The dynamics of surprise and curiosity
Noordewier, M.K.

CHAPTER 3

Surprise: unfolding of facial expressions

This chapter is based on: Noordewier, M. K., & Van Dijk, E. (2016). Surprise: unfolding of facial expressions. *Manuscript submitted for publication.*

Acknowledgements: Thanks to Anne-Linn Beekhof and Frederique Arntz for manually coding the faces. Thanks to Carlo Konings for his help with programming the studies and converting the data. Thanks to Suzanne Kuiper for her help with editing the videos. Thanks to Peter Lewinski for his advice on FaceReader.

Surprise: Unfolding of Facial Expressions

When people are confronted with unexpected, inconsistent, or disfluent stimuli, they experience surprise (e.g., Meyer, Niepel, Rudolph, & Schützwohl, 1991; Meyer, Reisenzein, & Schützwohl, 1997; Noordewier & Breugelmans, 2013; Noordewier, Topolinski, & Van Dijk, 2015; Reisenzein, 2000b). When surprised, ongoing thoughts and activities are interrupted and attention is directed at the surprising stimulus to make sense of it (e.g., Camras et al., 2002; Horstmann, 2006; Meyer et al., 1991, 1997; Reisenzein, 2000b; Scherer, 2001). After sense-making, surprise dissipates and other affective states follow depending on the nature of the surprising event (e.g., Ekman, 2003; Noordewier & Breugelmans, 2013; Tomkins, 1984).

Responses to a surprising stimulus are thus dynamic and unfold from initial interruption (i.e., responses to the unexpectedness of the event) to cognitive mastering (i.e., responses to the valence of the event; Meyer et al., 1991, 1997; Noordewier & Breugelmans, 2013; Noordewier et al., 2015). Therefore, to study surprise rather than its consequences, it is key to take the temporal dynamics of sense-making into account (cf. Noordewier et al., 2015; see also Noordewier & Breugelmans, 2013; Tomkins, 1984). We aimed to do exactly this and systematically tested the temporal unfolding of facial expressions in response to surprising stimuli, to distinguish surprise from the state that follows it. Moreover, we aimed to provide more insight into what surprise expression looks like and what this might mean in terms of the valence of surprise.

Facial expressions are particularly suited to reveal the unfolding of responses, because they can capture initial responses to a surprising stimulus as well as changes in responses over time (cf. Noordewier et al., 2015). In a first study on this, expressions of people who were positively surprised in TV-shows were analyzed (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. Faces were more negative in the first moments as compared to later; a pattern that was assumed to reflect unfolding of responses, from interruption to mastering (Noordewier & Breugelmans, 2013).

In line with this, a facial electromyography study (fEMG; Reisenzein et al., 2006, Study 7) showed that participants who were surprised with an unanticipated photograph of themselves had a slight increase of corrugator activity (i.e., frown; also found in Topolinski & Stack, 2015; see also Schützwohl & Reisenzein, 2012), which was after 1-3 seconds followed by an increase in zygomaticus activity (i.e., smile). While in this study Reisenzein et al. aimed to test the occurrence of the surprise expression (raised eyebrows, eye-widening, jaw drop; Darwin, 1872/1998; Ekman, Friesen, & Hager, 2002) rather than its temporal dynamics per se, it supports the notion that initial responses to surprising stimuli differ from later responses. Interestingly, these studies also point to two other elements of the responses to surprise stimuli: The initial expression is more negative than the later expression and it may involve frowning.

First, regarding the valence of the expression, several lines of work indeed support the notion that surprise or interruption is relatively negative (see also Hajcak, 2012; Mendes et al., 2007; Miceli & Castelfranchi, 2015; Noordewier & Breugelmans, 2013; Topolinski & Strack, 2015). In fact, from the point of view of cognitive consistency theories and personal control perspectives, surprise reflects inconsistency, disruption, and lacking 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; Miceli & Castelfranchi, 2015; Proulx, Inzlicht, & Harmon-Jones, 2012; Rutjens, Van Harreveld, Van der Pligt, Kreemers, & Noordewier, 2013).

