How Do Static and Dynamic Emotional Faces Prime Incremental Semantic Interpretation?: Comparing Older and Younger Adults

Katja Münster (Katja.Muenster@uni-bielefeld.de)²,³

Maria Nella Carminati (mcarmina@techfak.uni-bielefeld.de)¹,³

Pia Knoeferle (knoeferl@cit-ec.uni-bielefeld.de)¹²,³

¹ SFB 673 “Alignment in Communication”
² Cognitive Interaction Technology Excellence Center
³ Department of Linguistics
CITEC, Inspiration 1, Bielefeld University
33615 Bielefeld, Germany

Abstract

Using eye-tracking, two studies investigated whether a dynamic vs. static emotional facial expression can influence how a listener interprets a subsequent emotionally-valenced utterance in relation to a visual context. Crucially, we assessed whether such facial priming changes with the comprehender’s age (younger vs. older adults). Participants inspected a static (Experiment 1, Carminati & Knoeferle, 2013) or a dynamic (Experiment 2) facial expression that was either happy or sad. After inspecting the face, participants saw two pictures of opposite valence (positive and negative; presented at the same time) and heard an either positively or negatively valenced sentence describing one of these two pictures. Participants’ task was to look at the display, understand the sentence, and to decide whether the facial expression matched the sentence. The emotional face influenced visual attention on the pictures and during the processing of the sentence, and these influences were modulated by age. Older adults were more strongly influenced by the positive prime face whereas younger adults were more strongly influenced by the negative facial expression. These results suggest that the negativity and the positivity bias observed in visual attention in young and older adults respectively extend to face-sentence priming. However, static and dynamic emotional faces had similar priming effects on sentence processing.

Keywords: Eye-tracking; sentence processing; emotional priming; dynamic vs. static facial expressions

Introduction

Monitoring people’s gaze in a visual context provides a unique opportunity for examining the incremental integration of visual and linguistic information (Tanenhaus et al., 1995). Non-linguistic visual information can rapidly guide visual attention during incremental language processing in young adults (e.g., Chambers, Tanenhaus, & Magnuson, 2004; Knoeferle et al., 2005; Sedivy et al., 1999; Spivey et al., 2002). Similar incremental effects of visual context information emerged in event-related brain potentials (ERPs) for older adults (e.g., Wassenaar & Hagoort, 2007). However, the bulk of research has focused on assessing how object- and action-related information in the visual context influences spoken language comprehension.

By contrast, we know little about how social and visual cues of a speaker in the visual context (e.g., through his/her dynamic emotional facial expression) can affect a listener’s utterance comprehension¹. In principle, a speaker’s facial expression of emotion could help a listener to rapidly interpret his/her utterances. With a view to investigating sentence processing across the lifespan and in relation to emotional visual cues, we assessed whether older adults exploit static and dynamic emotional facial cues with a similar time course and in a similar fashion as younger adults. The rapid integration of multiple emotional cues (facial, pictorial and sentential) during incremental sentence processing seems particularly challenging, yet such integration appears to occur effortlessly in natural language interaction. Here we examine how this integration is achieved using a properly controlled experimental setting.

To motivate our studies in more detail, we first review relevant literature on emotion processing, on the recognition of dynamic facial emotion expressions, and on emotion processing in young relative to older adults.

Affective Words and Face-Word Emotion Priming

Humans seem to attend more readily to emotional compared with neutral stimuli. For instance, participants in a study by Kissler, Herbert, Pyke, and Junghofer (2007) read words while their event-related brain potentials were measured. Positive and negative compared with neutral words elicited enhanced negative mean amplitude ERPs, peaking at around 250 ms after word onset. On the assumption that enhanced cortical potentials index increased attention, valenced relative to neutral information seems to immediately catch our attention (see e.g., Kissler & Keil, 2008 for evidence on endogenous saccades to emotional vs. neutral pictures; Nummenmaa, Hyönä, & Calvo, 2006 for eye-tracking...
evidence on exogenous attentional capture by emotional vs. neutral pictures; Lamy, Amunts, & Bar-Haim, 2008 for evidence on emotional vs. neutral facial expressions.

A further paradigm for examining emotion processing is emotional priming\(^2\). In emotional priming, emotionally congruent (vs. incongruent prime-target pairs) elicited faster response times when participants had to detect an odd face among other faces (e.g., a picture of an emotional face in an array of neutral faces, Lamy et al., 2008). Reaction times were shorter when an emotional facial expression was followed by a similar emotional expression (compared with a neutral one) on the next trial. Thus, “implicit memory for a recently attended [static] facial expression of emotion speeds the search for a target displaying the same facial emotion” (Lamy et al., 2008, p. 152). Such priming did not occur when the target was a neutral face.

