% CI: 0.58–0.86; $p < 0.001$, permutation tests) and 2b (mean: 0.76; bootstrapped 95% CI: 0.56–0.87; $p < 0.001$, permutation tests).

More importantly, the peak of the skew-Cauchy fitted mean CCF is clearly shifted from zero time lag (see the peak of the black line in Fig. 4A). For Experiment 2a, where we inserted a 100 msec time lag in the videos, we detected a significant time lag in the affect ratings (mode: −100 msec; mean: −87 msec; bootstrapped 95% CI: −160 to −20 msec; minus sign represents a lag instead of a lead; see the blue violin plot in Fig. 4B). In Experiment 2b, where we inserted a 200 msec time lag in the videos, we detected a significant time lag in the affect ratings (mode: −200 msec; mean: −193 msec; bootstrapped 95% CI: −300 to −100 msec; the green violin plot in Fig. 4B). When we split the no-lag ratings into two halves (Monte Carlo) and compared them using cross-correlation (the 0 msec condition), we verified that the measured time lag was narrowly tuned to zero (mode: 0 msec; mean: −4 msec; bootstrapped 95% CI: −20 to 20 msec).

To show that our method can distinguish a 100 msec lag from a 200 msec or a 0 msec one, we quantified the effect size (Cohen's d) for the difference in measured time lags between lag conditions (Fig. 4b). We found a Cohen's d of 3.38 between the 0 msec and the 100 msec lag conditions, a Cohen's d of 1.71 between the 100 msec and the 200 msec lag condition, and a Cohen's d of 3.51 between the 0 msec and 200 msec lag condition. These Cohen's d values all indicate very large effect sizes (Cohen, 2013). Furthermore, we found that the linear regression function fitted on the bootstrapped distributions of 0 msec, 100 msec and 200 msec data has a mode slope of 1 (see the black diagonal fitted regression line in Fig. 4C), which shows that that the measured time lag using our method matches the physical lag inserted. These results suggest that our method can resolve lags as small as 100 msec or less with high precision.

We also compared the measured time lag in the context-only condition in Experiment 1 (Fig. 2B) to that of the 0 msec, 100 msec and 200 msec lag conditions in Experiment 2 (Fig. 4). As a reminder, the measured time lag in the context only condition was near zero (Fig. 2B). We found a Cohen's d of 0.75 between this context-only condition and the 100 msec lag condition, and a Cohen's d of 1.77 between the context-only condition and the 200 msec lag condition. These Cohen's d values indicate medium to large effect sizes (Cohen, 2013). Because we have obtained the bootstrapped distribution of measured lags for 0 msec (Fig. 4C, gray dot cloud), 100 msec (Fig. 4B, blue dot clouds) and 200 msec (Fig. 4C, green dot clouds) lag conditions, we computed the Bayes factor to evaluate which of these distributions fits the context-only distribution (Fig. 4C, red dot cloud) the best. We found a Bayes factor (BF_{12}) of 5.11 in the comparison between the 0 msec (H_1) and the 100 msec (H_2) distributions, and a Bayes factor (BF_{12}) of 9.78 in the comparison between 0 msec (H_1) and the 200 msec (H_2) distributions. These Bayes factors suggest moderate (Lee & Wagenmakers, 2014) to substantial (Jeffreys, 1961; Kass & Raftery, 1995) statistical evidence in favor of the 0 msec (H1) distribution compared to the 100 msec or the 200 msec (H2) distributions. Although the labelling of Bayes factors varies slightly across different references (Jeffreys, 1961; Kass & Raftery, 1995; Lee & Wagenmakers, 2014), we opt to focus the literal interpretation of the Bayes factors, the fact that the measured time lag in context-only condition is 5.11 times more likely to be under to the 0 msec lag distribution than the 100 msec lag distribution. The Bayes factors show that latency of context-only affect perception is much more likely to be drawn from a population with 0 msec lag than a population with 100 or 200 msec lag, confirming that the context information is available with a remarkably short latency.


A
CCF between lagged (-0.1 s) and no-lag
Normalized correlation coefficients
1.0
0.8
0.6
0.4
0.2
0
-1.0 -0.8 -0.6 -0.4 -0.2 0 0.2 0.4 0.6 0.8 1.0
Time lag (sec)
Lagged is behind
Lagged is ahead
Empirical
Permuted
1.0
0.5
0
-10 -5 0 5 10


B
d = 3.51
d = 3.38
d = 1.71
Measured time lag using cross-correlation (sec)
Lagged is ahead
Lagged is behind
0.20
0.10
0.00
-0.10
-0.20
-0.30
-0.40
0
-0.1
-0.2
Physical time lag inserted in videos (sec)

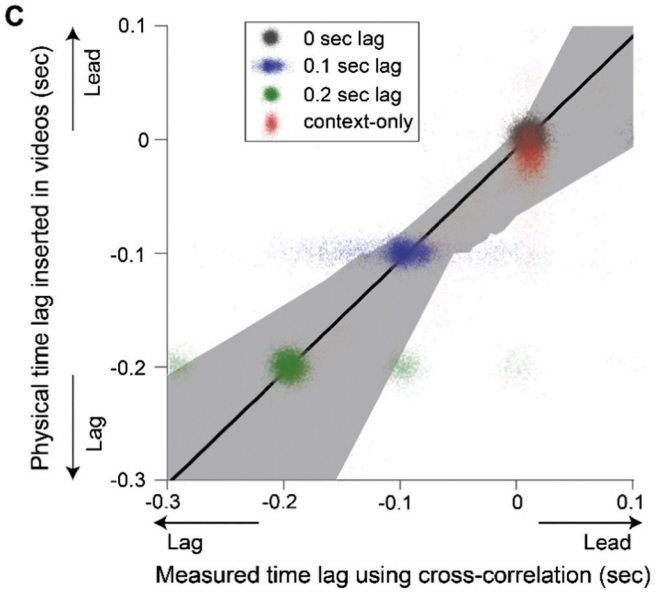

C
Physical time lag inserted in videos (sec)
Lead
Lag
0.1
0
-0.1
-0.2
-0.3
0 sec lag
0.1 sec lag
0.2 sec lag
context-only
-0.3 -0.2 -0.1 0 0.1
Lag
Lead
Measured time lag using cross-correlation (sec)