So, even if the surprising stimulus is positive, people first experience this brief phase of inconsistency and lack of meaning (i.e., surprise), before they can appreciate the outcome as is it (i.e., the state after surprise). Importantly, this means that to understand surprise, it should not be confused with its consequences and only by taking time into account, surprise can be distinguished as initial interruption from subsequent states that follow after sense-making. This temporal dynamics perspective also explains why other researchers suggested that surprise feels positive (e.g., Fontaine et al., 2007; Mellers, Schwartz, Ho, & Ritov, 1997; Valenzuela, Strebel, & Mellers, 2010), as they measure retrospective evaluations or feelings *after* cognitive mastering (see Noordewier et al., 2015, for a similar reasoning).

Second, the expression after a surprise might thus involve frowning (Topolinski & Strack, 2015; see also Reisenzein et al., 2006;

Schützwohl & Reisenzein, 2012). This seems inconsistent with the "typical" surprise expression (raised eyebrows, eye-widening, jaw drop). Previous research already showed, however, that this "typical" expression is in fact rarely and only partly observed in a minority of surprised people (mostly raising eyebrows only; 4-25% in Reisenzein et al., 2006; 34% in Reisenzein, 2000; 10-33% in Schützwohl & Reisenzein, 2012). Yet, it should be noted that in these studies it is not always clear whether participants who did not show a surprised face showed a neutral face or other facial action. Besides frowns, smiles were also observed (2-86% in Reisenzein et al., 2006; 26-71% in Schützwohl & Reisenzein, 2012; 8.6-12.1% in Scherer, Zenter, & Stern, 2002); and in infant studies, freezing (Camras et al., 2002; Scherer et al., 2002) and signs of interest were also found (Camras et al., 2002).

There is thus some indication that the expression after a surprise involves frowning, yet other facial action has also been observed. Importantly, because most studies did not take time into account, the possibility that different facial actions follow each other remains untested. To better understand the expression after a surprise, it is therefore important to incorporate time when coding different facial actions.

The Current Studies

To clarify the temporal unfolding of facial expressions in response to a surprising stimulus, we developed two repetition-change studies—a standardized 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 3.1 and

3.2) as well as a negative surprise (Experiment 3.2) and recorded facial expressions using webcams. Using computerized and manual coding, we measured the valence of facial expression over time as well as different facial action elements.

We predict that if surprise is relatively negative and different from subsequent states after sense-making, 1) initial expressions to positive surprises are more negative than later expressions and 2) responses to positive and negative surprises are initially similar and only start to differentiate depending on the nature of the event after some time. We report all manipulations, all measures, and all data exclusions. Sample sizes are at least 50 per cell (Simmons, Nelson, & Simonsohn, 2011, 2013), yet, we collected more data to be able to account for data exclusion as a result of coding errors and participants not giving permission to use their material.

Experiment 3.1: 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 (47 females, 24 males; M_{age} = 22.32 years, SD_{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 positive surprising target.

Procedure and Materials. The study started with a cover story to explain the use of the webcam and to induce a social context. Participants were told that they would participate in a study on eyemovement and attention to pictures and in order to analyze their eye-movements, we would record them with a webcam. Then, 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 this setting and we reasoned that one explanation could be that it is not social enough (e.g., Friedlund, 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 extra 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). They were then 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 many 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 neutral part of the experiment.

After instructions, participants continued to the main part of the experiment in which they were surprised. 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 and ended with a question asking participants to indicate whether the last picture in the 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 gave each trial a certain rhythm, which strengthens the expectancy of what will follow (buildings and a question).

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 question, we showed them a gif-file of a puppy, 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 seconds 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 we asked them to report their age and gender and whether they 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 they were 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 analyzing 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 (see below).

Participant Selection, Target Evaluation, Footage Editing. We excluded participants who did not give permission to use their footage (N = 8) and who participated before in a similar (pilot) study (N = 2). Next, we excluded participants who wore glasses (N = 8; glasses may hinder classification in FaceReader; Noldus, 2012, p. 16) and those who resulted in other coding errors (N = 1, extreme yawning). We analyzed the data of the remaining 52 participants (18 males, 34 females; M_{age} = 21.83 years, SD_{age} = 4.79). We first checked the ratings of the target. As expected, the target was rated as relatively surprising (M = 6.00, SD = 1.12) and as relatively positive (M = 5.85, SD = 1.36).

Next, we edited the videos such that they started two seconds before display of the surprising stimulus (baseline) until eight second after the surprise. 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 separately when the surprising event took place. We then converted each video such that they were chronologically similar and as such, comparable in the analyses. Then, we analyzed the footage in FaceReader and using manual coding.