In sum, emotional stimuli receive more attention than neutral stimuli; however psycho- and neurolinguistic research on emotional priming has focused on words. By contrast, we know little about how a smiling or a sad speaker face primes (visual attention during) spoken comprehension. Facial emotional expressions are part of communication and could thus play an important (rapid) role even in incremental sentence processing (much like extralinguistic cues from objects and events). If we observe rapid and incremental face-priming effects on ensuing visual attention to events during sentence comprehension, existing accounts of situated language processing will need to accommodate them (e.g., Knoeferle & Crocker, 2007).

Dynamic vs. Static Emotional Faces

Another novel aspect of our research is the direct comparison of dynamic and static prime faces. Research on emotion recognition and emotional priming has used mostly static pictures of emotional faces. By contrast, everyday social signals are dynamic. Notwithstanding, it has been shown that people can quickly and correctly decode static facial expressions (Kills et al., 2003).

However, higher recognition accuracy for dynamic than static stimuli has been reported in numerous studies (see Harwood, Hall, & Shrinkfield, 1999 for identification of emotions from moving and static videotaped and photographic displays from written and pictorial labels of emotions; Kozel & Gitter, 1968 for identification of different emotions from video vs. visual only vs. audio only vs. still pictures). Recio, Sommer, and Schacht (2011) measured ERPs while participants performed a categorization task for happy, angry and neutral faces (static vs. dynamic). An early posterior negativity and a late positive complex were both enhanced and prolonged for dynamic compared to static facial expressions. At the same time, response times were faster and accuracy higher for dynamic compared with static faces (see also Trautmann, Fehra, & Hermann, 2009 for related fMRI evidence). Against this background, we predict higher accuracy and faster response times with dynamic than static faces for the present studies.

The Nature of Emotion Processing Across the Ages

Evidence shows that the recognition of emotional stimuli is not invariant across the lifespan. Several ERP studies have found that the late positivity mean amplitude ERPs were more positive-going for negatively- than for positively-valenced words in young adults (e.g., Bernat, Bunce, & Shevrin, 2001; see Kanske & Kotz, 2007, Experiment 2). This ‘negativity bias’ found in young people generalizes to faces. For example, young adults preferentially attend to negative (afraid) faces (Isaacowitz et al., 2006).

By contrast, there is evidence showing that older people focus more on positive and less on negative information (‘positivity effect’, socio-emotional selectivity theory, Mather & Carstensen, 2005). In Mather and Carstensen’s (2003) study, older adults responded faster to a visually-presented dot probe when it appeared where a neutral face had been than where a negative face had been (see also e.g., Isaacowitz et al., 2007; Ruffman et al., 2008). Moreover, positive information (faces, pictures, life events) is memorized better than negative information in older age (Isaacowitz et al., 2006; Kennedy, Mather, & Carstensen, 2004; Mather & Carstensen, 2003). Thus, we can expect to see differences in how younger and older adults process emotional information. In particular, we expect the effects of negative and positive facial and sentence information to show opposite directionality.

The Present Research

We investigated how static (Experiment 1) versus dynamic (Experiment 2) emotional facial expressions prime the interpretation of positively and negatively valenced sentences, which were about emotionally valenced pictures. A further central aim was to assess potential differences in such priming effects for younger compared to older adults. Participants saw a picture of a person’s facial expression (Experiment 1) or watched a video of a person’s facial expression changing naturally from neutral to either happy or sad (Experiment 2). They were told this was the face of the speaker. Following this prime, two event photographs appeared on-screen, and shortly after, participants heard either a related positively- or negatively-valenced sentence (Table 1). The sentence always referred to one of the two event photographs. Participants indicated as fast and as accurately as possible whether the prime face matched the sentence by pressing a “yes”- or “no”-button.

During this task, we measured their eye movements to the event photographs, and response latencies in the face-sentence verification task. A priming effect in this task could manifest itself in the eye movements or in the response latencies or in both measures. If the emotional face

---

\(^2\) Priming: what people perceive at one moment in time (dubbed the ‘prime’) influences the perception and recognition of subsequent information (often dubbed ‘target’).
primes sentence processing and visual attention to the target photographs, then we should find more/earlier looks to a referenced photo when its valence matched (vs. mismatched) the valence of the prime face. Response times should further be faster and accuracy higher for congruent trials (i.e., when both prime and target are either positive or negative in valence), irrespective of age.

