

Surprise: unfolding of facial expressions

Marret K. Noordewier & Eric van Dijk

To cite this article: Marret K. Noordewier & Eric van Dijk (2019) Surprise: unfolding of facial expressions, *Cognition and Emotion*, 33:5, 915-930, DOI: [10.1080/02699931.2018.1517730](https://doi.org/10.1080/02699931.2018.1517730)

To link to this article: <https://doi.org/10.1080/02699931.2018.1517730>



© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 06 Sep 2018.



Submit your article to this journal [↗](#)



Article views: 961



View Crossmark data [↗](#)

Surprise: unfolding of facial expressions

Marret K. Noordewier and Eric van Dijk

Faculty of Social and Behavioural Sciences, Social and Organizational Psychology, Leiden University, Leiden, The Netherlands

ABSTRACT

Responses to surprising events are dynamic. We argue that initial responses are primarily driven by the unexpectedness of the surprising event and reflect an interrupted and surprised state in which the outcome does not make sense yet. Later responses, after sense-making, are more likely to incorporate the valence of the outcome itself. To identify initial and later responses to surprising stimuli, we conducted two repetition-change studies and coded the general valence of facial expressions using computerised facial coding and specific facial action using the Facial Action Coding System (FACS). Results partly supported our unfolding logic. The computerised coding showed that initial expressions to positive surprises were less positive than later expressions. Moreover, expressions to positive and negative surprises were initially similar, but after some time differentiated depending on the valence of the event. Importantly, these patterns were particularly pronounced in a subset of facially expressive participants, who also showed facial action in the FACS coding. The FACS data showed that the initial phase was characterised by limited facial action, whereas the later increase in positivity seems to be explained by smiling. Conceptual as well as methodological implications are discussed.

ARTICLE HISTORY

Received 17 July 2015
Revised 22 August 2018
Accepted 22 August 2018

KEYWORDS

Surprise; unexpectedness;
facial expression; FACS;
repetition-change

When people are confronted with unexpected stimuli, they experience surprise (e.g. Meyer, Niepel, Rudolph, & Schützwohl, 1991; Meyer, Reisenzein, & Schützwohl, 1997; Noordewier & Breugelmans, 2013; Noordewier, Topolinski, & Van Dijk, 2016; Reisenzein, 2000b). Surprise is characterised by the interruption of ongoing thoughts and activities, a feeling of surprise, and the direction of attention at the surprising stimulus to make sense of it (e.g. Camras et al., 2002; Horstmann, 2006; Meyer et al., 1991; Meyer et al., 1997; Reisenzein, 2000b; Scherer, 2001). Once people make sense of the surprising stimulus, other affective states follow depending on the nature of the surprising event (Ekman, 2003; Noordewier & Breugelmans, 2013; Tomkins, 1984). Then, people become for instance happy with a positive surprise or disappointed with a negative surprise (see also Noordewier et al., 2016).

Responses to surprising stimuli thus unfold depending on the dynamics of sense-making

(Noordewier et al., 2016). Initial responses are primarily driven by the unexpectedness of the surprising outcome and reflect an interrupted and surprised state. Later responses are more likely to incorporate the valence of the surprising outcome itself, as it reflects the state after sense-making when the outcome is understood (Meyer et al., 1991, 1997; Noordewier et al., 2016; Noordewier & Breugelmans, 2013). The unfolding logic thus situates surprise in the initial phase that is characterised by interruption and feeling surprised (Meyer et al., 1991, 1997; Noordewier et al., 2016; Noordewier & Breugelmans, 2013), and sets it apart from the affective state that follows it once people understand what has happened (Tomkins, 1984). So, even if the surprising stimulus is positive, people first experience this brief phase of interruption and surprise, before they can appreciate and welcome the outcome as is it. This differentiation does not mean that correlates of surprise cannot linger after sense-making (e.g. residue of arousal), but it does

CONTACT Marret K. Noordewier  m.k.noordewier@fsw.leidenuniv.nl

© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

mean that to understand surprise, we should distinguish between initial and later responses to surprising stimuli (cf. Noordewier et al., 2016; see also Noordewier & Breugelmans, 2013; Tomkins, 1984).

The temporal dynamics perspective on surprise is important as it can address disagreement in the literature on what surprise feels like. That is, sometimes surprise is depicted as a positive state (Fontaine, Scherer, Roesch, & Ellsworth, 2007; Valenzuela, Strebel, & Mellers, 2010), whereas others have argued it feels bad (Miceli & Castelfranchi, 2015; Noordewier et al., 2016; Noordewier & Breugelmans, 2013; Topolinski & Strack, 2015), or is without any particular valence (Mellers, Fincher, Drummond, & Bigony, 2013; Reisenzein & Meyer, 2009; Reisenzein, Horstmann, & Schützwohl, 2017; Reisenzein, Meyer, & Niepel, 2012; Russell, 1980). Importantly, studies that show that surprise is positive did not focus on the initial interruption after unexpectedness, but instead allowed participants to first make sense of the outcome. Participants for instance report being happy with a surprising gift (Valenzuela et al., 2010) or elated with an unexpected financial gain (e.g. Mellers, Schwartz, Ho, & Ritov, 1997).

So, the fact that people can eventually enjoy a positive surprise does not mean that the initial surprise reaction was positive. Indeed, from the point of view of cognitive consistency theories and personal control perspectives, surprise reflects inconsistency, disruption, and lack of structure. Because this conflicts with people's need for a predictable and coherent world, this may feel relatively negative (Abelson et al., 1968; Gawronski & Strack, 2012; Kay, Whitson, Gaucher, & Galinsky, 2009; Mendes, Blascovich, Hunter, Lickel, & Jost, 2007; Miceli & Castelfranchi, 2015; Noordewier et al., 2016; Noordewier & Breugelmans, 2013; Proulx, Inzlicht, & Harmon-Jones, 2012; Rutjens, van Harreveld, van der Pligt, Kreemers, & Noordewier, 2013; Topolinski & Strack, 2015).

While the theory distinguishes between initial and subsequent responses to surprising stimuli, empirical evidence is scarce. The current studies systematically tested the temporal unfolding of facial expressions in response to surprising stimuli. Facial expressions are particularly suitable to reveal the unfolding of responses because they can capture initial responses to a surprising stimulus as well as dynamic changes in responses over time (e.g. Noordewier et al., 2016; Noordewier & Breugelmans, 2013; see also Dukes, Clément, Audrin, & Mortillaro, 2017). We aimed to answer the following questions: Regarding the general valence of expression, does it take time to

show positive expressions after a positive surprise and are facial expressions after a positive and negative surprises initially similar? In addition, regarding specific facial action, what do facial expressions after a surprising event look like?

Unfolding of expressions

What do we know about the facial expression after a surprise? Regarding the unfolding of the valence of facial expressions after a surprise, a first study analysed how participants perceived expressions of people who were positively surprised in TV-shows (Noordewier & Breugelmans, 2013). Screenshots taken right after the surprise and subsequently at one-second intervals were evaluated in terms of feelings and type of situation the person in the picture was in. Results showed that faces were rated less positive in the first moments as compared to later; a pattern that was assumed to reflect the unfolding of responses from surprise to the appreciation of the outcome itself (Noordewier & Breugelmans, 2013).

In line with this, a facial electromyography study (fEMG; Reisenzein, Bördgen, Holtbernd, & Matz, 2006, study 7) showed that participants who were surprised with an unanticipated photograph of themselves, initially showed a slight increase of corrugator activity (i.e. AU4/brow lowering; which is also found in Topolinski & Strack, 2015; see also Schützwohl & Reisenzein, 2012), which was after 1–3 s followed by an increase in zygomaticus activity (i.e. AU12/smile). While in this study Reisenzein et al. aimed to test the occurrence of the surprise expression (raised eyebrows, eye-widening, jaw drop; Darwin, 1872/1999; Ekman, Friesen, & Hager, 2002) rather than the temporal dynamics of facial action per se, it supports the notion that initial responses to surprising stimuli differ from later responses.

These findings support the logic that people first respond to the unexpectedness of an event and only later, responses differentiate based on the valence of the event. This results in the prediction that in terms of general valence, it takes time to respond positively to a positive surprise. Responses to positive and negative surprises should therefore be initially similar, but after some time start to differentiate depending on the valence of the event.

Regarding specific facial action, predictions are somewhat more difficult to make. The facial EMG studies point to the possibility that people may show brow lowering after a surprise (Reisenzein

et al., 2006; Topolinski & Strack, 2015; see also Schützwohl & Reisenzein, 2012). In fact, brow lowering was also sometimes observed in the studies on the occurrence of the surprise expression (9/13% in Experiment 3/8; Reisenzein et al., 2006; 20% in Schützwohl & Reisenzein, 2012). This seems inconsistent with the “prototypical” surprise expression which involves raised eyebrows in addition to eye-widening, and a jaw drop. Previous research already showed, however, that this three-component surprise expression is in fact observed in only a small minority of surprised people (0–5% in Reisenzein et al., 2006; Reisenzein, 2000a; Schützwohl & Reisenzein, 2012). Instead, people generally show raising of the eyebrows only (on average 9% in Reisenzein et al., 2006; 9.5% in Reisenzein, 2000; and 25% in Schützwohl & Reisenzein, 2012).