FaceReader. After uploading videos, FaceReader can analyse 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 database is made and intensity of expressions 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, during, and after the surprise. We focused on two output measures: *valence* and *surprise*. FaceReader was set to analyze 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 an average intensity score on valence and surprise for each 0.5-second (mean of 12 or 13 frames) for each participant. After restructuring the data, the final data consisted of 21 data points (resulting in the within participants factor Time) for each participant for both valence and surprise on which we ran repeated measures ANOVAs. In all analyses (also Experiment 3.2), we performed Greenhouse-Geisser corrections where necessary (visible in adjusted degrees of freedom). Note that when we refer to seconds, the negative numbers refer to seconds before the surprising stimulus (i.e., baseline) and the positive numbers refer to seconds after the surprising stimulus.

Figure 3.1: Valence of facial expression in response to a surprising puppy as a function of Time (Experiment 3.1).

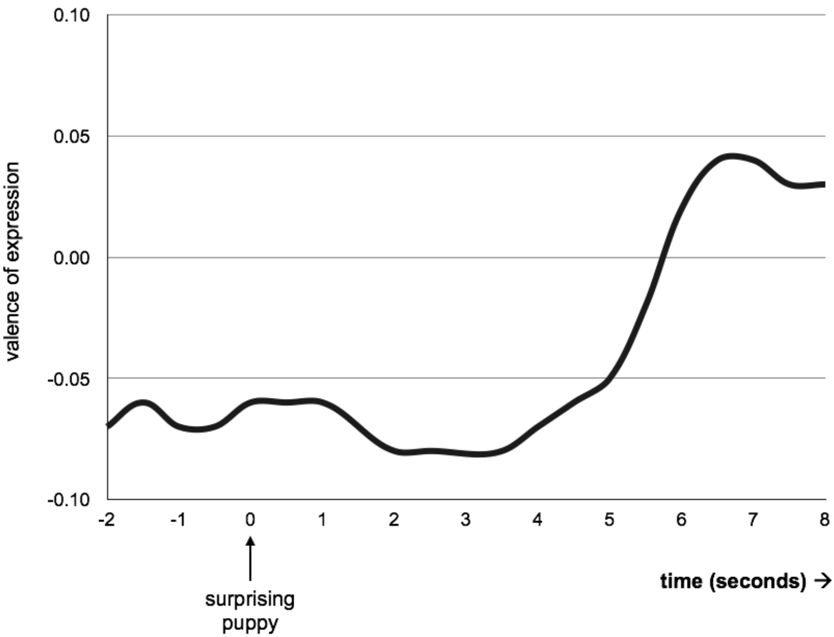
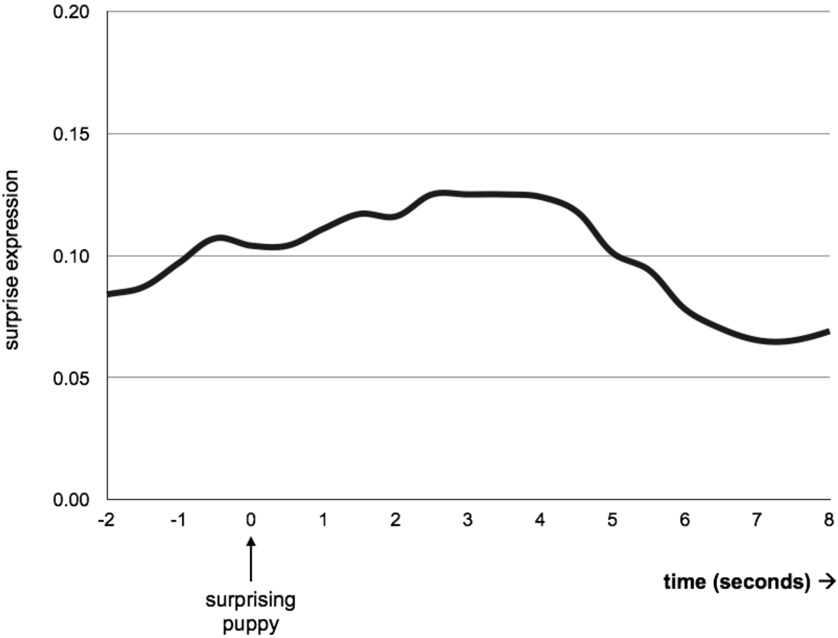


Figure 3.2: Surprise expression in response to a surprising puppy as a function of Time (Experiment 3.1).