We expect age effects in response times and accuracy with slower and less accurate responses for older than younger adults (see, e.g., Mather & Carstensen, 2003; Salthouse, 2010). As for eye movements, if the negativity bias for younger adults generalizes to face-sentence priming, we should observe an enhancement of looks to the negative picture when prime and sentence are both negative. We should not observe this enhancement, or this enhancement should be smaller, when the sentence is positive. Crucially, the opposite behavior (i.e., an enhancement for positive face-sentence pairs) is expected for the older adults.

Considering the age biases, older adults should answer positively congruent trials faster and more accurately than negatively congruent trials. By contrast, younger adults should demonstrate the opposite response time and accuracy pattern or no bias. A negativity bias for younger adults should be evident in faster and more accurate responses to negatively than positively congruent trials.

Additionally, we predicted faster response times and more accurate responses for Experiment 2 than for Experiment 1, if the dynamic facial expression results in a processing advantage over the static facial expression.

**Experiment**

**Participants**
32 older (60–72 years, \( M = 64 \)) and 32 younger (19–29 years, \( M = 23 \)) adults participated in Experiment 1. 16 younger (18–30 years, \( M = 24 \)) and 16 older adults (60–80 years, \( M = 68 \)) participated in Experiment 2. All had German as their only mother tongue and normal or corrected-to-normal vision. All were unaware of the experiment purpose and gave informed consent.

**Materials and Design**
Materials and design were identical for both experiments, except that Experiment 1 used static emotional faces and Experiment 2 dynamic facial expressions as primes. There were 28 experimental target items consisting of a picture pair and corresponding sentence pair. Each picture pair had one positive and one negative picture, selected based on valence ratings (Lang, Bradley, & Cuthbert, 2008, the International Affective Picture System, IAPS). The experimental pictures were balanced for screen position. Within each item pair, they were controlled for arousal and visual similarity.

Each picture in a pair was associated with a corresponding negative or positive sentence (Table 1). The sentences were recorded in neutral intonation and at a relatively low pace, leaving a pause between phrases. The onsets of the critical word regions were aligned in each positive/negative sentence pair. Sentence pairs were matched for syllable length. We crossed the picture-sentence combinations with either a positive or negative static (Experiment 1) or dynamic (Experiment 2) prime face, in a 2 (prime face: negative vs. positive) x 2 (sentence: negative vs. positive) x 2 (picture: negative vs. positive) design. The experimental faces consisted of photographs or videos of sad and happy facial expressions. In Experiment 2, the face models first made a neutral face and then naturally changed into either a happy or a sad expression. A proportion of the filler items had neutral faces; for these, models were instructed to keep a constant neutral face. Experiment 1 and 2 used the same models, ensuring that the emotional prime face only varied in its form of presentation.

In addition to the 28 experimental items, we included 56 filler items. Each filler item also consisted of a picture pair, a sentence about one of the pictures, and prime faces (28 neutral; 14 positive; and 14 negative).

**Procedure**
An Eyelink 1000 Desktop Mounted System monitored participants’ eye movements. Only the right eye was tracked, but viewing was binocular. Prior to the experiment, participants gave informed consent, read the instructions and completed eight practice trials. After this the eye tracker was re-calibrated and the experiment began. Each trial started with a static facial expression (Experiment 1) or a video (Experiment 2). For Experiment 2, the facial expression stayed neutral (1.3 seconds) and then changed into the desired emotional expression (3.7 seconds). The