Based on these findings, we thus might expect that surprised participants show brow raising and/or brow lowering. The brow raising is expected to occur in a minority of cases (9–25%; see above), whereas the proportion of brow lowering is harder to specify. When brow lowering was mentioned in frequency coding, it was observed in a minority of cases (9–20%; see above). Frequency of brow lowering is not reported in fEMG studies, however, as fEMG data represents averaged muscle action rather than an absolute rate of occurrence. Studies on the occurrence of the surprise expression also revealed other facial actions, such as smiling (7–96% in Reisenzein et al., 2006; 26–71% in Schützwohl & Reisenzein, 2012; 9–12% in Scherer, Zentner, & Stern, 2004) and in infant studies, freezing (i.e. facial stilling, Camras et al., 2002; Scherer et al., 2004) and signs of interest have been reported (Camras et al., 2002).

The current literature thus provides a mixed picture regarding specific facial action after a surprise. Yet, it seems reasonable to assume that smiles are correlated with the appreciation of an outcome itself which means that for positive surprises, they most likely only occur after some time (similar to Reisenzein et al., 2006, as discussed above). Most brow action is expected to occur before any of these smiles. Brow raising is related to the surprise expression and hence, the surprise phase. Similarly, brow lowering also fits this initial phase, as it has been related to sense-making concepts like orientation (Van Dillen, Harris, Van Dijk, & Rotteveel, 2015; Yartz & Hawk, 2002), error monitoring processes (Elkins-Brown, Saunders, & Inzlicht, 2016), mental effort (e.g. Van Boxtel & Jessurun, 1993), and negative affect (Cacioppo, Petty,

Losch, & Kim, 1986; Nohlen, Van Harreveld, Rotteveel, Barends, & Larsen, 2016; Topolinski & Strack, 2015; Topolinski, Likowski, Wyers, & Strack, 2009). Thus, if brow raising or brow lowering is observed, it most likely occurs before any smiling in the case of positive surprises. In the case of negative surprises, brow lowering might (continue to) appear after some time as a correlate of not appreciating the outcome.

In sum, and somewhat tentative, regarding specific facial action we predict that initial expressions are more likely to involve brow action, which may either involve brow raising in a minority of cases and/or brow lowering in an unknown proportion. The relative proportion of brow raising and brow lowering is also unknown and it is an open question whether brow raising and brow lowering will be observed alone or in combination with each other. Following possible brow action, later expressions are more likely to involve smiles in the case of positive surprise and possibly (continue to) involve brow lowering in the case of negative surprise. Similar to previous research (e.g. Reisenzein et al., 2006; Schützwohl & Reisenzein, 2012), we do not expect to find strong evidence for a surprise expression (i.e. a combination of brow raise, eye-widening, and jaw drop).

The current studies

To reveal the temporal unfolding of facial expressions in response to a surprising stimulus, we developed two repetition-change studies – a standardised and well-validated procedure to induce surprise (e.g. Camras et al., 2002; Meyer et al., 1997; Reisenzein et al., 2006). We tested our predictions using positive surprises (Experiments 1 and 2) as well as a negative surprise (Experiment 2) and recorded facial expressions using webcams. We used computerised coding to assess overall valence of expression and manual coding using FACS (Facial Action Coding System: Ekman et al., 2002) to assess specific facial actions.

Our hypotheses are as follows: In term of *valence of the expression* (measured with computerised coding), we predict that after a positive surprise, initial expressions are less positive than later expressions (H1). Next, after a positive and a negative surprise, expressions are initially similar (H2a) and only start to differentiate depending on the nature of the event after some time (H2b). In terms of *specific facial action* (measured with FACS), we predict that initial expressions are more likely to involve brow

raising (H3a) or lowering (H3b), while later expressions are in more likely to involve smiles in the case of positive surprises (H4) and brow lowering in the case of negative surprises (H5). In terms of *surprise expression*, we do not expect to find strong evidence for the “prototypical” expression, such that surprise expression after baseline measured with computerised coding does not increase, or only weakly, increases (H6a) and the combination of brow raise, eye-widening, and jaw drop measured with FACS is unlikely (less than 5%; H6b).

In the studies, we report all manipulations, all measures, and all data exclusions. In addition, we aimed for sample sizes of at least 50 per cell (as advised by Simmons, Nelson, & Simonsohn, 2013) and continued data collection within the given time available in the lab (approximately two weeks for Experiment 1 and one week for Experiment 2) to be able to account for possible data exclusion as a result of coding errors and participants not giving permission to use their material.

Experiment 1: a surprising puppy

In the first study, we tested our unfolding hypothesis by positively surprising participants with a puppy.

Method

A total of 71 participants (24 males, 47 females; $M_{\text{age}} = 22.32$, $SD_{\text{age}} = 4.87$) were assigned to a within participants design in which we compared facial expressions in response to neutral stimuli (baseline) and to a positively surprising target.

Procedure and materials

The study started with a cover story for using the webcam and to induce a social context. Participants were told that they would participate in a study on eye-movement and attention to pictures and in order to analyse their eye-movements, we would record them with a webcam.

We wanted to make the context somewhat more social than the more typical lab setting, where participants are in a lab cubicle on their own. A pilot test showed that participants were not very expressive in such individual settings and we reasoned that one explanation could be that it is not social enough (e.g. Fridlund, 1991). Therefore, we told participants that recent research suggested that there are reasons to believe that people perform better on

attention tasks when they do this with other people and that we were interested to test whether it is necessary to see the other person or not. We told them that they would be connected to another participant via the webcam, like on Skype. This story was most likely credible to participants, as in the two preceding, but unrelated, experiments in the experimental session they were also connected to other participants (in one experiment for real, in the other also as part of a cover story).

All participants were presented with a pre-recorded video of a confederate with the request to look at the other person and to connect with this person by for instance waving. The confederate waved and, on the footage, we saw participants doing so too, which leads us to believe that we created a credible social context. A picture (i.e. a still frame) of the confederate remained in the top right corner of the screen throughout the non-surprise part of the experiment.

After instructions, participants continued to the main part of the experiment. Surprise was induced using a repetition-change procedure. On a computer screen, participants were presented with a series of trials with sequential presentation of affectively neutral stimuli: buildings. Each trial presented four pictures of buildings (i.e. building-building-building-building) at one-second intervals. To engage participants in the trials, they were asked to indicate whether the last picture in each trial contained any green. On a keyboard, they could press either “a” or “l”, for yes and no, respectively. Participants were given one second to press the key. So, all elements in the trials took one second, which induces a certain rhythm and strengthens the expectancy about what would follow.

After four practice trials, fourteen experimental trials followed. The last trial was the critical surprise trial. In this trial, instead of presenting participants with the “did it contain green”-question, we unexpectedly showed them a gif-file of a puppy (i.e. graphic interchange format: multiple image frames are played in sequence, creating a moving picture), in which the puppy moved its head and paw towards the camera (see imgfave.com/view/1494654). The gif repeated three times, which took 9 s in total. After the surprise trial, the experiment automatically continued to some background questions. Participants were asked to indicate (translated from the original Dutch) “To what extent were you surprised by the puppy?” (from 1 = *not at all* to 7 = *extremely*) and “What did you think of the puppy?” (from 1 = *negative* to 7 =

positive). Then, they were asked to report their age and gender and whether they had participated before in a comparable study before (*yes/no*; we ran a pilot study a couple of months before this study). Finally, participants were fully debriefed and asked for permission to use their recorded footage (*yes/no*).

Results and discussion

The analyses consisted of different steps. First, we selected participants. Then, we checked our manipulation. Finally, after editing the footage, we tested our unfolding predictions by analysing the footage in two ways. First, the facial expressions were coded using Noldus' FaceReader (version 5; see Noldus.com/FaceReader). Next, the facial expressions were also coded manually using FACS.

Participant selection, target evaluation, and footage editing

We excluded 8 participants who did not give permission to use their footage and 2 who had participated before in a similar (pilot) study. Next, we excluded 8 participants who wore glasses (which may hinder classification in FaceReader; Noldus, 2012, p. 16) and 1 who produced a substantial amount of facial behaviour uncodeable by FaceReader (i.e. extreme yawning). We analysed the data of the remaining 52 participants (18 males, 34 females; $M_{\text{age}} = 21.83$, $SD_{\text{age}} = 4.79$). We first checked the ratings of the target. As expected, the target was rated as surprising ($M = 6.00$, $SD = 1.12$) and as positive ($M = 5.85$, $SD = 1.36$).

Next, we edited the videos such that they ran from 2 s before display of the surprising stimulus (baseline) until 8 s after. We did this based on event markers that were saved during the experiment: We saved the start and stop time of the experiment and we saved the time of critical trials. Based on the total duration of the video, we could then calculate for each participant when the surprising event had been shown. We first analysed the videos with FaceReader (computerised facial coding) and then coded it with FACS. For ease of presentation, we report the FACS results first, because these results have implications for the FaceReader results.