Valence. The repeated measures ANOVA (N = 51[5]) showed an effect of Time on valence of expressions, $F(1.66, 83.15) = 4.14$, p = .026, $\eta_p^2 = .08$ (see Figure 3.1). When comparing the valence of expressions relative to baseline (second -2) with within subjects contrasts, we found that expressions were more positive at second 6 until second 8, Fs(1, 50) = 5.16-5.94, ps = .018-.028, η_p^2s = .09-.11.

[5] The number of participants in the analyses is reported when, as a consequence of missing data, it deviates from the total number (i.e., a repeated measures ANOVA and Cochran's Q analyses excludes participants when there are missing data).

Expressions were marginally more positive at second 5.5, $F(1, 50) = 3.03$, $p = .088$, $\eta_p^2 = .06$.

Surprise. The repeated measures ANOVA ($N = 51$) showed a marginal effect of Time on the surprise expression, $F(3.41, 170.72) = 2.47$, $p = .056$, $\eta_p^2 = .05$ (see Figure 3.2). Comparing surprise expressions relative to baseline (second -2) with within subjects contrasts showed marginally more surprise at second 2.5 until 4, Fs(1, 50) = 2.84-3.03, ps = .088-.098, η_p^2s = .05-.06.

In sum, the results show that it takes time to respond positively to a positive surprise and before that, people show some surprise expressions. While there thus seems to be some surprise expression, this is not very strong and the question remains what the facial expression after a surprising stimulus looks like. In Figure 3.1, there is a small decline visible in valence of expression just after the surprise. This decline is not statistically different compared to baseline. Yet, we considered the possibility that expressions are too subtle for FaceReader to detect. We therefore decided to also manually code different facial expressions elements.

Manual Coding. Two independent coders who were blind to the research question and hypotheses of the study were trained using material of a pilot study to code different expression elements. Then, we created screenshots of the 52 videos of the current study at 0.5-second intervals. Each screenshot was coded in terms of absence or presence of a frown, smile, eyebrow raise, jaw drop, and eye-widening. A screenshot was coded a “0” when an element was absent and it was coded on a scale from 1 (*very weak*) to 5 (*very strong*) when an element was present.[6] We also included the option "unclear", which we treated as missing data (0.004% of a total of 12,480 codings).

The interraterreliability of the data was calculated with correlations and because this was not consistently above r = .70 (particularly with brow raise, eye-widening), we recoded the data in terms of absence (0) and presence (1 = 1 to 5 intensity score) of the expressions after surprise (comparable absence/presence coding has also been used in Reisenzein et al., 2006). This recoding was done after one coder individually checked her ratings and corrected 0.05% of the brow raise and frown codings. To check reliability, we calculated percentage agreement between coders. Reliability for frown, smile, brow raise, and jaw drop was good, as it ranged between 70% and 100% (except for brow raise on seconds 0-1 and 2-3.5, where it was between 67-69%). The eye-widening agreement was too low (between 48-63%) and therefore excluded from further analyses. Finally, disagreement on all screenshots after the surprising stimulus was solved through discussion and we analyzed these 100% agreement data. Then, we tested whether Time affected the frequency of each expression element with Cochran's Q tests (see Figure 3.3). When an effect was found, we subsequently

---

[6] Valence was also coded ("How does this person feel?" on a scale from -2 *negative* to +2 *positive*, with 0 representing neutral). Results replicated the FaceReader data pattern, such that a repeated measures ANOVA showed a main effect of Time, $F(1.68, 85.83) = 11.89$, $p < .001$, $\eta_p^2 = .19$. Relative to baseline (second -2), expressions were coded more positive after second 4 until second 8, Fs(1, 51) = 4.04-17.68, ps = .000-.050, η_p^2s = .07-.26. However, because the correlations between the ratings of the two coders were not consistently high (i.e., they ranged between .26 and .85, with 66% < .70), we excluded these results from the main analyses.

compared individual seconds with McNemar tests. Note that we now only focus on what happens after surprise (baseline is excluded).