Fig. 4. Results of Experiment 2. (A) Experiment 2A: skew-Cauchy fitted CCF (solid black line) between lagged (−100 msec) affect ratings of the target and no-lag affect ratings of the target as a function of the time displacement/lag between them. Blue shaded area represents bootstrapped 95% confidence interval of measured time lags identified from the skew-Cauchy fitted CCF. The dashed line (inset) represents the permuted null cross correlation functions generated by shuffling the video clip labels of continuous ratings. Shaded gray region around black lines represents bootstrapped 95% confidence intervals of mean cross correlation coefficients averaged across clips and targets. (B) Violin plots of the measured time lags for Experiment 2a (−0.1 s physical lag in blue), Experiment 2b (−0.2 s physical lag in green), and the split-half analysis (0 s physical lag in gray). Error bars represent bootstrapped 95% confidence intervals of the mean measured time lag (same as the blue region in panel A). (C) The fitted linear relationship between physical time lag inserted in videos and measured time lag using our cross-correlation method. Gray shaded region shows the 95% confidence intervals of fitted linear regression functions using the bootstrapped data from 0 msec, 100 msec lag and 200 msec lag conditions. We then predicted the physical time lag of the context-only condition using the fitted linear regression function and the measured time lag of the context-only condition from Experiment 1 (in red). The measured and the predicted physical time lag for the context-only condition are both located closer to the measured time lags of the 0 msec condition than the 100 or 200 msec lag conditions. Data points are jittered for visualization purposes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Discussion

Our results provide the first measure of the speed of context-based dynamic emotion perception and show that the context is processed with a remarkably short latency, essentially as fast as using all available information including facial expressions. Our continuous inferential affective tracking technique (IAT), in combination with cross-correlation analysis (Experiment 1), has high precision in detecting a sub-second temporal lags as established by empirical experimental manipulations (Experiment 2). These results contrast with previous theories of inferred emotion from context, which implicitly or explicitly suggest that perceiving emotion from context is slower than emotion directly from faces (Bar, 2004; Skerry & Saxe, 2014). Our results support the alternative view that emotion from dynamic contextual information might be automatic and immediate. Seemingly complex context-dependent emotional inference and recognition is far more efficient than previously assumed.

The dynamic tracking method itself does not set limits on the precision of detecting a time lag between conditions. Although participants' affect ratings could be sluggish because of a large latency in the motor movement or mouse kinematics, this applies to all conditions (fully informed, and context-only). These sources of latency and temporal blurring do not affect our speed measurement because we focused on the difference between conditions, and the motor latency, for example, would therefore cancel-out in the comparison. Despite the 10 and 50 Hz sampling rate of the behavioral data, comparing between conditions can reveal a reliable temporal lag of 100 msec or less. Averaging across trials, for example, allows one to measure temporal differences in reaction time much finer than the resolution of the device itself (e.g., Donders, 1969).

Although the current study only concerns valence and arousal, our tracking method has also been demonstrated with ratings in discrete emotion categories rather than affect (Chen & Whitney, 2020). The tracking method was adapted from IAT and it was called inferential emotion tracking (IET). With IET, we showed that context remains essential for emotion recognition regardless of whether the emotion is reported as dimensional or categorical. Based on the current study, it is therefore plausible that context information is used very quickly for both affective and categorical emotion perception.

It is worth noting that the striking findings in this study are related to the failure to demonstrate a consistent lag between conditions different from zero. It is unquestionable that one cannot conclude that the null hypothesis is true when one fails to reject it. This is where Experiment 2a and 2b come in, to guide our interpretation of the measured CCF lag in Experiment 1, by showing that our method can reliably detect a sub-second lag. One possibility is that the time lag between fully informed and context-only affect ratings is smaller than the limit that we have tested (< 100 msec). Another possibility is that Experiment 1 involves a comparison between different conditions, which is inherently noisier than a comparison within the same condition. With the above constraints considered, we can still safely conclude that emotion inferred from context information alone does not yield a time lag as strong and consistent as the 100 msec lag tested in Experiment 2a. Context information is therefore available as early as any emotion-related visual information, and it has a very fast influence on perceived affect.

Although our study has shown that inferring emotion from only contextual information is processed with little delay, our results do not speak to what specific information in the context-only condition is essential for such fast emotion inference. The information in blurred masks on its own contains some colour or residual outline motion, but that cannot be used to perceive emotion accurately, as previously demonstrated (Chen & Whitney, 2019). However, the blurred mask is embedded in the scene context and it may interact with the context in a way that provides useful information. Scenes with other characters as part of the context may provide more information to allow for faster inference of emotion. However, we did not find a large or easily interpretable difference in measured time lag between videos with one character only or with more than one character (see Fig. S3,

supplemental materials). The overall valence of the affect ratings also did not change the results much: the relative latency of inferential emotion tracking was near zero for videos that were relatively more positive or negative (Fig. S4, supplemental materials). One might expect that familiarity with video content might play a role in determining the lag. We therefore analyzed the data excluding trials in which participants reported familiarity with the video. We found that the IET latency was similar regardless of familiarity with the videos (Fig. S5, supplemental materials).