FACS

We included FACS coding to test specific facial action after a surprise (H3-6). This coding also allows us to assess the frequency with which the facial actions of

interest would occur after surprise, rather than the aggregated valence and surprise scores from FaceReader. Using Elan software (see <https://tla.mpi.nl/tools/tla-tools/elan/>), two licensed FACS coders coded the onset and offset of a subset of Action Units (AUs). The AUs typically associated with surprise were coded: inner/outer brow raise (AU1, AU2), eye-widening (AU5), and jaw drop (AU26). In addition, brow lowering (AU4) and AUs associated with happiness/smiling were coded: cheek raise and lip corner puller (AU6, AU12). Finally, AU0 was coded if none of the specific action units of interest were observed. When coding the AUs of interest, all FACS combination rules were taken into account, but because we did not expect systematic variation in other AUs, they were not coded.

One coder (first author) coded all videos and a second coder (not involved in the project) coded a subset of 15% of the videos for a reliability check. The AU-agreement index was .91 (i.e. total number of AUs agreed upon divided by total number of AUs coded by both coders). Next, we also checked the reliability of the timing of the AUs. Allowing a 0.2 s variation, this timing-agreement index was .93 (i.e. total number of onset and offset times agreed upon divided by total number of onset and offset times coded by both coders; note that this only includes the AUs the coders agreed upon). Disagreement was resolved through discussion and timing differences were averaged (except one large difference of 2.7 s, which we discussed and resolved). The final data set consisted of frequency, onset, offset, and duration for each coded AU.

To test our predictions that initial expressions are more likely to involve brow raising (H3a) or lowering (H3b), while later expressions are more likely to involve smiles (H4), and the "prototypical" surprise expression is unlikely (H6b), we computed the average onset times of the different AUs (no participant showed any of the coded AUs multiple times, so frequency and onset data are unaffected by that; for all descriptives, see Table 1). Given the relatively low occurrence rate of most of the AUs, we only report the frequency and average onset time; we do not perform additional statistical tests.

We found that those who raised their inner/outer brow ($N_{\text{AU1}} = 5$, $N_{\text{AU2}} = 4$) did this on average at second 5.01 ($SD = 2.93$) and 4.66 ($SD = 3.26$) after the surprising stimulus, respectively. Participants who widened their eyes ($N_{\text{AU5}} = 3$) did this on average at second 4.94 ($SD = 0.78$) after the surprising stimulus.

Table 1. Frequency, mean onset (SD), and duration (SD) of different Action Units (AU) observed after a positively surprising stimulus (Experiment 1).

	<i>N</i>	Onset	Duration
AU1 inner brow raise	5	5.01 (2.93)	2.46 (2.08)
AU2 outer brow raise	4	4.66 (3.26)	2.08 (2.18)
AU5 eye-widening	3	4.94 (0.78)	1.47 (1.94)
AU4 brow lowering	9	4.78 (1.60)	2.01 (2.21)
AU6 cheek raise	1	4.91 (–)	1.12 (–)
AU12 lip corner puller	22	4.61 (1.23)	4.27 (1.84)

Notes: AU26 (jaw drop) was never observed. Different AUs sometimes stem from the same participant. A total of $N = 6$ showed at least one of the prototypical surprise AUs (AU1, AU2, or AU5). Of these 6, $N = 1$ additionally showed AU12 and $N = 4$ additionally showed AU4. In addition, a total of $N = 22$ showed one of the enjoyment AUs (i.e. AU6 was observed in combination with AU12). Of these 22, $N = 1$ additionally showed surprise AUs (see above) and $N = 3$ additionally showed AU4.

Participants who lowered their brows ($N_{AU4} = 9$) did this on average at second 4.78 ($SD = 1.60$) after the surprising stimulus. The participant who raised his/her cheek ($N_{AU6} = 1$) did this at second 4.91 and participants who pulled their lip corner ($N_{AU12} = 22$) did this on average at second 4.61 ($SD = 1.23$) after the surprising stimulus. No participant showed a jaw drop (AU26) and 23 participants showed none of the AUs of interest.

We thus most often observed the lip corner puller (AU12) and when looking at the average onset time, it does not seem to occur later than action in the brows. Importantly, however, of those who showed brow lowering ($N = 9$) only 3 participants also showed lip corner puller, which was observed later than the brow lowering in 2 participants. In the 7 other brow lowering cases, we observed no facial action ($N = 2$), inner brow raise before ($N = 1$), or inner/outer brow raise after ($N = 2$; one with and one without eye-widening). In addition, lip corner puller ($N = 22$), was most often observed without other facial action ($N = 17$). In the other 5 lip corner puller cases, we observed inner/outer brow raise ($N = 1$) and brow lowering before ($N = 2$; see above); or brow lowering ($N = 1$) and cheek raise ($N = 1$) after. For an overview of all first vs. later AUs of participants who sequentially showed multiple AUs, see Table 2.

In sum, these data show no support for distinct brow action in the initial phase after surprise. It does show that after some time, a proportion of the

Table 2. Sequential AUs (Experiment 1).

First	Later
AU1	AU4
AU1/2	AU5
–	AU12
AU4	AU1/2
–	AU5
–	AU1/2 & AU5
–	AU12
–	AU12
AU12	AU4
–	AU6

Note: A total of $N = 10$ participants sequentially showed multiple AUs (i.e. AUs with different onset times). Each row is one participant.

participants smiled. Importantly, a subset of participants (44%) do not show FACS-action of interest. In the subsequent FaceReader analyses, we first describe analyses on the entire sample. Then, to understand the relationship between the FACS data and the FaceReader data, we also compare those who showed FACS-action with those who do not (see “Integrating FaceReader and FACS data” at the end of the Results section).

FaceReader

After uploading videos, FaceReader analyses facial expressions in terms of basic emotions (i.e. happiness, sadness, anger, surprise, fear, and disgust) and general valence (happiness minus negative emotions, excluding surprise). FaceReader first locates the face and then creates a face model based on 500 key points. The face is then compared to a database of 10,000 manually coded faces. The deviation of the face relative to the database is made and intensity of expressions is calculated. For each frame, FaceReader computes intensity scores for expressions of basic emotions (0 to 1) and valence (–1 to 1; for more information, see noldus.com/facereader; for validation see Den Uyl & Van Kuilenburg, 2005; Van Kuilenburg, Wiering, & Den Uyl, 2005; Lewinski, den Uyl, & Butler, 2014; for studies using FaceReader see e.g. Chentsova-Dutton & Tsai, 2010; Garcia-Burgos & Zamora, 2013).

The FaceReader data allowed us to compare the unfolding of responses within participants; comparing expressions before and after the surprise. We focused on two output measures: *valence* and *surprise*. FaceReader was set to analyse 25 frames per second and to calibrate each participant individually, filtering out person-specific biases (e.g. looking angry or happy by nature). We reduced this large data set (i.e. 250

data points per participant for both valence and surprise) by computing a baseline score for valence and surprise by averaging the scores of the two seconds before the surprising target (mean of 50 frames; seconds $-2-0$). In addition, we computed an average intensity score on valence and surprise for each 2-seconds interval after the surprise (seconds 0–2: time 1, seconds 2–4: time 2, seconds 4–6: time 3, seconds 6–8: time 4)¹ for each participant. Note that when FaceReader cannot find the face (e.g. due to a hand in front of the face or strong upward or downward movement of the participant), it does not produce any data. In the current study, this was the case for 110 frames, which is 0.008% of the total of 13,000 frames (i.e. 52 participants times 250 frames). The average intensity scores exclude these missing data points.

Next, we checked for outliers. Values that were 3.3 standard deviations above or below the mean were recoded as 1% higher than the next-highest non-outlier value (i.e. this lowers the impact of extreme values, while preserving the distribution of the data; see Seery, Leo, Lupien, Kondrak, & Almonte, 2013).² The final data consisted of 5 data points for each participant for both valence and surprise, resulting in the within-subjects factor Time (i.e. baseline and times 1–4 after the surprising target). On these data, we ran repeated measures ANOVAs to test the prediction that initial expressions are less positive than later expressions (H1), while surprise was not expected to increase (strongly) after baseline (H6a). We first

checked for the effect of Time and when a statistically significant effect was found, we conducted within-subjects contrasts comparing the expressions after surprise with the first data point of the baseline.

Valence. The repeated measures ANOVA showed a marginal effect of Time on valence of expressions, Wilks' Lambda = .85, $F(4,48) = 2.19$, $p = .084$, $\eta_p^2 = .15$ (see Figure 1(a)). Comparing the valence of expressions relative to baseline with within-subjects contrasts, we found that expressions were more positive at time 4, $F(1,51) = 5.89$, $p = .019$, $\eta_p^2 = .10$, whereas they did not differ from baseline at times 1–3, F_s between .21 and 2.15, p_s between .149 and .650, $\eta_p^2_s$ between .004 and .04.