<table>
<thead>
<tr>
<th>Table 1: Sentence Structure and Example Sentences in German with a literal translation into English</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Positive Sentence</strong></td>
</tr>
<tr>
<td>IP</td>
</tr>
<tr>
<td><em>Es ist offensichtlich, dass</em></td>
</tr>
<tr>
<td>It is obvious that</td>
</tr>
<tr>
<td>NP1</td>
</tr>
<tr>
<td><em>die Kleine</em></td>
</tr>
<tr>
<td>the little (one)</td>
</tr>
<tr>
<td><strong>NP2</strong></td>
</tr>
<tr>
<td><em>die Melone</em></td>
</tr>
<tr>
<td>the melon</td>
</tr>
<tr>
<td><strong>ADJ</strong></td>
</tr>
<tr>
<td><em>heiter</em></td>
</tr>
<tr>
<td>cheerfully</td>
</tr>
<tr>
<td><strong>VERB</strong></td>
</tr>
<tr>
<td><em>verspeist.</em></td>
</tr>
<tr>
<td>eats.</td>
</tr>
<tr>
<td><strong>Negative Sentence</strong></td>
</tr>
<tr>
<td>IP</td>
</tr>
<tr>
<td><em>Es ist offensichtlich, dass</em></td>
</tr>
<tr>
<td>It is obvious that</td>
</tr>
<tr>
<td>NP1</td>
</tr>
<tr>
<td><em>die Blonde</em></td>
</tr>
<tr>
<td>the blonde (woman)</td>
</tr>
<tr>
<td><strong>NP2</strong></td>
</tr>
<tr>
<td><em>die Migräne</em></td>
</tr>
<tr>
<td>the migraine</td>
</tr>
<tr>
<td><strong>ADJ</strong></td>
</tr>
<tr>
<td><em>gereizt</em></td>
</tr>
<tr>
<td>fretfully</td>
</tr>
<tr>
<td><strong>VERB</strong></td>
</tr>
<tr>
<td><em>verflucht.</em></td>
</tr>
<tr>
<td>curses.</td>
</tr>
</tbody>
</table>
prime face was then disappeared and the valenced target photographs appeared; 1500 ms later, the sentence was presented. Participants verified via a button press on a Cedrus (RB 834) response box whether (“yes” or “no”) the face and sentence matched in valence. The timeout was 1500 ms after sentence end for young, and 3000 ms for older adults. Participants were advised to answer as quickly and accurately as possible. The (left/right) position of the yes/no-answer button was counterbalanced across participants.

Analysis

We divided the sentence into critical regions. The first region (the first noun phrase, NP1) extends from the onset of NP1 until the onset of NP2 (Table 1). It represents the first point in time at which the sentence disambiguates the target picture. We also analyzed gaze over a longer time period ('long region') to uncover effects during the sentence. This period comprised the entire embedded sentence starting from its first disambiguating word (NP1). For each region, we computed the mean log gaze probability ratio according to the formula: $\ln \left( \frac{p(\text{negative picture})}{p(\text{positive picture})} \right)$. $\ln$ refers to the logarithm and $p$ refers to probability. This ratio expresses the bias of inspecting the negative relative to the positive picture. The ratio does not violate the independence and homogeneity of variance assumptions, which makes it suitable for comparing looks to two scene regions with parametric tests such as Analyses of Variance (ANOVAs, see, e.g., Arai, Van Gompel, & Scheepers, 2007). More looks to the negative (vs. positive) picture are indexed by a positive log ratio. More looks to the positive (vs. negative) picture are indexed by a negative log ratio.

We computed mean log gaze probability ratios for each region separately by participants and items. These means were then subjected to ANOVAs with participants and items as random effects. We report ANOVAs on the combined eye-movement data for both groups. Unless otherwise stated, group was a between-participant factor in the analysis by participants and a within-item factor in the analysis by items.

Reaction times were computed from NP1 onset. Accuracy scores (excluding trials with timeouts and incorrect responses) were computed for each group by condition. In an additional analysis, we combined the data of the two experiments and used Experiment (1 vs. 2) as a factor to detect a possible difference between the two experiments.

Results

Main results for the eye-movement analysis: The results from Experiment 1 showed that fixations on the pictures were increased when the speaker's (static) face was emotionally congruent (vs. incongruent) with the sentence. Crucially, this enhancement was modulated by age. The effect for the older adults was more pronounced with positive faces, whereas the effect was stronger for younger adults with negative faces (Carminati & Knoeferle, 2013). Here we report in detail the new data from Experiment 2, as well as the between-experiment comparison. Figure 1 illustrates the results from Experiment 2 and specifically how the dynamic face affected looks to the pictures, independent of sentence valence. For the long region (Fig 1), older adults looked more at the positive picture after seeing a positive (vs. negative) prime face, and inspected the negative picture more after seeing a negative (vs. positive) face. By contrast, younger adults preferred to inspect the negative picture independent of face valence (face x group interaction in the item analyses, $p < .05$). Older adults further had a numerically bigger preference for the positive picture after a positive face than for the negative picture after a negative face (Figure 1).