Our findings are consistent with and extend a large body of work showing that the perceptual organization and integration of visual contextual information is fast and automatic. There are many extensively investigated visual processes that require the integration of some form of visual context and they have been suggested to be pre-attentive and automatic (for a review see Albright & Stoner, 2002). These rapid processes include but are not limited to analysis of shadows (Rensink & Cavanagh, 2004), perceptual filling-in (Mattingley, Davis, & Driver, 1997), texture segmentation (Zhaoping, 2000), figure-ground segregation (Kimchi & Peterson, 2008), etc. The perception of facial attributes such as expressions and attractiveness is also influenced by the context of other faces presented in the recent past or simultaneously (e.g. Liberman, Manassi, & Whitney, 2018; Wedell, Parducci, & Geiselman, 1987). Some studies have shown that neurons at early stages of cortical processing are involved in detecting contextual cues and representing the modulated information (Albright & Stoner, 2002). This evidence suggests that context-based processing could be primitive and efficient. Our results extend this to the some of the highest levels of visual cognition, including inferential emotion perception.

Our findings speak to the question of what enables emotion inference from context to be so fast. Context could be faciliatory for emotion recognition in a temporal and spatial manner. Emerging evidence in neuroscience supports that the human brain performs mental inference based on predictive encoding (Friston, 2010). This account has subsequently been extended to social and emotion perception (Otten, Seth, & Pinto, 2017). According to this account, extracting sequential regularities embedded in the temporal context to form predictions about upcoming events is an essential cognitive and neural process. These sequential regularities in the recent past can serve as the temporal context to constrain and shape inferences of others' emotions (Kimura, Kondo, Ohira, & Schröger, 2012). Similarly, spatial context may contain heuristics based on behavioral regularities in the social environment, which can then provide shortcuts for emotion processing (Marsh, 2002). For example, emotions like panic tend to spread in crowds, and get intensified beyond what any individual face can signal. Furthermore, detailed contextual information may facilitate the understanding of others' emotional states by actively engaging other empathic and interoceptive processes (Melloni, Lopez, & Ibanez, 2014).

Our results have implications for the underlying neural mechanisms of emotion perception. A common view is that contextual information modulates the neural processing of facial expressions through feedback connections (Wieser & Brosch, 2012). Our study suggests an alternative possibility, albeit speculative, that there might be a parallel pathway, independent of the pathway of facial analysis, for processing and extracting affective information from visual background context. This context-pathway is supported by work showing that independent and unique variance in emotion perception is carried by face and contextual information (Chen & Whitney, 2019) and it would likely involve brain regions that support the analysis of objects, scenes, bodies, and actions that constitute the interpretation of visual context. Using our approach, future neuroimaging experiments could generate encoding and decoding models to isolate the neural mechanism of context-based emotion perception.

Taken together, our findings reinforce the idea that context plays a critical role in supporting the rapid and robust understanding of others' emotion. This has practical implications for affective computing, which stresses the importance of fast and accurate emotion recognition. In light of our results, the implementation of a context processing stream should not be regarded as peripheral and superfluous, but rather essential. Emotion inference from context is a seemingly complex and challenging problem as visual scene context is heterogenous and the processing of it seems computationally expensive. However, we have shown that the human brain resolves it with remarkably speed and efficiency in a relatively effortless manner. To understand and exploit the brain's full potential in the realm affective computing, it is therefore important to shift our focus towards studying the cognitive and neural mechanisms underlying context-based emotion perception.

Declaration of Competing Interest

The authors declare no competing interests.

Acknowledgements

All authors contributed to the development of the study concept and the study design. Testing and data collection were performed by Z. Chen. Z. Chen performed the data analysis and interpretation under the supervision of D. Whitney. Z. Chen drafted the manuscript, and D. Whitney provided critical revisions. All authors approved the final version of the manuscript for submission. This work was supported in part by the National Institute of Health (grant no. 1R01CA236793-01) to D.W. We would like to thank Zhihang Ren for helpful discussions on data analysis. We also thank Meer Wu and Chloe Li for assistance with data analysis.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2020.104549.

References

Albright, T. D., & Stoner, G. R. (2002). Contextual influences on visual processing. *Annual Review of Neuroscience, 25*, 339–379. https://doi.org/10.1146/annurev.neuro.25.112701.142900.

Aviezer, H., Bentin, S., Dudarev, V., & Hassin, R. R. (2011). The automaticity of emotional face-context integration. *Emotion, 11*(6), 1406–1414. https://doi.org/10.1037/a0023578.

Bahrami, W., Rangin, H., & Rangin, K. (2010). A two-parameter generalized skew-Cauchy distribution. *Journal of Statistical Research of Iran JSRI, 7*(1), 61–72.

Bar, M. (2004). Visual objects in context. *Nature Reviews. Neuroscience, 5*(8), 617–629. https://doi.org/10.1038/nrn1476.

Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in emotion perception. *Current Directions in Psychological Science, 20*(5), 286–290. https://doi.org/10.1177/0963721411422522.

Bayazit, M., & Önöz, B. (2007). To prewhiten or not to prewhiten in trend analysis? *Hydrological Sciences Journal, 52*(4), 611–624. https://doi.org/10.1623/hysj.52.4.611.

Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews. Neuroscience, 6*(8), 641–651. https://doi.org/10.1038/nrn1724.

Chen, Z., & Whitney, D. (2019). Tracking the affective state of unseen persons. *Proceedings of the National Academy of Sciences of the United States of America, 116*(15), 7559–7564. https://doi.org/10.1073/pnas.1812250116.

Chen, Z., & Whitney, D. (2020). *Inferential emotion tracking (IET) reveals the critical role of context in emotion recognition* (Manuscript in press in Emotion).

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic press.

Cowen, A. S., Laukka, P., Elfenbein, H. A., Liu, R., & Keltner, D. (2019). The primacy of categories in the recognition of 12 emotions in speech prosody across two cultures. *Nature Human Behaviour, 3*(4), 369–382. https://doi.org/10.1038/s41562-019-0533-6.

de Gelder, B., de Borst, A. W., & Watson, R. (2015). The perception of emotion in body expressions. *Wiley Interdisciplinary Reviews: Cognitive Science, 6*(2), 149–158. https://doi.org/10.1002/wcs.1335.