Surprise. The repeated measures ANOVA showed no effect of Time on the surprise expression, Wilks' Lambda = .90, $F(4,48) = 1.41$, $p = .244$, $\eta_p^2 = .11$ (means ranged between .07 and .12; SD_s between .11 and .20).

Integrating FaceReader and FACS data

Taken together, the FACS data show that initially, there is limited facial action. The FaceReader data showed that expressions at time 4 were more positive than at baseline, although the overall effect of time was only marginal. Importantly, however, the FACS data showed that only a subset of participants showed facial action of interest. To further understand the relation between the FACS data and the FaceReader data, we created a variable indicating whether

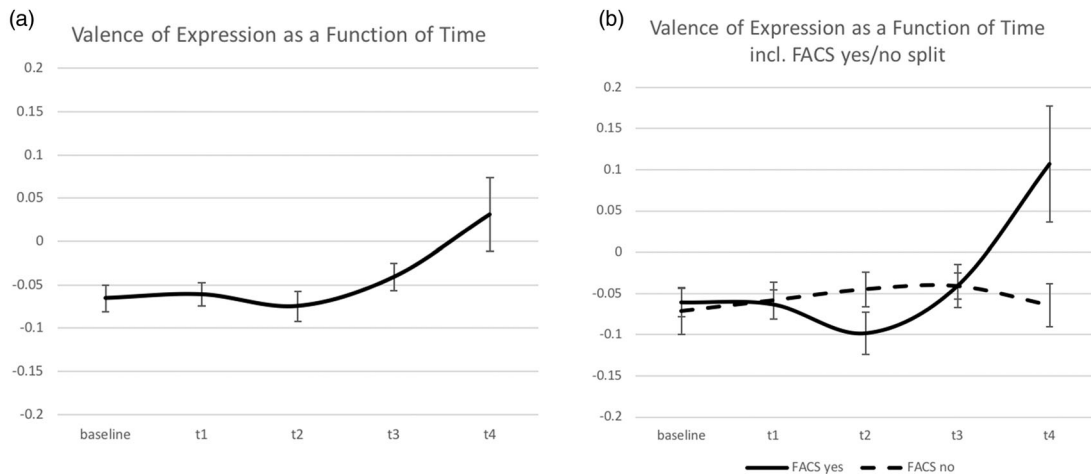


Figure 1. (a) Valence of facial expression in response to a surprising puppy as a function of Time (Experiment 1). The baseline is a 2-seconds interval before the surprise and times 1–4 are 2-seconds intervals after the surprise. Error bars indicate ± 1 SE. (b) Valence of facial expression in response to a surprising puppy as a function of Time and FACS-action (yes/no; Experiment 1). Error bars indicate ± 1 SE.

participants showed FACS-action (yes/no) and re-analysed the valence FaceReader data. This confirmed that those who show facial action in the FACS coding ($N=29$; 56%) show an effect of Time on valence, Wilks' Lambda = .66, $F(4,25) = 3.27$, $p = .028$, $\eta_p^2 = .34$, whereas those who do not show facial action in the FACS coding ($N=23$; 44%) also show no effect of Time on valence ($F < 1$; see Figure 1(b)). Subsequent baseline comparison with within-subjects contrasts showed that expressions in the FACS-action subgroup were more positive at time 4, $F(1,28) = 6.28$, $p = .018$, $\eta_p^2 = .18$, and marginally more negative at time 2, $F(1,28) = 4.18$, $p = .051$, $\eta_p^2 = .13$; other effects $ps > .42$).

These results show that the FaceReader findings are explained by the subset of facially expressive participants and that in this selection, there is marginal support for negative unfolding at time 2 and no support in both FaceReader and FACS data for the "prototypical" surprise expression. The increase in positivity in the FaceReader data seems to be characterised by an increase in "smiles" (AU12). Note, though, that the timing of AU12 does not line up perfectly with the timing of the increase in positivity in the FaceReader data (i.e. AU12 starts on average at second 4.63, while the increase in positivity happens at time 4, which is between seconds 6 and 8). A plausible explanation for this difference is that the FACS data refer to the onset of facial action, while the FaceReader data refer to the average intensity of the expression. Therefore, AU12 might start just after time 2, but may reach its apex at a later stage, affecting the aggregated intensity scores more. We will further discuss this in the General Discussion.

Experiment 2: a surprising person

Experiment 2 tests the unfolding logic by surprising people in a person-perception setting (see also Proulx, Slegers, & Tritt, 2017). We assumed that this setting is more social and self-relevant than the buildings and the unexpected puppy in Experiment 1, which might intensify responses (e.g. Fridlund, 1991; Jakobs, Manstead, & Fischer, 1999, 2001; Scherer, 2001; Soussignan et al., 2013, for a similar argument in the context of surprise, see Reisenzein et al., 2006; Schützwohl & Reisenzein, 2012). In this study we also included a negative surprise. We again used a repetition-change method and showed participants a series of neutral faces, followed by a face that deviated from the preceding faces and thus was unexpected.

This was either a positive or a negative face, which allows us to compare initial and later responses to a positive vs. a negative target.

Method

We randomly assigned 128 participants (59 males, 69 females; $M_{\text{age}} = 21.20$, $SD_{\text{age}} = 2.25$) to a positive versus negative surprise condition. The study was presented as a test of factors driving first impressions of unknown others. To this end, participants were asked to evaluate pictures of 20 faces, with equal numbers of males and females, all showing a neutral expression. Pictures were selected from the Radboud Faces Database (RAFD; Langner et al., 2010). Each neutral face was shown 5 s after which the question "What is your impression of this person?" appeared on the screen. Participants could answer "positive" or "negative" with respectively green and blue response buttons (i.e. the left and right ctrl buttons on a keyboard were covered with green and blue stickers).

After 20 trials the critical surprise trial showed either a positive or a negative target face for 8 s. The positive target was a woman with a pig nose mask showing a funny face. The negative target was a man with wounds on his face. Both targets did not show any positive or a negative expression, to prevent that participants would mimic the face. After the critical trial, the programme automatically continued to background questions. Participants were asked to report to what extent they were surprised by the target (from 1 = *not at all*, to 7 = *extremely*), to evaluate the target (from 1 = *negative* to 7 = *positive*), and to report their age and gender. Finally, they were fully debriefed and asked for permission to use their footage (yes/no).

Results and discussion

The analyses were done following the same steps as in Experiment 1.

Participant selection and footage editing

We excluded participants who did not give us permission to use the footage ($N=5$), who wore glasses ($N=7$) or because of coding errors (i.e. $N=3$; video could not open and $N=1$; only half of the face was recorded). We report analyses of the remaining 112 participants (53 males, 59 females, $M_{\text{age}} = 21.14$, $SD_{\text{age}} = 2.27$).

First, we checked the ratings of the target. As expected, the positive target was rated more positive ($M = 5.70$, $SD = 1.69$) than the negative target ($M = 2.60$, $SD = 1.26$), $t(110) = 10.89$, $p < .001$, $d = 2.08$. Yet, the positive target was rated as equally surprising ($M = 5.72$, $SD = 1.38$) as the negative target ($M = 6.02$, $SD = 1.18$), $t(110) = -1.24$, $p = .22$, $d = -.23$. So, based on this we conclude that our stimuli represented a positive surprise in the positive surprise condition and a negative surprise in the negative surprise condition.

Next, we edited the videos in the same way as in Experiment 1, such that they showed participants 2 s before the surprise (baseline) until 8 s after the surprise. This footage was first coded with FaceReader and then coded with FACS.

FACS

The videos were coded, blind to condition, using FACS in the same way as in Experiment 1. The AU-agreement index was .86 and the timing-agreement index was .79. Disagreement was resolved through discussion and timing differences were averaged.

We computed the average onset times of the different AUs within both the positive and the negative surprise condition to test the predictions that initial expressions are more likely to involve brow raising (H3a) or lowering (H3b), while later expressions are more likely to involve smiles in the case of positive surprises (H4) and brow lowering in the case of

negative surprises (H5). Moreover, the “prototypical” surprise expression was predicted to be unlikely (H6b). In the positive target condition, there were four cases where the same action unit was observed multiple times in one participant. This did not happen in the negative target condition. To avoid confounding frequency and mean onset and duration, we report the means for the first occurrence of the AU only. All other means, including the means with these double AUs, can be found in Table 3. Note that while we statistically compare the AU-frequencies between the positive and the negative surprise condition, we only report the AU-frequencies and average onset time within each condition. Similar to Experiment 1, we did not conduct additional statistical tests, given the relatively low occurrence rate of most of the AUs.

Within the *positive surprise condition*, participants who raised their inner/outer brow ($N_{AU1} = 3$, $N_{AU2} = 3$) did this on average at second 2.44 ($SD = 1.16$) and 2.84 ($SD = 1.77$) after the surprising stimulus, respectively. The participant who widened his/her eyes ($N_{AU5} = 1$) did this on second 0.90 after the surprising stimulus. The participant who lowered his/her brows ($N_{AU4} = 1$) did this at second 0.71 after the surprising stimulus. Participants who raised their cheek ($N_{AU6} = 5$) and pulled their lip corners ($N_{AU12} = 35$) did this on average at second 3.47 ($SD = 0.71$) and 3.78 ($SD = 1.01$) after the surprising stimulus, respectively. No participant showed a jaw drop

Table 3. Frequency, mean onset (SD), and duration (SD) of different Action Units (AU) observed after a positively vs. a negatively surprising target (Experiment 2).