More importantly, Figure 2 shows how looks to the sentence-matching picture were modulated by age in the long region (face x sentence x age interaction): Younger participants were more likely to look at the negative picture after they had inspected a negative (vs. positive) prime face if the sentence was also negative (pairwise comparison: $p < .05$), but the opposite pattern was absent (i.e., no difference) if the sentence was positive; by contrast, older adults were more likely to inspect the positive picture after a positive (vs. negative) facial expression but only if the sentence was positive (pairwise comparison, $p < .05$). This effect was also reliable early, in the NP1 region ($p < .05$ by participants). In short, as in Experiment 1 (static faces) we see face-sentence priming only for negative face-sentence pairs in young, and only for positive face-sentence pairs in older adults.
Finally, analyses on the combined data from Experiments 1 and 2 confirmed all the effects found in the analyses on the separate experiments; importantly no interactions with experiment were observed.

**Response times:** The results did not differ between experiments and we report the new results for Experiment 2. Response times were slower for older than young adults (p < .01); slower for negative than positive sentences (p < .05); and slower for negative sentences in the older than the younger group (sentence x group interaction, p < .05). Participants’ verification times were also faster for incongruous than congruous face-sentence pairs (p < .05).

**Main results for the accuracy analysis:** Accuracy results did not differ between experiments and we report the new results for Experiment 2. Figure 3 shows that younger people were more accurate than older people (p < .05). However, older adults’ accuracy was higher than younger adults’ for positive compared with negative sentences. Thus, older adults seem to have benefitted more from positive sentences in answering the verification question. Interestingly, responses were more accurate when face and sentence valence mismatched than when they matched (Figure 3), and this mismatch advantage was more pronounced in older adults. Young adults only displayed a mismatch advantage for negative sentences.

![Figure 3: Accuracy Scores](image)

**Discussion**

An emotional speaker face primed both older and young adults’ visual attention to valenced pictures as soon as it became clear which picture the sentence referred to. However, crucially, this influence was modulated by age. Priming occurred only with the negative face-sentence combinations in young, and for positive face-sentence combinations in older adults. This confirms our hypotheses and suggests that visual attention, reflecting sentence interpretation, is guided by a negativity bias in young and by a positivity bias in older adults (Figure 2).

Moreover, younger participants showed an overall visual preference for the negative picture, regardless of face valence, but older adults, were clearly influenced by the prime face in the expected direction (Figure 1). Furthermore, older people’s positive picture preference was numerically bigger than their negative picture preference, providing further evidence for a positivity bias (Figure 1).

Somewhat unexpectedly, all participants responded significantly faster and more accurately to incongruent than congruent face-sentence valence items (Figure 3). This “mismatch” effect was stronger in older than young adults. One, admittedly speculative, reason for this unexpected pattern is that for some kinds of information in visual context, dissimilarities with language may be easier to verify than similarities. Increased response latencies for matches compared with mismatches have also been reported by Vissers et al. (2008) when young adults verified a spatial description against a line drawing. This mismatch effect does also not depend on the dynamics of the prime face, as both experiments yielded the same results. Contrary to our initial predictions, dynamic (vs. static) emotional facial expressions did not enhance the post-comprehension processing of the sentence, and they did not enhance eye movement behavior either. Thus, although dynamic facial expressions are recognized faster and more accurately than static facial expressions (Recio, Sommer, & Schacht, 2011), this ‘recognition advantage’ for dynamic expressions does not seem to generalize to the specific context of our study.

However, age plays a crucial role in emotional priming. One possible account for older adults’ focus shifts towards positive events is different fixation strategies for identifying emotions. Perhaps older adults extract different information from faces than younger people (Mill et al., 2009). In addition, our results support existing findings that younger adults are more sensitive to negative than positive stimuli (e.g., Holt, Lynn, & Kuperberg, 2008; Taylor, 1991) in the sense that they are more facilitated by the negative face in processing the negative sentence. The decline in older adults’ emotion processing skills and general cognitive functions (Mill et al., 2009) could go hand in hand with a change in fixation strategies in causing the change from a negativity bias towards a focus on positive information.

Overall thus the observed face-sentence priming effects corroborate and extend existing findings about age differences in emotion recognition. Emotional primes, regardless of whether they are static or dynamic, can facilitate the interpretation of an affective sentence. Crucially, age modulated this facilitation, with older adults’ showing increased facilitation from positive and younger adults from negative face primes.

**Acknowledgements**

This research was funded by the SFB 673 “Alignment in Communication” and by the Cognitive Interaction Technology Excellence Center (German Research Foundation, DFG). We thank participants for their support.

**References**


Bernat, E., Bunce, S., & Shevrin, H. (2001). Event related potentials differentiate positive and negative mood adjectives during both supraliminal and subliminal visual