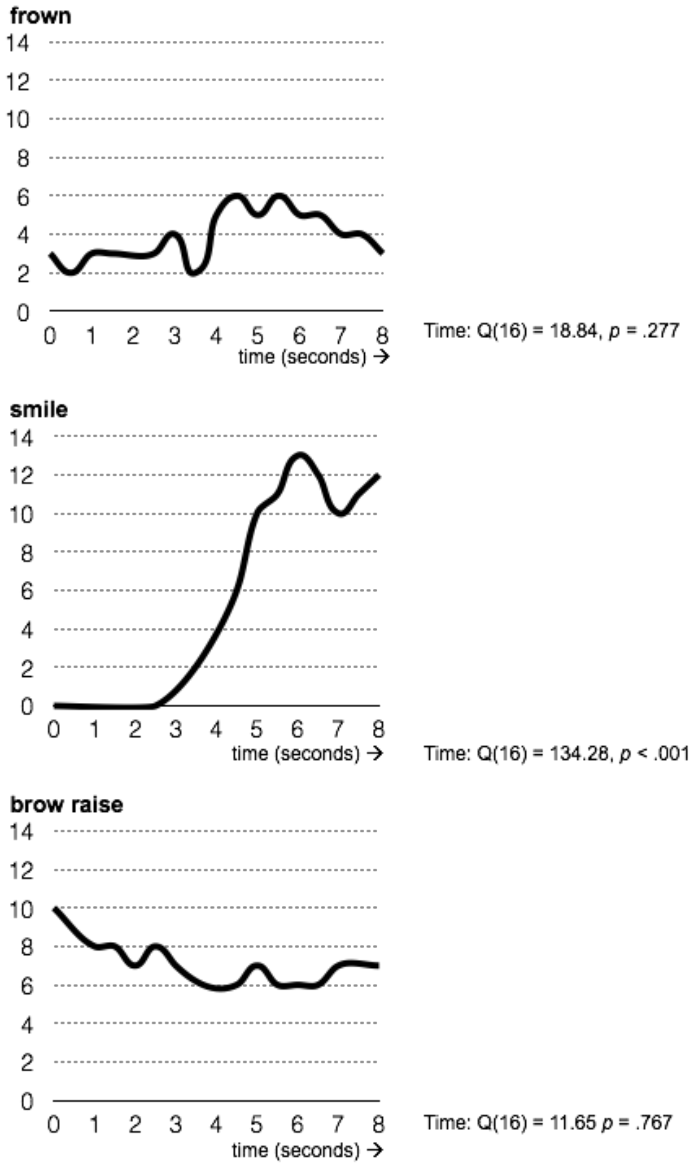
Frown. There was no effect of Time on the number of frowns, $Q(16) = 18.84$, $p = .277$ ($N = 51$).

Smile. There was an effect Time on the number of smiles, $Q(16) = 134.28$, $p < .001$. McNemar comparisons showed that, relative to second 0, there were more smiles from second 4.5 until second 8, *p*s between < .001 and .031, whereas before that, the number of smiles did not differ, *p*s > .124.

Surprise elements. There was no effect of Time on number of brow raises, $Q(16) = 11.65$, $p = .767$ ($N = 50$). There was also no effect of Time on jaw drops, $Q(16) = 12.00$, $p = .744$, which were hardly observed (2 times or less; note that eye-widening was excluded from the analyses; see above).

Taken together, these results show that it takes time to respond positively to a positive surprise. This positivity seems to be characterized by an increase in smiles. In addition, while FaceReader showed some indication of surprise expressions in the first couple of seconds, the manual coding of the separate facial actions did not confirm this result. The second study aimed to induce more intense expressions and also includes a negative surprise.

Figure 3.3: Number of frowns, smiles, and brow raises in response to a surprising puppy as a function of Time (Experiment 3.1).



Experiment 3.2: A Surprising Person

Experiment 3.2 tests the unfolding logic by surprising people in a person-perception setting. We assumed that this setting is more social and self-relevant than the buildings and puppy in Experiment 3.1, which might intensify responses. 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 and as such, we now also included negative surprise condition, which allows us to compare unfolding of responses to surprise to responses to a positive vs. a negative target.

Method

A total of 128 participants (69 females, 59 males; M_{age} = 21.20 years, SD_{age} = 2.25) were randomly assigned 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, they were asked to evaluate pictures 20 faces. Pictures were selected from de Radboud Faces Database (RAFD; Langner, Dotsch, Bijlstra, Wigboldus, Hawk, & Van Knippenberg, 2010). We selected equal numbers of males and females, all showing a neutral expression. Each neutral face was shown five seconds 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 eight seconds. 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 this expression. After the critical trial, the program 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 3.1.