Dean, R. T., & Dunsmuir, W. T. M. (2016). Dangers and uses of cross-correlation in analyzing time series in perception, performance, movement, and neuroscience: The importance of constructing transfer function autoregressive models. *Behavior Research Methods, 48*(2), 783–802. https://doi.org/10.3758/s13428-015-0611-2.

Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica, 30*, 412–431. https://www.ncbi.nlm.nih.gov/pubmed/5811531.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6*(3–4), 169–200. https://doi.org/10.1080/02699939208411068.

Falagiarda, F., & Collignon, O. (2019). Time-resolved discrimination of audio-visual emotion expressions. *Cortex; A Journal Devoted to the Study of the Nervous System and Behavior, 119*, 184–194. https://doi.org/10.1016/j.cortex.2019.04.017.

Fischer, J., & Whitney, D. (2011). Object-level visual information gets through the bottleneck of crowding. *Journal of Neurophysiology, 106*(3), 1389–1398. https://doi.org/10.1152/jn.00904.2010.

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

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford: Oxford University Press.

Kass, R. E., & Raftery, A. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773–795. https://doi.org/10.1080/01621459.1995.10476572.

Kimchi, R., & Peterson, M. A. (2008). Figure-ground segmentation can occur without attention. *Psychological Science, 19*(7), 660–668. https://doi.org/10.1111/j.1467-9280.2008.02140.x.

Kimura, M., Kondo, H., Ohira, H., & Schröger, E. (2012). Unintentional temporal context–based prediction of emotional faces: An electrophysiological study. *Cerebral Cortex, 22*(8), 1774–1785. https://doi.org/10.1093/cercor/bhr244.

Kohler, C. G., Turner, T. H., Gur, R. E., & Gur, R. C. (2004). Recognition of facial emotions in neuropsychiatric disorders. *CNS Spectrums, 9*(4), 267–274. https://doi.org/10.1017/s1092852900009202.

Lee, M. D., & Wagenmakers, E. J. (2014). *Bayesian cognitive modeling: A practical course.* Cambridge: Cambridge University Press.

Liberman, A., Manassi, M., & Whitney, D. (2018). Serial dependence promotes the stability of perceived emotional expression depending on face similarity. *Attention, Perception & Psychophysics, 80*(6), 1461–1473. https://doi.org/10.3758/s13414-018-1533-8.

Liedtke, C., Kohl, W., Kret, M. E., & Koelkebeck, K. (2018). Emotion recognition from faces with in- and out-group features in patients with depression. *Journal of Affective Disorders, 227*, 817–823. https://doi.org/10.1016/j.jad.2017.11.085.

Marsh, B. (2002). Heuristics as social tools. *New Ideas in Psychology, 20*(1), 49–57. https://doi.org/10.1016/S0732-118X(01)00012-5.

Mattingley, J. B., Davis, G., & Driver, J. (1997). Preattentive filling-in of visual surfaces in parietal extinction. *Science, 275*(5300), 671–674. https://doi.org/10.1126/science.275.5300.671.

Meeren, H. K. M., van Heijnsbergen, C. C. R. J., & de Gelder, B. (2005). Rapid perceptual integration of facial expression and emotional body language. *Proceedings of the National Academy of Sciences of the United States of America, 102*(45), 16518–16523. https://doi.org/10.1073/pnas.0507650102.

Melloni, M., Lopez, V., & Ibanez, A. (2014). Empathy and contextual social cognition. *Cognitive, Affective, & Behavioral Neuroscience, 14*(1), 407–425. https://doi.org/10.3758/s13415-013-0205-3.

Mumenthaler, C., & Sander, D. (2015). Automatic integration of social information in emotion recognition. *Journal of Experimental Psychology. General, 144*(2), 392–399. https://doi.org/10.1037/xge0000059.

Otten, M., Seth, A. K., & Pinto, Y. (2017). A social Bayesian brain: How social knowledge can shape visual perception. *Brain and Cognition, 112*, 69–77. https://doi.org/10.1016/j.bandc.2016.05.002.

Poncet, F., Baudouin, J.-Y., Dzhelyova, M. P., Rossion, B., & Leleu, A. (2019). Rapid and automatic discrimination between facial expressions in the human brain. *Neuropsychologia, 129*, 47–55. https://doi.org/10.1016/j.neuropsychologia.2019.03.006.

Probst, W. N., Stelzenmüller, V., & Fock, H. O. (2012). Using cross-correlations to assess the relationship between time-lagged pressure and state indicators: an exemplary analysis of North Sea fish population indicators. *ICES Journal of Marine Science: Journal Du Conseil, 69*(4), 670–681. https://doi.org/10.1093/icesjms/fss015.

Razavi, S., & Vogel, R. (2018). Prewhitening of hydroclimatic time series? Implications for inferred change and variability across time scales. *Journal of Hydrology, 557*, 109–115. https://doi.org/10.1016/j.jhydrol.2017.11.053.

Rensink, R. A., & Cavanagh, P. (2004). The influence of cast shadows on visual search. *Perception, 33*(11), 1339–1358. https://doi.org/10.1068/p5322.

Righart, R., & de Gelder, B. (2006). Context influences early perceptual analysis of faces–an electrophysiological study. *Cerebral Cortex, 16*(9), 1249–1257. https://doi.org/10.1093/cercor/bhj066.

Righart, R., & de Gelder, B. (2008). Recognition of facial expressions is influenced by emotional scene gist. *Cognitive, Affective, & Behavioral Neuroscience, 8*(3), 264–272. https://doi.org/10.3758/CABN.8.3.264.

Shumway, R. H., & Stoffer, D. S. (2014). *Time series analysis and its applications*. Springer. https://play.google.com/store/books/details?id=N-EYswEACAAJ.

Sims, C. A. (1988). Bayesian skepticism on unit root econometrics. *Journal of Economic Dynamics & Control, 12*(2), 463–474. https://doi.org/10.1016/0165-1889(88)90050-4.