	Positive target			Positive target (first occurrence AU only)			Negative target		
	<i>N</i>	Onset	Duration	<i>N</i>	Onset	Duration	<i>N</i>	Onset	Duration
AU1 inner brow raise	4	3.08 (1.60)	3.18 (3.02)	3	2.44 (1.16)	4.04 (3.03)	6	3.86 (2.27)	2.50 (1.67)
AU2 outer brow raise	4	3.38 (1.81)	2.8 (2.90)	3	2.84 (1.77)	3.64 (3.02)	7	3.54 (1.69)	2.25 (1.77)
AU5 eye-widening	1	0.90 (-)	0.20 (-)	1	0.90 (-)	0.20 (-)	3	3.09 (0.88)	0.75 (0.95)
AU4 brow lowering	2	3.81 (4.39)	0.70 (0.56)	1	0.71 (-)	0.31 (-)	14	3.47 (1.26)	2.73 (1.55)
AU6 cheek raise	5	3.47 (0.71)	3.84 (1.41)	5	3.47 (0.71)	3.84 (1.41)	1	3.86 (-)	1.85 (-)
AU12 lip corner puller	36	3.88 (1.15)	3.25 (1.68)	35	3.78 (1.01)	3.32 (1.64)	8	4.73 (2.49)	1.84 (1.50)

Notes: AU26 (jaw drop) was never observed. In four cases, AUs were observed multiple times in one participant. This was all in the positive target condition. The second positive target column shows the frequency and means based on first occurrence of AU only. Different AUs sometimes stem from the same participant: A total of $N = 14$ showed at least one of the prototypical surprise AUs (AU1, AU2, or AU5; 11 in the negative surprise condition). Of these 14, $N = 5$ additionally showed enjoyment AUs (AU6 or AU12; 2 in the negative surprise condition) and $N = 4$ additionally showed AU4 (all in the negative surprise condition). A total of $N = 43$ showed one of the enjoyment AUs (8 in negative surprise condition; AU6 was always observed in combination with AU12). Of these 43, $N = 5$ showed surprise AUs (see above) and $N = 3$ additionally showed AU4 (2 in the negative surprise condition).

(AU26) and 25 participants showed none of the AUs of interest.

Within the *negative surprise condition*, participants who raised their inner/outer brow ($N_{AU1} = 6$, $N_{AU2} = 7$) did this on average at second 3.86 ($SD = 2.27$) and 3.54 ($SD = 1.69$) after the surprising stimulus, respectively. Participants who widened their eyes ($N_{AU5} = 3$) did this on average at second 3.09 ($SD = 0.88$) after the surprising stimulus. Participants who lowered their brows ($N_{AU4} = 14$) did this on average at second 3.47 ($SD = 1.26$) after the surprising stimulus. The participant who raised his/her cheek ($N_{AU6} = 1$) did this at second 3.86 and participants who pulled their lip corners ($N_{AU12} = 8$) did this on average on 4.73 ($SD = 2.49$) after the surprising stimulus, respectively. No participant showed a jaw drop (AU26) and 27 participants showed none of the AUs of interest.

When we compare the frequency of AUs in the positive and the negative target condition (analyses with first AU only; results including the four double AUs were similar), we see that in the positive target condition AU12 is more often observed, $\chi^2(1, N = 139) = 20.82$, $p < .001$, whereas in the negative target condition AU4 is more often observed, $\chi^2(1, N = 139) = 14.18$, $p < .001$. For the other AUs, no difference between conditions were observed ($ps > .12$).

In addition, when we look at participants who sequentially showed multiple AUs ($N = 13$, for an overview of all first vs. later AUs, see Table 4), we see that AU12 follows various AUs ($N = 4$ in the positive surprise condition, $N = 3$ in the negative surprise condition), whereas when AU12 is the first expression, it is only followed by AU6 ($N = 3$ in the positive surprise condition, $N = 1$ in the negative surprise condition). While these frequencies are too low for drawing

strong conclusions, these cases support the notion that smiling follows other facial action rather than the other way around.

In sum, like in Experiment 1, these data show no support for distinct brow action after surprise. Overall, in the positive target condition more participants smiled and in the negative target conditions, more participants showed brow lowering.

FaceReader

FaceReader was set to analyse 30 frames per second and to calibrate each participant individually, filtering out person-specific biases. We again computed an average intensity score on *valence* and *surprise* for 2-seconds intervals (note that in this study we used 30 and not 25 frames per second like in Experiment 1, as this made it easier to compute means with equal number of frames for 0.5-second interval analyses; see below). There were missing data for 38 frames, which is 0.001% of the total of 33,600 frames (i.e. 112 participants times 300 frames). After restructuring and checking extreme values (see Experiment 1), the final data consisted of 5 data points for each participant on valence and surprise (i.e. averaged baseline and times 1–4 at 2-seconds intervals after the surprising target). On these data, we ran repeated measures ANOVAs (see Figure 2(a)), followed by within-subjects contrasts (Time) and between condition comparisons (Target) to test the prediction that after a positive and a negative surprise, expressions are initially similar (H2a) and only start to differentiate depending on the nature of the event after some time (H2b). Moreover, we tested whether surprise increases after baseline (H6a).

Valence. The repeated measures ANOVA showed a Time X Target interaction on valence of expression, Wilks' Lambda = .90, $F(4,107) = 3.03$, $p = .021$, $\eta_p^2 = .10$ (see Figure 2(a)). Furthermore, there was no main effect of Time, Wilks' Lambda = .94, $F(4,107) = 1.60$, $p = .179$, $\eta_p^2 = .06$, and a marginal main effect of Target, $F(1,110) = 3.58$, $p = .061$, $\eta_p^2 = .03$. To interpret the interaction, we separately compared the effect of Time within the positive and negative target condition.

Within the positive target condition, there was a marginal main effect of Time, Wilks' Lambda = .85, $F(4,56) = 2.48$, $p = .055$, $\eta_p^2 = .15$. Simple contrasts showed that expressions were more positive relative to baseline at times 2–4: F_s between 5.66 and 9.46,

Table 4. Sequence of multiple AUs (Experiment 2).

	First	Later
Positive surprise	AU1	AU2 & AU12
	AU1/2	AU12*
	AU4	AU4 & AU12
	AU5	AU1/2* & AU12
	AU12 (3x)	AU6 (3x)
Negative surprise	AU2/12	AU6
	AU4	AU1
	–	AU12
	–	AU12
	AU5	AU4
–	AU12	

Notes: A total of $N = 13$ participants sequentially showed multiple AUs (i.e. AUs with different onset times). Each row is one participant, unless specified differently with frequencies between brackets. AUs with * occurred twice in the same participant.