Participant Selection and Footage Editing. First, we excluded participants who did not give us permission to use the footage ($N = 5$), who wore glasses ($N = 8$) or because of other coding errors (i.e., $N = 2$; 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, $M_{age} = 21.14$ years, $SD_{age} = 2.27$).

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 = -0.23. So, based on this we conclude that our stimuli represented a positive versus a negative surprise.

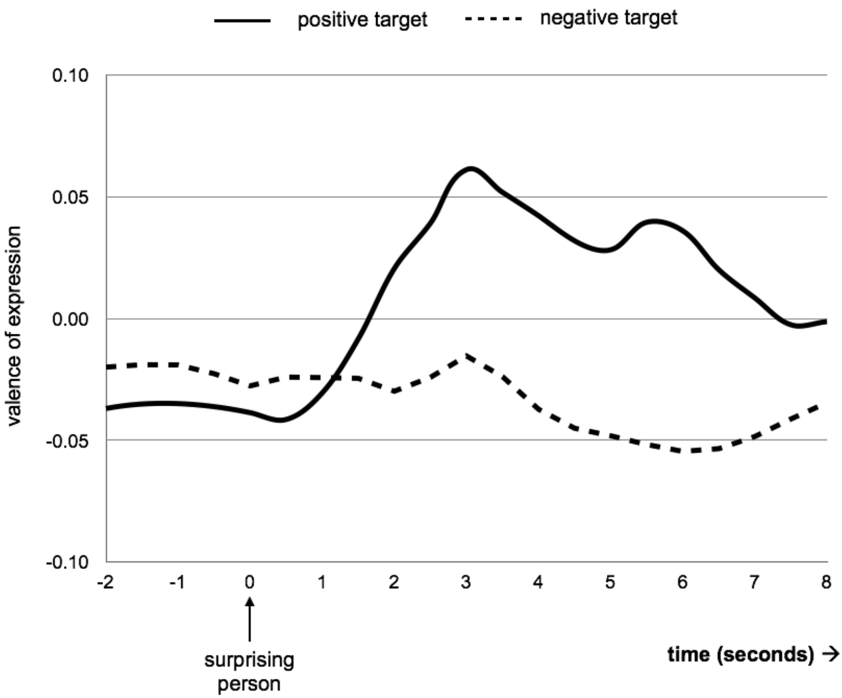
Next, we edited the videos in the same way as in Experiment 3.1, such that they showed participant two seconds before the surprise (baseline) and eight seconds after the surprise. This footage was first coded with FaceReader.

FaceReader. FaceReader was set to analyze 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 each 0.5-second. After restructuring, the final data had 21 data points (Time) for each participant on valence and surprise on which we ran repeated measures ANOVAs (see Figure 3.4), followed by within subjects contrasts (Time) and between condition comparisons (Target).

Valence. The repeated measures ANOVA showed a main effect of Time on valence of expressions, $F(3.07, 337.43) = 2.59$, $p = .051$, $\eta_p^2 = .02$, and a Time x Target interaction, $F(3.07, 337.43) = 4.76$, $p = .003$, $\eta_p^2 = .04$. To interpret the interaction, we compared the effect of Time within the positive and negative target condition separately.

Within the positive target condition, there was a main effect of Time, $F(2.84, 159.89) = 4.55$, $p = .005$, $\eta_p^2 = .07$. Simple contrasts showed that expressions were more positive relative to baseline from second 1.5 until second 7: Fs between 3.97 and 9.34, ps between .003 and .051, η_p^2s between .06 and .14, and marginally more positive at second 8, $F(1,59) = 3.16$, $p = .081$, $\eta_p^2 = .05$. Within the negative target condition, there was a marginal main effect of Time, $F(2.63, 134.05) = 2.18$, $p = .102$, $\eta_p^2 = .04$. Simple contrasts showed that expressions were more negative relative to baseline from second 4.5 until 7.5, *F*s between 3.89 and 5.38, *p*s between .024 and .054, η_p^2s between .07 and 95.

Figure 3.4: Valence of facial expression in response as a function of Target (positive vs. negative) and Time (Experiment 3.2).



So, 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 to more negative expressions in the negative target condition. Interestingly, expressions seemed to unfold slower in the negative as compared to the positive target condition. Moreover, overall, the unfolding seemed faster than in Experiment 3.1. We will discuss this in more detail in the General Discussion.