Skerry, A. E., & Saxe, R. (2014). A common neural code for perceived and inferred emotion. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 34*(48), 15997–16008. https://doi.org/10.1523/JNEUROSCI.1676-14.2014.

Ventura-Bort, C., Löw, A., Wendt, J., Dolcos, F., Hamm, A. O., & Weymar, M. (2016). When neutral turns significant: Brain dynamics of rapidly formed associations between neutral stimuli and emotional contexts. *The European Journal of Neuroscience, 44*(5), 2176–2183. https://onlinelibrary.wiley.com/doi/abs/10.1111/ejn.13319.

Wedell, D. H., Parducci, A., & Geiselman, R. E. (1987). A formal analysis of ratings of physical attractiveness: Successive contrast and simultaneous assimilation. *Journal of Experimental Social Psychology, 23*(3), 230–249. https://doi.org/10.1016/0022-1031(87)90034-5.

Wieser, M. J., & Brosch, T. (2012). *Faces in context: A review and systematization of contextual influences on affective face processing* (vol. 3). https://doi.org/10.3389/fpsyg.2012.00471.

Yang, Y.-H., & Yeh, S.-L. (2018). Unconscious processing of facial expression as revealed by affective priming under continuous flash suppression. *Psychonomic Bulletin & Review, 25*(6), 2215–2223. https://doi.org/10.3758/s13423-018-1437-6.

Zhaoping, L. (2000). Pre-attentive segmentation in the primary visual cortex. *Spatial Vision, 13*(1), 25–50. https://doi.org/10.1163/156856800741009.

Supplemental materials for:

Inferential affective tracking reveals the remarkable speed of context-based emotion perception

Zhimin Chen [a], and David Whitney [a, b, c]

[a] Department of Psychology, University of California, Berkeley, CA 94720;

[b] Vision Science Program, University of California, Berkeley, CA 94720;

[c] Helen Wills Neuroscience Institute, University of California, Berkeley, CA 94720

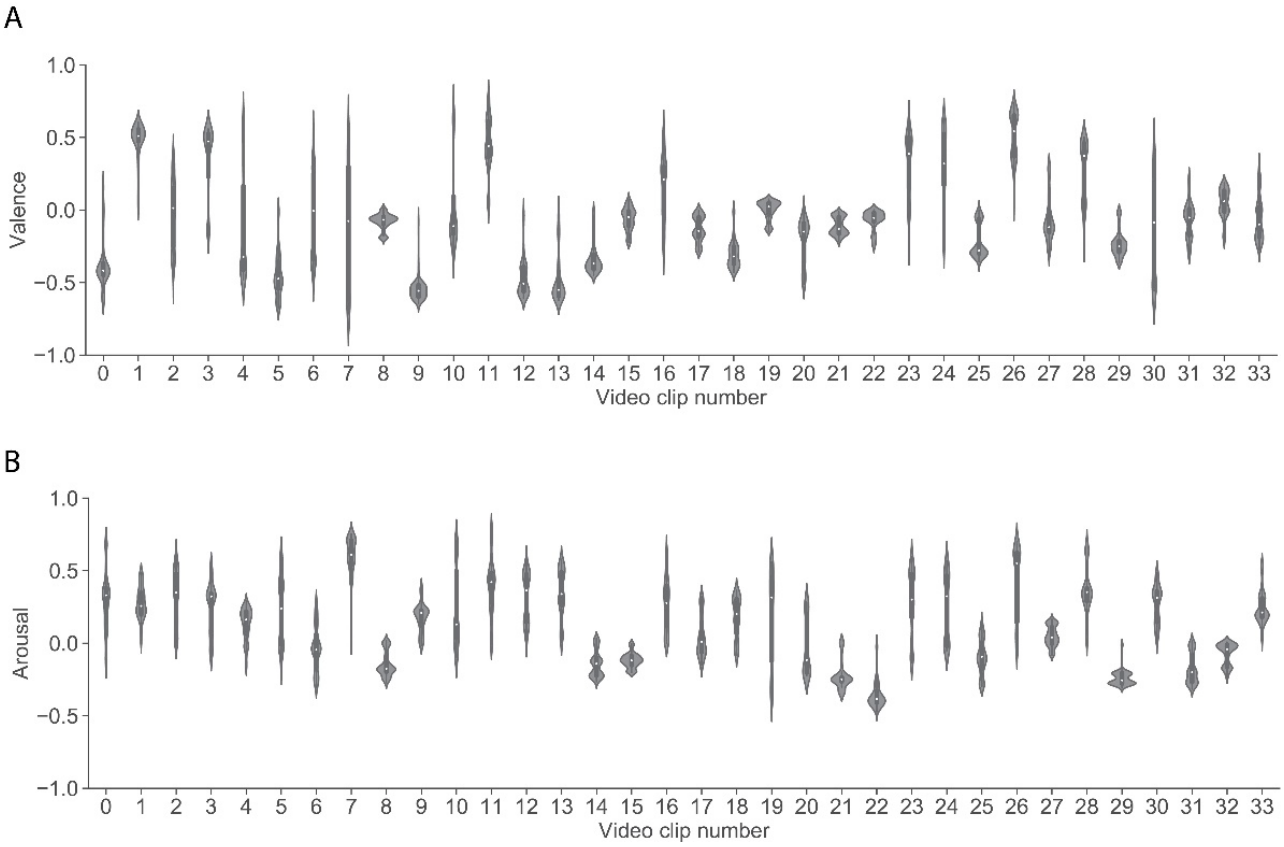


Fig. S1. The distribution of valence (A) and arousal (B) ratings in all 34 video clips. To effectively evaluate tracking performance, we used video clips that contained variations in emotion. This means that many videos in our dataset contained more than 1 emotion and some transitions from positive to negative emotions or from negative to positive. To quantify this, we averaged the valence/arousal ratings of the target character within every 1-second bin along the time course of every video. The violin plots show the distribution of valence and arousal for all binned time points within a given video (34 videos in total). In most of the videos, the affect was quite heterogeneous.