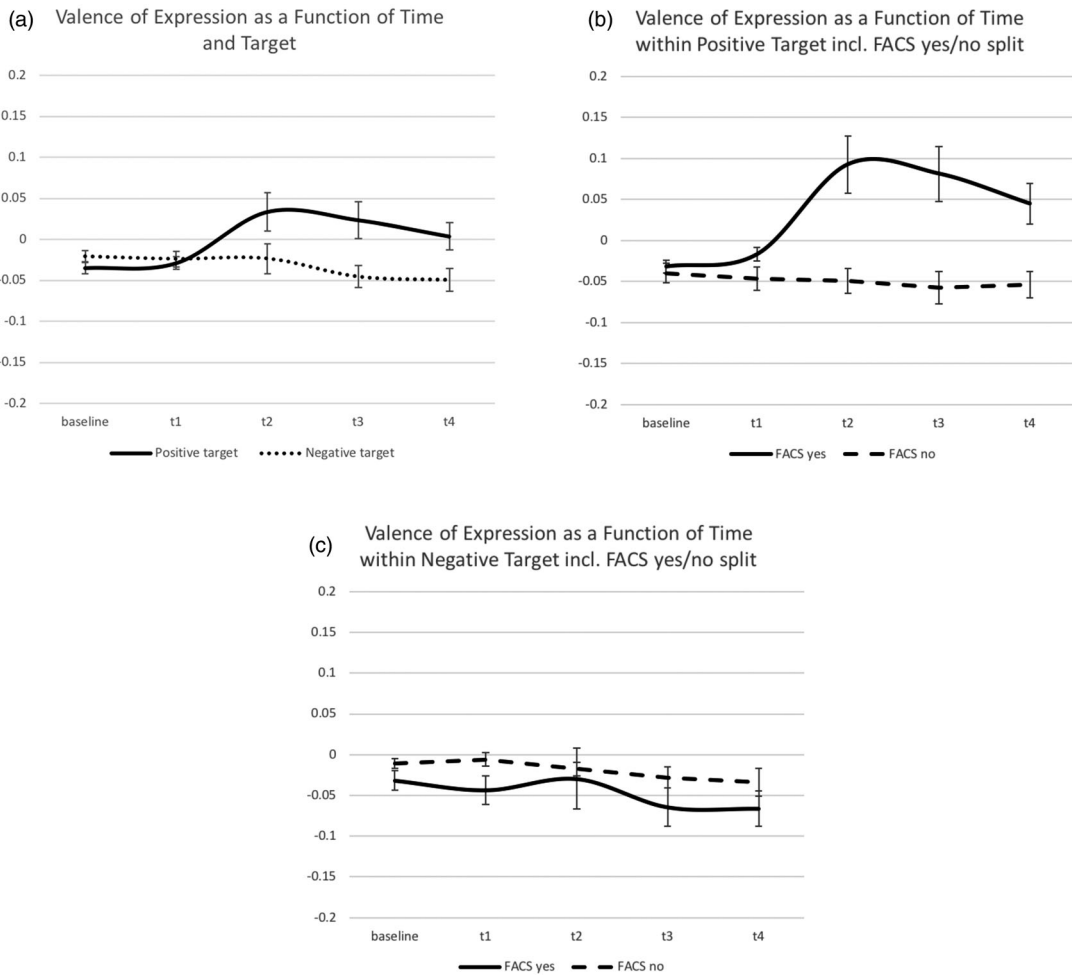


Figure 2. (a) Valence of facial expression as a function of Target (positive vs. negative) and Time (Experiment 2). The baseline is a 2-second interval before the surprise and times 1–4 are 2-second intervals after the surprise. Error bars indicate ± 1 SE. (b) Valence of facial expression within the Positive Target condition as a function of Time and FACS action (yes/no; Experiment 2). Error bars indicate ± 1 SE. (c) Valence of facial expression within the Negative Target condition as a function of Time and FACS action (yes/no; Experiment 2). Error bars indicate ± 1 SE.

p s between .003 and .021, η_p^2 s between .09 and .14, but not at time 1 ($F < 1$). Within the negative target condition, there was no main effect of Time, Wilks' Lambda = .89, $F(4,48) = 1.54$, $p = .206$, $\eta_p^2 = .11$. Simple contrasts, however, showed that expressions were more negative relative to baseline at times 3 and 4, F s = 4.51/5.34, p s = .039/.025, η_p^2 s = .08/.10 (times 1 and 2 compared to baseline, $F < 1$).

Next, we compared the valence of expressions between the two target conditions with independent samples t -tests. At baseline and time 1, conditions did not differ ($p > .13$), whereas the valence of expressions were marginally more positive in the positive vs. negative target condition at time 2, $t(110) = 1.89$, $p = .061$,

$d = .36$ and significantly more positive at times 3/4, t s = 2.51/2.39, p s = .014/.018, d s = .49/.48

Thus, facial expressions were initially similar in the positive and negative target condition. Over time, they unfolded to more positive expressions in the positive target condition and there is some indication that they unfolded to negative expressions in the negative target condition. Interestingly, the unfolding seemed to occur faster than in Experiment 1. We will discuss this in more detail in the General Discussion.

Surprise. No effects were observed on the surprise expression (all p s $> .129$; all means ranged between .03 and .06).

Integrating FaceReader and FACS data

Taken together, and similar to Experiment 1, the FACS data show that initially there is limited facial action. The FaceReader data show that immediate responses to a positive or a negative surprise do not differ, while with time, the expressions in the positive target condition become more positive and there is some indication that expressions in the negative target condition become more negative. Importantly, like in Experiment 1, we see that only a subset of participants showed facial action of interest ($N = 60$; 54%). Therefore, using the same yes/no FACS-action selection as in Experiment 1, we re-analysed the FaceReader data.

The group with facial action in the FACS coding ($N = 60$, 54%) showed a Time X Target interaction, Wilks' Lambda = .80, $F(4,55) = 3.35$, $p = .016$, $\eta_p^2 = .20$ (see Figure 2(b,c)), a marginal main effect of Time, Wilks' Lambda = .87, $F(4,55) = 2.09$, $p = .094$, $\eta_p^2 = .13$, and a main effect of Target, $F(1,58) = 9.07$, $p = .004$, $\eta_p^2 = .14$. Contrary, the group without facial action in the FACS coding ($N = 52$, 46%), showed no Time X Target interaction and no main effect of Time ($F_s < 1$). We did see a marginal main effect of Target, $F(1,50) = 4.01$, $p = .051$, $\eta_p^2 = .07$.

Subsequent analyses of expressions in the FACS-action subgroup in the positive target condition showed a main effect of Time, Wilks' Lambda = .67, $F(4,31) = 3.75$, $p = .013$, $\eta_p^2 = .33$. Compared to baseline, expressions were more positive at times 1–4, F_s between 4.32 and 13.66, p_s between .001 and .045, η_p^2 s between .11 and .29. When we look more detailed at the differences in time 1 with 0.5 s intervals, we see that at second 1.5 the expressions are more positive than baseline, $F(1,34) = 10.53$, $p = .003$, $\eta_p^2 = .24$, and marginally more positive at second 1, $F(1,34) = 3.52$, $p = .069$, $\eta_p^2 = .09$. Between seconds 0 and 1, expressions do not differ as compared to baseline, $F_s < 1$. Next, within the negative target condition, there was no main effect of Time ($F < 1$), but time 4 was marginally lower than baseline ($p = .087$; other effects $p_s > .13$).

Finally, we also checked the differences between the positive and negative target conditions, with this yes/no FACS-action split. This showed that the expressions of those who showed FACS-action were more positive in the positive vs. negative target condition at times 2–4 ($p_s < .023$), but not at baseline and time 1 ($p_s > .17$). Expressions of those without FACS-action were (marginally) more negative in the positive vs. negative target condition at baseline and

times 1–2 (p_s between .018 and .074). We do not know how to explain these differences.

All in all, these results show, in line with Experiment 1, that the FaceReader findings are explained by a subset of expressive participants. In this subset, we see an increase in positivity of expressions in the positive target condition, which seems to be explained by "smiling"/AU12. Any marginal decrease in positivity in the negative surprise condition can probably be explained by brow lowering/AU4. As such, the data show limited facial action in the initial phase, whereas with time, expressions to a positive target become more positive in a subset of the participants.

General discussion

Responses to surprising events are dynamic (Meyer et al., 1991, 1997; Noordewier et al., 2016; Noordewier & Breugelmans, 2013). Initially, people are in an interrupted and surprised state due to the unexpectedness of the surprising event, whereas later, after sense-making, their responses can incorporate the valence of the surprising outcome itself (see also Meyer et al., 1997; Noordewier et al., 2016; Noordewier & Breugelmans, 2013). To study surprise and distinguish it from the state that follows it, we tested the temporal unfolding of facial expressions in response to a surprising stimulus. We conducted two repetition-change studies and analysed the general valence of facial expression using computerised facial coding and specific facial action using FACS.

For general valence, the computerised coding showed that initial expressions after positive surprises are less positive than later expressions (supporting H1). Moreover, expressions after a positive and negative surprise are initially similar (supporting H2a) and only start to differentiate after some time depending on the valence of the event (particularly in the positive surprise condition, supporting H2b). Importantly, however, these findings are particularly observed within the subset of facially expressive participants (i.e. among those who also showed facial action in the FACS coding).

In terms of specific facial action, the FACS data showed limited facial action in the initial phase and as such, no systematic brow raising or lowering was found (no support for H3a/b). Moreover, the increase in positive expressions in the positive target conditions seemed to be best explained by an increase in smiles (Experiments 1 and 2), while in the negative

target condition brow lowering could underlie the more negative expressions (Experiment 2). These smiling and brow lowering findings are in line with H4 and H5, but additional data are needed to confirm the systematic timing of these facial actions. Finally, like previous research (e.g. Reisenzein et al., 2006; Schützwohl & Reisenzein, 2012), we do not find evidence for a prototypical surprise expression (i.e. no increase in surprise expression after baseline, measured with computerised coding; no combination of brow raise, eye-widening, and jaw drop, measured with FACS; supporting H6a/b).

Interestingly, when comparing the two studies in terms of speed of unfolding, the expressions seemed to unfold faster in the study with the surprising faces than the study with the surprising puppy. When we compare the increase of the positivity in the facial expression (see Figures 1 and 2 and simple contrast results) we see that in the positive surprise condition of Experiment 2 (surprising faces) the expression is already more positive than baseline at time 2 (i.e. between seconds 2 and 4) and marginally more positive at second 1 in subsequent analyses of the subset of facially expressive participants, whereas in Experiment 1 (surprising puppy), this difference occurs at time 4 (i.e. between seconds 6 and 8). A plausible explanation may be found in the relation between expectancy and surprise. The surprising puppy in Experiment 1 was categorically different from the preceding repetition trials (buildings), whereas the surprising positive/negative faces in Experiment 2 were categorically similar to the preceding repetition trials (neutral faces). Categorical similarity of surprise to the preceding context may make the surprising event easier to categorise, which facilitates sense-making and thus, faster responses to the actual meaning of the target. Moreover, faces are probably more self-relevant to participants than a puppy, which could have contributed further to faster unfolding. Another explanation is that in Experiment 1, we used a video clip, whereas in Experiment 2, a picture was used. Some of the participants in Experiment 1 could have been waiting for the surprising event to end (i.e. the moving puppy). Note, though, that the puppy was still moving when the expression became more positive than baseline, so this waiting notion can only partially explain the difference.