Next, we compared the valence of expressions between the two target conditions. Two seconds after the surprise, facial expressions started to differentiate, such that at seconds 2-3.5, expressions in the positive target condition became (marginally) more positive than in the negative target condition, *t*s between 1.80 and 1.98 *p*s between .051 and .068, *d*s between 0.34 and 0.38. From second 4 until 7, conditions differed statistically, such that *t*s were between 2.08 and 2.69, and *p*s between .006 and .026, *d*s between 0.42 and 0.52.

Surprise. No effects were observed on the surprise expression (all *F*s < 1; all means ranged between 0.03 and 0.07).

Manual Coding. We also coded the videos manually in the same way as in Experiment 3.1 (same coders, same method, including recoding in terms of 0 = absence and 1 = presence). We made screenshots at 0.5-second intervals of all 112 videos. These screenshots were coded in terms of frown, smile, brow raise, jaw drop, and eye-widening. The "unclear" option was also included, which resulted in 0.003% missing of a total of 26,880 codings.

Agreement ranged between 70% and 100% (with the exception of frowning on seconds 0.5/1, where it was 65/69%; and eye-widening on seconds 1.5/3 where it was 69/68%, respectively) and disagreement was solved through discussion. Then, we compared the frequency of each expression element between conditions with Chi-square tests.

Next, we tested whether Time affected the frequency of each expression element (within each conditions, where relevant) with Cochran's Q tests (see Figure 3.5). When an effect of Time was found, we subsequently compared individual seconds with McNemar tests[7].

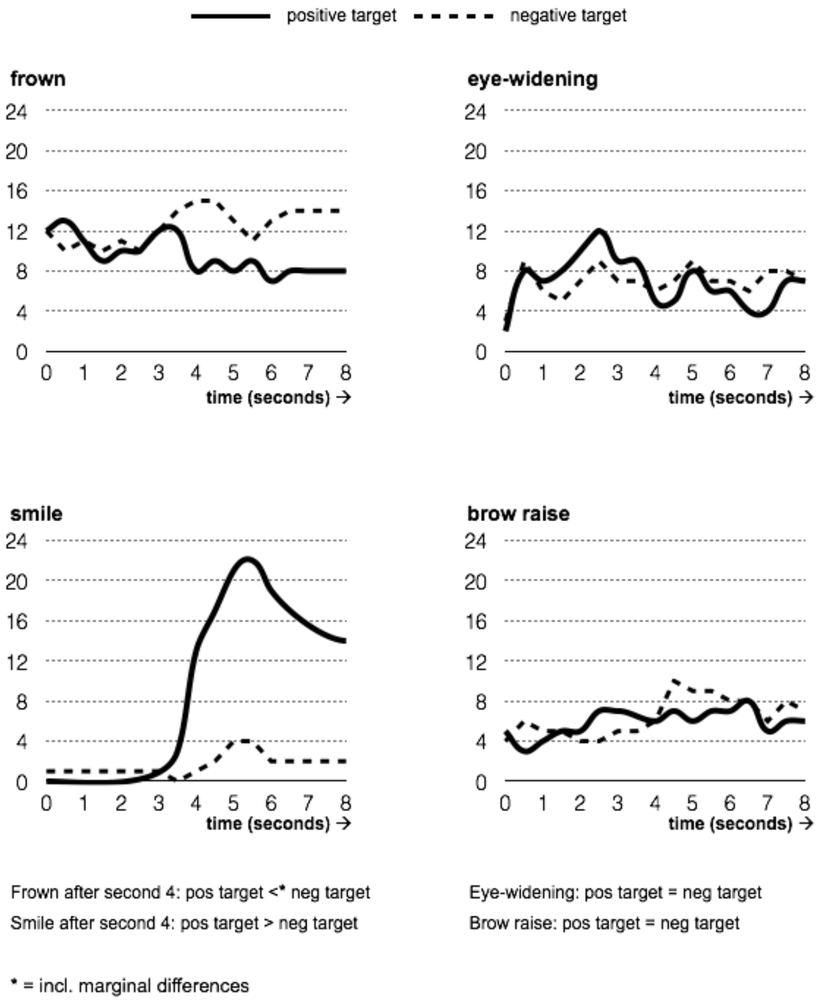
Frown. Chi-square tests showed that there were (marginally) fewer frowns in positive target condition than the negative target condition from second 4 until second 8, χ^2s(1, *N* = 107-111) between 2.83 and 4.16, *p*s between .041 and .092, except second 5 and 5.5, χ^2s(1) = 2.69/0.66, *p*s = .101/.415. Before second 4, equal number of frowns were observed in the positive and negative target condition, χ^2s < .68 and *p*s > .40. Within the positive target condition, we found and effect of Time on the number of frowns, *Q*(16) = 35.38, *p* = .004 (*N* = 51). McNemar tests within the positive target condition showed that relative to second 0, the number of frowns did not differ, *p*s > .288, but there were (marginally) fewer frowns relative to second 0.5 at second 6 and 6.5, *p*s = .031/.063 (other comparisons relative to second 0.5, *p*s > .124). Within the negative target condition, there was no effect of Time, *Q*(16) = 16.46, *p* = .421 (*N* = 49). So, initially, we observed the same