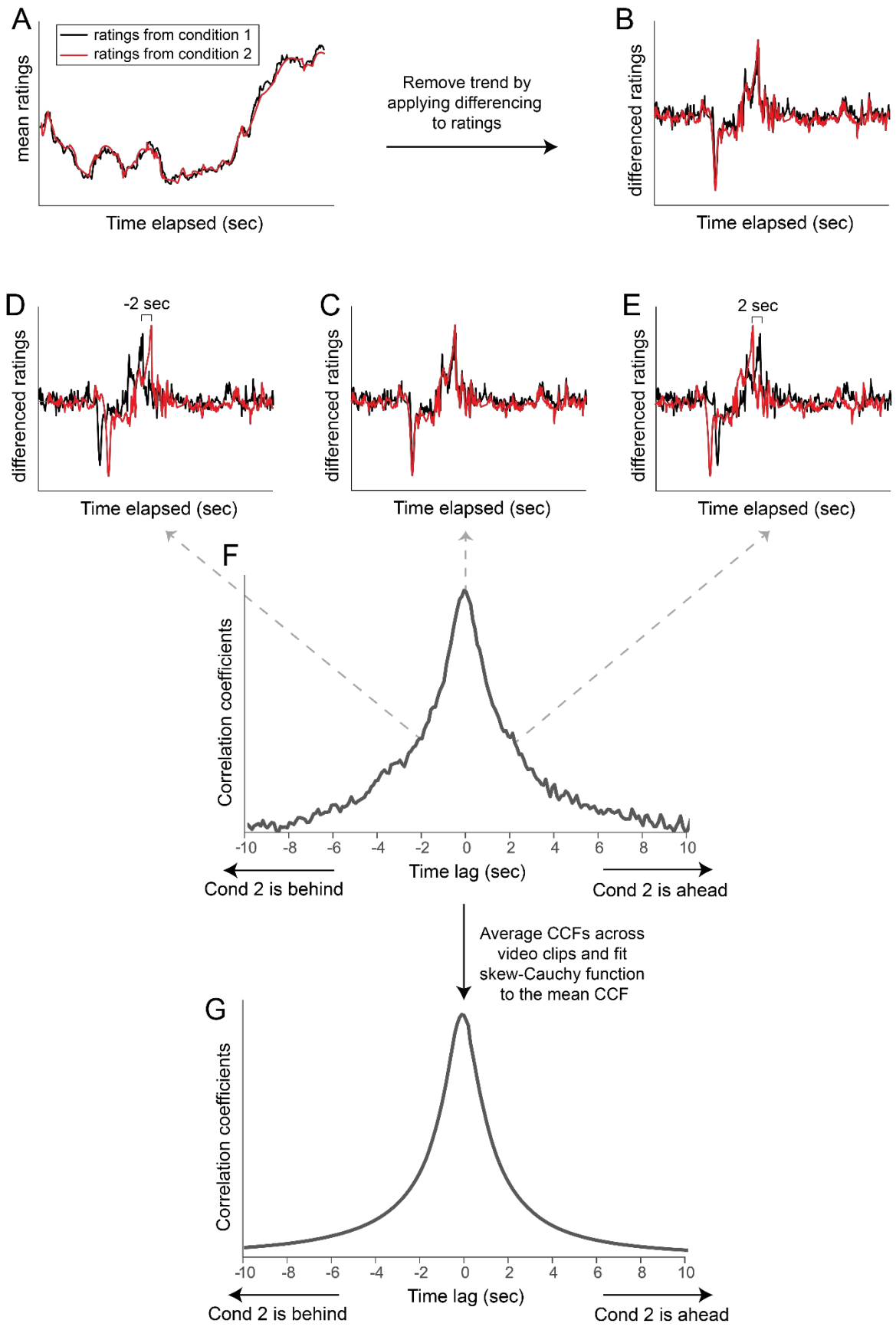

A
ratings from condition 1
ratings from condition 2
mean ratings
Time elapsed (sec)
Remove trend by applying differencing to ratings
B
differenced ratings
Time elapsed (sec)
D
-2 sec
differenced ratings
Time elapsed (sec)
C
differenced ratings
Time elapsed (sec)
E
2 sec
differenced ratings
Time elapsed (sec)
F
Correlation coefficients
-10 -8 -6 -4 -2 0 2 4 6 8 10
Time lag (sec)
Cond 2 is behind
Cond 2 is ahead
Average CCFs across video clips and fit skew-Cauchy function to the mean CCF
G
Correlation coefficients
-10 -8 -6 -4 -2 0 2 4 6 8 10
Time lag (sec)
Cond 2 is behind
Cond 2 is ahead

Fig. S2. Calculating a cross correlation function (CCF) to measure lag in emotion tracking. (A) We obtained mean affect ratings for each individual video in each condition by averaging across responses from all participants. The visualizations here are for one single video; the same analysis was performed all 34 videos. (B) We then transformed the data for each video to make it stationary by applying differencing (Sims, 1988), which involves subtracting every value xt from xt+1 to obtain successive differences between adjacent values in time. As an alternative, establishing stationarity with a pre-whitening approach (Shumway & Stoffer, 2011) did not change the results. The validity of the approach is confirmed by the near flat and zero CCF functions in the permuted null distributions (see Fig. 2A). (C, D, E, F) We computed the correlation coefficient between ratings from both conditions after shifting one series of ratings with different time displacement relative to the other. (e.g. -2 sec for D, 0 sec for C, 2 sec for E). This was performed for all possible time displacements to obtain the continuous cross correlation function in F for every video. (G) The CCFs for individual videos are averaged to obtain the mean CCF across videos. A skew-Cauchy distribution (Bahrami, Rangin, & Rangin, 2010) was fit to the mean CCF in order to capture the shape of the CCF and the time lag that has the highest correlation.