Furthermore, the FACS results warrant some discussion. We included FACS coding to test whether specific AUs would occur immediately after surprise (i.e. the presence of brow raising or lowering) versus after

people had some time to make sense of the outcome (i.e. zygomaticus action in the case of a positive surprise; brow lowering in the case of a negative surprise). In addition, we aimed to assess the frequency with which the specific facial actions of interest would occur. Previous research already showed people might initially lower (Topolinski & Strack, 2015) or raise their brows (Reisenzein et al., 2006), and that the “prototypical” surprise expression is rare (raised eyebrows, eye-widening, and jaw drop; Reisenzein et al., 2006). Contrary to predictions, we did not find evidence for systematic brow lowering or raising after a surprising event. In fact, we find only limited facial action in the initial phase after the surprise.

One explanation for this limited facial action in the initial phase is that overall, it is a challenge to get expressive participants with stimuli being presented on a computer. Other more intense and/or social settings might produce more brow action (e.g. Jakobs et al., 1999, 2001; Scherer, 2001; Soussignan et al., 2013; see also Schützwohl & Reisenzein, 2012). Another possibility is that brow action is only in specific situations associated with surprise, such as brow lowering when people need to (extra) mental effort to deal with the surprise (e.g. see Van Boxtel & Jessurun, 1993). Finally, the limited facial action could also point to another possible response. In infant studies, surprise has been connected to freezing (Camras et al., 2002; Scherer et al., 2004), which is a passive, defensive, response to a stressful event. Freezing has been characterised by reduced body motion and physiological changes like a reduced heart rate (Roelofs, Hagens, & Stins, 2010). Importantly, infant studies found support for facial sobering in response to surprise, which was defined as the “sudden cessation of any facial movement” (Camras et al., 2002 p. 183). In future studies, we aim to test whether surprise-induced freezing can be observed in adults as well.

Next, it is important to focus on a more general question regarding the relationship between the current FACS and FaceReader data. An important difference between the two types of measures is that FaceReader provides aggregated intensity scores, while FACS provides absolute data on the presence of certain facial action (with additionally also an intensity score, which was omitted here). This means that, similar to other aggregated facial expression measures like fEMG, it is possible that an overall difference in expression as a function of experimental conditions is based on the expressions of a subset of participants. This is supported by the current results,

where overall valence effects in the FaceReader is explained by a subset of expressive participants and the increase in positivity is most likely based on those participants who showed “smiling” activity (AU12/zygomaticus). Similarly, the fEMG brow lowering effects that occurred as a function of surprise (Reisenzein et al., 2006; Topolinski & Strack, 2015) might be the product of corrugator activity in only a subset of participants. This difference in data type is key for the choice of coding in research. Aggregated measures are suitable for research questions focused at general responses to stimuli (e.g. do people show more positive expressions after A than B?), whereas FACS is more suited for research questions directed at the relative prevalence of certain expressions (e.g. does brow lowering systematically occur after an unexpected event?). In addition, future research might also make further comparisons between FACS coding and computerised coding (e.g. in different situations and with additional AUs than the subset we coded here). With more research comparing manual coding to automated coding, we can learn more about the advantages and disadvantages of the different coding systems.

Finally, it is important to note that while we use facial expressions to show that initial and later responses after a surprise differ, we do not claim that these expressions can be directly translated to mental processes that underlie the expressions. Also, the direct link between expressions and feelings has been debated (e.g. Fernández-Dols & Crivelli, 2013; Reisenzein, Studtmann, & Horstmann, 2013; and for a broader perspective, see Lindquist, Siegel, Quigley, & Feldman Barrett, 2013; in reply to Lench, Flores, & Bench, 2011), which means that strong conclusions about how people feel after a surprise would require additional measures besides facial expression.

Taken together, our findings partly support the notion that responses to unexpected events unfold from an initial state that is characterised by interruption and surprise to later responses that incorporate the valence of the actual event. While these findings are found in the subset of facially expressive participants, they are important in the context of a broader question about the valence of surprise. Initial and later responses to surprising stimuli are different and should not be confused with each other to determine what surprise feels like (Noordewier et al., 2016). The current studies do not show evidence for a negative valence of surprise, but they do show that *if* positive responses are shown, they only occur after some

time. An implication of this unfolding of responses is that to study the subjective experience of surprise, one should focus on surprise “while it happens” rather than when people already had the opportunity to make sense of the event. Only then, we can distinguish actual surprise from the affective states that follow afterwards.

Notes

1. We chose to use 2-seconds intervals rather than smaller intervals to limit the number of comparisons and to avoid severe sphericity violations (i.e., more likely with more time-points).
2. We only used this procedure when extreme values affected the pattern of results. In both studies, this was not the case for the surprise values. For valence in Experiment 1, we recoded four values of three participants (all negative outliers; 0.02% of total). Without this recoding, the effect of Time is significant at $p = .050$ (rather than $p = 0.84$). For valence in Experiment 2, we recoded 13 values of nine participants (four negative, nine positive outliers; 0.02% of the total). Without this recoding, the main effect of Time within the positive target is $p = .050$ (rather than $p = 0.55$). In addition, within the positive target condition and the FACS-yes selection, the contrast between baseline and time 1 was marginal at $p = .063$ (rather than $p = .045$) and the difference between baseline and second 1 in the 0.5 seconds comparison is not significant at $p = .105$ (rather than $p = .069$).

Acknowledgements

Thanks to Xia Fang for her help with FACS coding. Thanks to Anne-Linn Beekhof and Frederique Arntz for their help with a previous version of the expression coding. Thanks to Carlo Konings for his help with programming the studies and converting the FaceReader data. Thanks to Suzanne Kuiper for her help with editing the videos. Thanks to Peter Lewinski for his advice on FaceReader.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

- Abelson, R. P., Aronson, E., McGuire, W., Newcomb, T., Rosenberg, M., & Tannenbaum, P. (1968). *Theories of cognitive consistency: A sourcebook*. Chicago: Rand McNally.
- Cacioppo, J. T., Petty, R. E., Losch, M. E., & Kim, H. S. (1986). Electromyographic activity over facial muscle regions can differentiate the valence and intensity of affective reactions. *Journal of Personality and Social Psychology, 50*, 260–268.
- Camras, L. A., Meng, Z., Ujiie, T., Dharamsi, S., Miyake, K., Oster, H., ... Campos, J. (2002). Observing emotion in infants: Facial