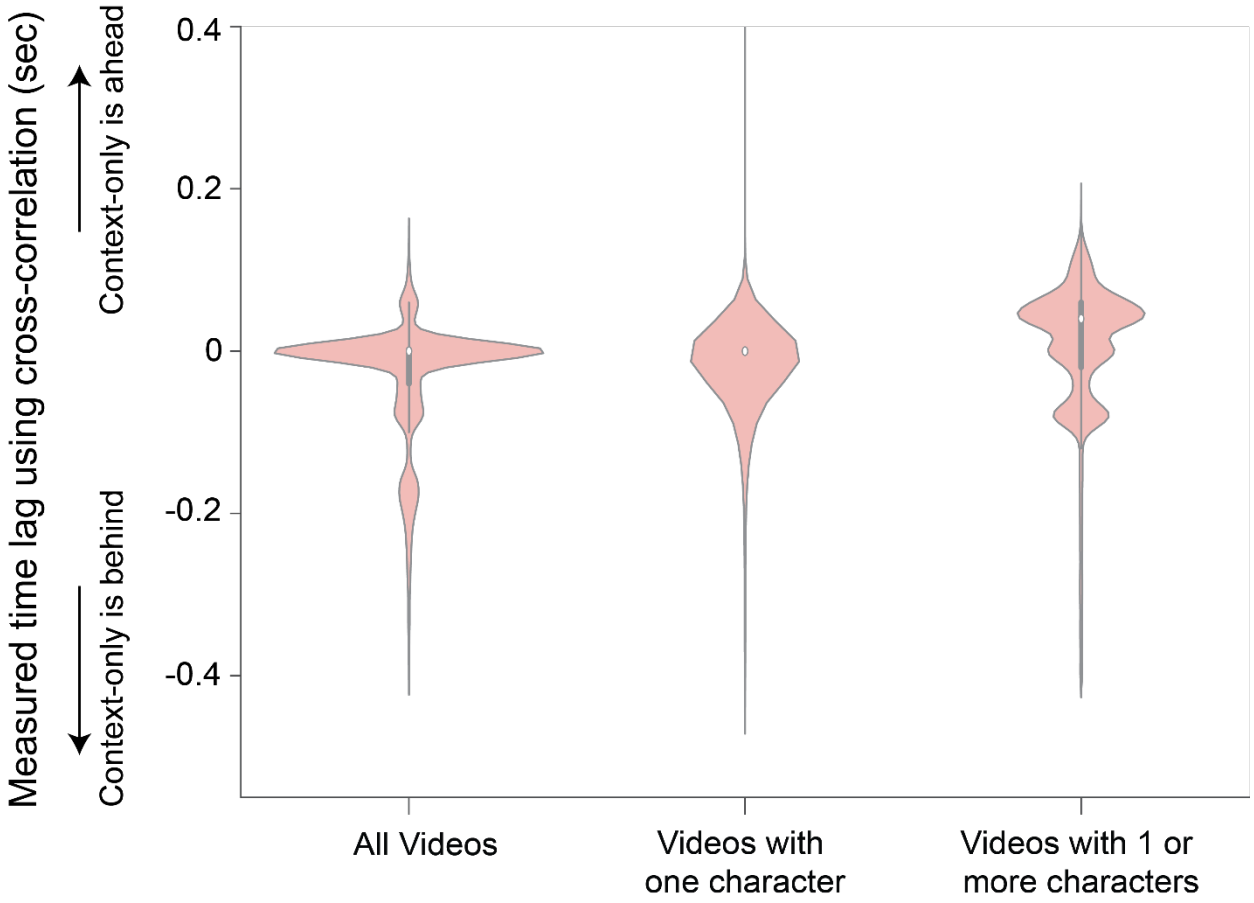


Fig. S3. Violin plots of the measured time lag in the context-only condition (Experiment 1) using all videos (n = 34), videos with only one character (n = 9), and videos with one or more characters (n = 25). We split the video clips into two groups depending on whether there was a partner character shown in the context, and we quantified the measured peak lags in that subset of videos. We found that in both sets of the videos, the distribution of measured time lags was near zero and it did not show a clear trend for substantial lags.