- expression, body behavior, and rater judgments of responses to an expectancy violating event. *Emotion*, 2, 179–193.
- Chentsova-Dutton, Y. E., & Tsai, J. L. (2010). Self-focused attention and emotional reactivity: The role of culture. *Journal of Personality and Social Psychology*, 98, 507–519.
- Darwin, C. (1872/1999). *The expression of the emotions in man and animals* (P. Ekman, Ed.). London: Fontana Press.
- Den Uyl, M., & Van Kuilenburg, H. (2005). The FaceReader: Online facial expression recognition. In L. P. J. J. Noldus, F. Grieco, L. W. S. Loijens, & P. H. Zimmerman (Eds.), *Proceedings of measuring behavior 2005, 5th international conference on methods and techniques in behavioral research* (pp. 589–590). Wageningen: Noldus Information Technology.
- Dukes, D., Clément, F., Audrin, C., & Mortillaro, M. (2017). Looking beyond the static face in emotion recognition: The informative case of interest. *Visual Cognition*, 25, 575–588.
- Ekman, P. (2003). *Emotions revealed. Understanding faces and feelings*. London: Weidenfeld & Nicolson.
- Ekman, P., Friesen, W. V., & Hager, J. C. (Eds.). (2002). *Facial action coding system*. Salt Lake City, UT: Research Nexus, Network Research Information.
- Elkins-Brown, N., Saunders, B., & Inzlicht, M. (2016). Error-related electromyographic activity over the corrugator supercilii is associated with neural performance monitoring. *Psychophysiology*, 53, 159–170.
- Fernández-Dols, J.-M., & Crivelli, C. (2013). Emotion and expression: Naturalistic studies. *Emotion Review*, 5, 24–29.
- Fontaine, J. R. J., Scherer, K. R., Roesch, E. B., & Ellsworth, P. C. (2007). The world of emotions is not two-dimensional. *Psychological Science*, 18, 1050–1057.
- Fridlund, A. J. (1991). Sociality of solitary smiling: Potentiation by an implicit audience. *Journal of Personality and Social Psychology*, 60, 229–240.
- García-Burgos, D., & Zamora, M. C. (2013). Facial affective reactions to bitter-tasting foods and body mass index in adults. *Appetite*, 71, 178–186.
- Gawronski, B., & Strack, F. (Eds.). (2012). *Cognitive consistency: A fundamental principle in social cognition*. New York: Guilford Press.
- Horstmann, G. (2006). Latency and duration of the action interruption in surprise. *Cognition and Emotion*, 20, 242–273.
- Jakobs, E., Manstead, A. S. R., & Fischer, A. H. (1999). Social motives and emotional feelings as determinants of facial displays: The case of smiling. *Personality and Social Psychology Bulletin*, 25, 424–435.
- Jakobs, E., Manstead, A. S. R., & Fischer, A. H. (2001). Social context effects on facial activity in a negative emotional setting. *Emotion*, 1, 51–69.
- Kay, A. C., Whitson, J. A., Gaucher, D., & Galinsky, A. D. (2009). Compensatory control: Achieving order through the mind, our institutions, and the heavens. *Current Directions in Psychological Science*, 18, 264–268.
- Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud faces database. *Cognition and Emotion*, 24, 1377–1388.
- Lench, H. C., Flores, S. A., & Bench, S. W. (2011). Discrete emotions predict changes in cognition, judgment, experience, behavior, and physiology: A meta-analysis of experimental emotion elicitation. *Psychological Bulletin*, 137, 834–855.
- Lewinski, P., den Uyl, T. M., & Butler, C. (2014). Automated facial coding: Validation of basic emotions and FACS AUs in FaceReader. *Journal of Neuroscience, Psychology, and Economics*, 7, 227–236.
- Lindquist, K. A., Siegel, E. H., Quigley, K. S., & Feldman Barrett, L. (2013). The hundred-year emotion war: Are emotions natural kinds of psychological constructions? Comment on Lench, Flores, and Bench (2011). *Psychological Bulletin*, 139, 255–263.
- Mellers, B., Fincher, K., Drummond, C., & Bigony, M. (2013). Surprise: A belief or an emotion? *Progress in Brain Research*, 202, 3–19.
- Mellers, B. A., Schwartz, A., Ho, K., & Ritov, I. (1997). Decision affect theory: Emotional reactions to the outcomes of risky options. *Psychological Science*, 8, 423–429.
- Mendes, W. B., Blascovich, J., Hunter, S. B., Lickel, B., & Jost, J. T. (2007). Threatened by the unexpected: Physiological responses during social interactions with expectancy-violation partners. *Journal of Personality and Social Psychology*, 92, 698–716.
- Meyer, W. U., Niepel, M., Rudolph, U., & Schützwohl, A. (1991). An experimental analysis of surprise. *Cognition and Emotion*, 5, 295–311.
- Meyer, W. U., Reisenzein, R., & Schützwohl, A. (1997). Towards a process analysis of emotions: The case of surprise. *Motivation and Emotion*, 21, 251–274.
- Miceli, M., & Castelfranchi, C. (2015). *Expectancy and emotion*. Oxford: Oxford University Press.
- Nohlen, H. U., Van Harreveld, F., Rotteveel, M., Barends, A. J., & Larsen, J. T. (2016). Affective responses to ambivalence are context-dependent: A facial EMG study on the role of inconsistency and evaluative context in shaping affective responses to ambivalence. *Journal of Experimental Social Psychology*, 65, 42–51.
- Noldus. (2012). *Facereader: Tool for automatic analysis of facial expression: Version 5.0*. Wageningen: Noldus Information Technology B.V.
- Noordewier, M. K., & Breugelmanns, S. M. (2013). On the valence of surprise. *Cognition and Emotion*, 27, 1326–1334.
- Noordewier, M. K., Topolinski, S., & Van Dijk, E. (2016). The temporal dynamics of surprise. *Social and Personality Psychology Compass*, 10, 136–149.
- Proulx, T., Inzlicht, M., & Harmon-Jones, E. (2012). Understanding all inconsistency compensation as a palliative response to violated expectations. *Trends in Cognitive Sciences*, 16, 285–291.
- Proulx, T., Slegers, W., & Tritt, S. M. (2017). The expectancy bias: Expectancy-violating faces evoke earlier pupillary dilation than neutral or negative faces. *Journal of Experimental Social Psychology*, 70, 69–79.
- Reisenzein, R. (2000a). Exploring the strength of association between the components of emotion syndromes: The case of surprise. *Cognition and Emotion*, 14, 1–38.
- Reisenzein, R. (2000b). The subjective experience of surprise. In H. Bless & J. P. Forgas (Eds.), *The message within: The role of subjective experience in social cognition and behaviour* (pp. 262–279). Philadelphia, PA: Psychology Press.
- Reisenzein, R., Bördgen, S., Holtbernd, T., & Matz, D. (2006). Evidence for strong dissociation between emotion and facial displays: The case of surprise. *Journal of Personality and Social Psychology*, 91, 295–315.
- Reisenzein, R., Horstmann, G., & Schützwohl, A. (2017). The cognitive-evolutionary model of surprise: A review of the

- evidence. *Topics in Cognitive Science*. Advance online publication. doi:10.1111/tops.12292
- Reisenzein, R., & Meyer, W. U. (2009). Surprise. In D. Sander & K. R. Scherer (Eds.), *Oxford companion to the affective sciences* (pp. 386–387). Oxford: Oxford University Press.
- Reisenzein, R., Meyer, W.-U., & Niepel, M. (2012). Surprise. In V. S. Ramachandran (Ed.), *Encyclopedia of Human Behavior* (2nd ed, pp. 564–570). New York City: Elsevier.
- Reisenzein, R., Studtmann, M., & Horstmann, G. (2013). Coherence between emotion and facial expression: Evidence from laboratory experiments. *Emotion Review*, 5, 16–23.
- Roelofs, K., Hagensmaars, M. A., & Stins, J. (2010). Facing freeze: Social threat induces bodily freeze in humans. *Psychological Science*, 21, 1575–1581.
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39, 1161–1178.
- Rutjens, B. T., van Harreveld, F., van der Pligt, J., Kreemers, L. M., & Noordewier, M. K. (2013). Steps, stages, and structure: Finding compensatory order in scientific theories. *Journal of Experimental Psychology: General*, 142, 313–318.
- Scherer, K. R. (2001). Appraisal considered as a process of multi-level sequential checking. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 92–120). New York: Oxford University Press.
- Scherer, K. R., Zentner, M. R., & Stern, D. (2004). Beyond surprise: The puzzle of infants' expressive reactions to expectancy violation. *Emotion*, 4, 389–402.
- Schützwohl, A., & Reisenzein, R. (2012). Facial expressions in response to a highly surprising event exceeding the field of vision: A test of Darwin's theory of surprise. *Evolution and Human Behavior*, 33, 657–664.
- Seery, M. D., Leo, R. J., Lupien, S. P., Kondrak, C. L., & Almonte, J. L. (2013). An upside to adversity? Moderate cumulative lifetime adversity is associated with resilient responses in the face of controlled stressors. *Psychological Science*, 24, 1181–1189.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2013). *Life after p-hacking*. Paper presented at the fourteenth annual meeting of the society for personality and social psychology, New Orleans, LA.
- Soussignan, R., Chadwick, M., Léonor, P., Conty, L., Dezechache, G., & Grèzes, J. (2013). Self-relevance appraisal of gaze direction and dynamic facial expressions: Effects of facial electromyographic and autonomic reactions. *Emotion*, 13, 330–337.
- Tomkins, S. S. (1984). Affect theory. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 163–196). Hillsdale, NJ: Erlbaum.
- Topolinski, S., Likowski, K. U., Wyers, P., & Strack, F. (2009). The face of fluency: Semantic coherence automatically elicits a specific pattern of facial muscle reactions. *Cognition and Emotion*, 23, 260–271.
- Topolinski, S., & Strack, F. (2015). Corrugator activity confirms immediate negative affect in surprise. *Frontiers in Psychology*, 6, 1–8.
- Valenzuela, A., Strebel, J., & Mellers, B. (2010). Pleasurable surprises: A cross-cultural study of consumer responses to unexpected incentives. *Journal of Consumer Research*, 36, 792–805.
- Van Boxtel, A., & Jessurun, M. (1993). Amplitude and bilateral coherency of facial and jaw-elevator EMG activity as an index of effort during a two-choice serial reaction task. *Psychophysiology*, 30, 589–604.
- Van Dillen, L. F., Harris, L., Van Dijk, W. W., & Rotteveel, M. (2015). Looking with different eyes: The psychological meaning of categorisation goals moderates facial reactivity to facial expressions. *Cognition and Emotion*, 29, 1382–1400.
- Van Kuilenburg, H., Wiering, M., & Den Uyl, M. (2005). A model-based method for facial expression recognition. In J. Gama, R. Camacho, P. Brazdil, A. Jorge, & L. Torgo (Eds.), *Lectures notes in computer science: Vol. 3720. Machine learning: ECML 2005* (pp. 194–205). Berlin: Springer-Verlag.
- Yartz, A. R., & Hawk, L. W. (2002). Addressing the specificity of affective startle modulation: Fear versus disgust. *Biological Psychology*, 59, 55–68.