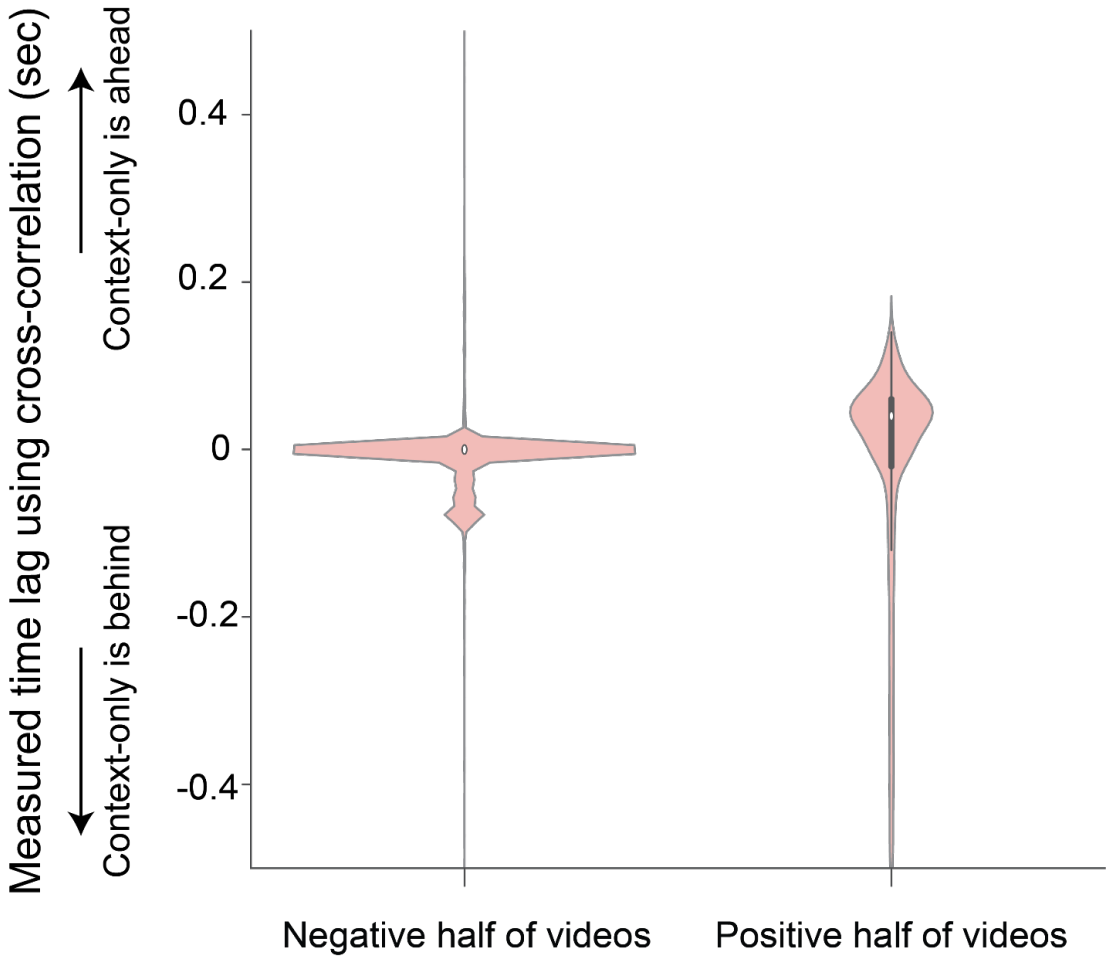


Fig. S4. The distribution of measured time lags in the context-only condition (Experiment 1) for the subsets of videos with relatively negative or positive valence. We split the video clips into two halves depending on whether affect ratings of the target characters were, on average, more negative or positive. We then quantified the measured time lags in these two groups of videos. We found that in both halves of the videos, the distribution of measured time lags was near zero.

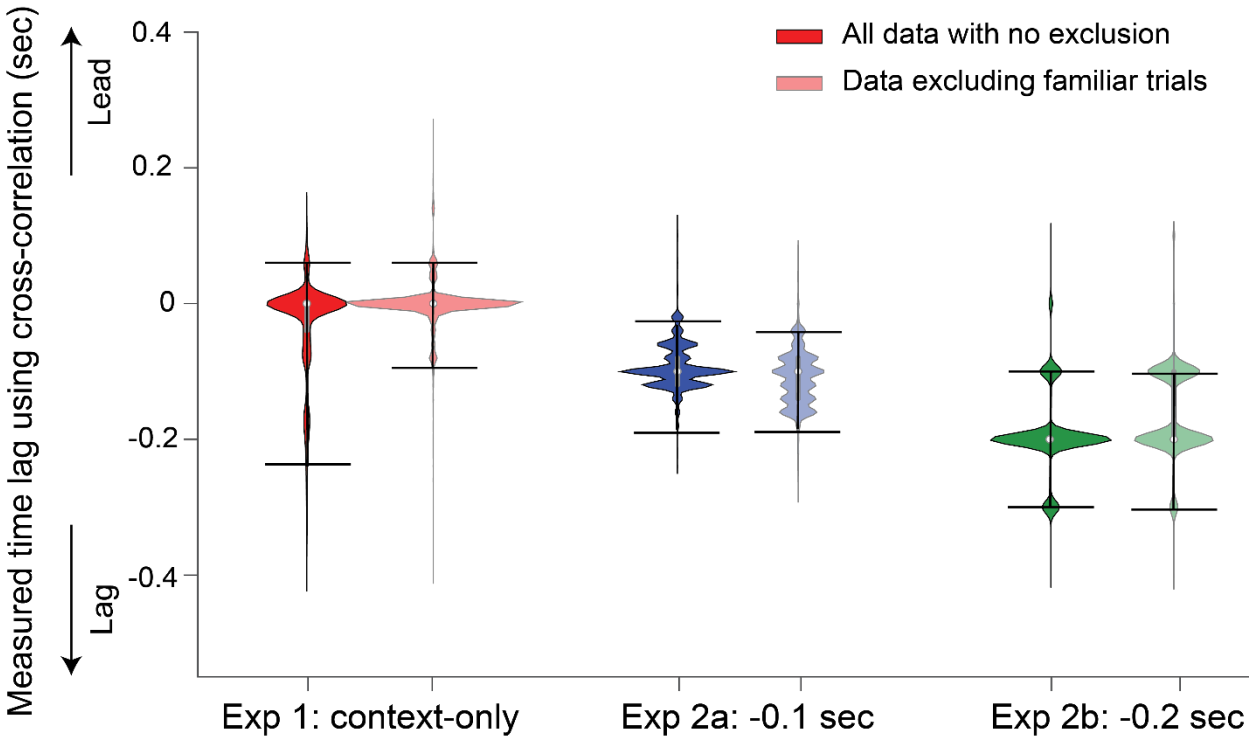


Fig. S5. The distribution of measured time lags for all experiments. On average, in 87% of all trials, participants reported that they had not seen the video content before participating in the experiments. We analyzed the data excluding trials in which participants reported familiarity. The distribution of measured time lags are similar to results obtained with data with no exclusion.

References

1. Sims, C. A. (1988). Bayesian skepticism on unit root econometrics. *Journal of Economic Dynamics & Control*, *12*(2), 463–474. https://doi.org/10.1016/0165-1889(88)90050-4

2. Shumway, R. H., & Stoffer, D. S. (2014). *Time Series Analysis and Its Applications*. Springer. https://play.google.com/store/books/details?id=N-EYswEACAAJ

3. Bahrami, W., Rangin, H., & Rangin, K. (2010). A Two-parameter Generalized Skew-Cauchy Distribution. *Journal of Statistical Research of Iran JSRI*, 7(1), 61-72.