---

[7] Similar to Experiment 3.1, valence of expressions was coded as well and results replicated the FaceReader data pattern, even though interraterreliability was not consistently high (correlations ranged between .32 and .83 with 62% < .7). A repeated measures ANOVA (*N* = 110) on the averaged ratings of both coders showed a Time x Target interaction, *F*(2.12, 228.65) = 15.94, *p* < .001, η_p^2 = .13, and a main effect of Time, *F*(2.12, 228,65) = 16.04, *p* < .001, η_p^2 = .13. Within the positive target condition, the expressions were more positive relative to baseline (second -2), after second 3 until second 8, *Fs*(1, 50) = 6.32-37.86, *p*s = .000-.015, η_p^2s = .10-.40. Within the negative target condition, the expressions did not change over time, *p*s > .558. Comparing positive and negative target conditions, we see that the facial expressions start to differ from second 4 until second 8, *t*s(110) = 3.10-3.75, *p*s < .003, *d*s > 0.59.

amount of frowning in both conditions. Over time there was (marginally) less frowning in the positive target condition as compared to the negative target condition.

Figure 3.5: Number of frowns, smiles, eye-widenings, and brow raises as a function of Target (positive vs. negative) and Time (Experiment 3.2).



Smile. Chi-square tests showed that there were more smiles in the positive target condition than the negative target condition from second 4 until second 8, χ^2s(1, *N* = 111-112) between 8.86 and 14.15, *p*s between < .001 and .003, and marignally more on second 3.5, χ^2(1, *N* = 111) = 2.72, *p* = .099. Before second 3.5, equal number of smiles were observed in the positive and negative target condition, χ^2s (1, *N* = 111-112) < 1.17 and *p*s > .280. Within the positive target condition, we found an effect of Time on the number of smiles, *Q*(16) = 202.16, *p* < .001. McNemar tests within the positive target condition showed that, relative to second 0, there were more smiles from second 4 until second 8, *p*s < .001, whereas before that, the number of smiles did not differ, *p*s > .249. Within the negative target condition, there was no effect of Time, *Q*(16) = 15.39, *p* = .496. So, initially, there were hardly any smiles in both conditions and over time this unfolded to more smiling in the positive target condition, but remained equally low in the negative target condition.

Surprise elements. Chi-square tests showed that conditions did not differ in number of brow raises, χ^2s (1, *N* = 108-111) < 1.69, *p*s > .192, or jaw drops (between 3 and 7 times observed), χ^2s (1, *N* = 111-112) < 1.01, *p*s > .316, or eye-widening, χ^2s (1, *N* = 109-112) < 2.21, *p*s > .136. Next, we tested the overall effect of Time, which showed that Time did not affect the number of brow raises, *Q*(16) = 15.32, *p* = .501 (*N* = 106), or jaw drops, *Q*(16) = 11.12, *p* = .802 (*N* = 109). There was, however, an effect of Time on eye-widening, *Q*(16) = 27.42, *p* = .037

(N = 107). Relative to second 0, there was more eye-widening from second 0.5 until second 8, *p*s between < .001 and .092, except on second 4 and 6.5, *p*s > .145.

Taken together, these results support our unfolding logic. Initially, responses to a positive or a negative surprise did not differ. Over time, expressions to a positive target became more positive, whereas to a negative target they stayed the same. Moreover, initially, there were equal numbers of frowns in both conditions, whereas later, there were less frowns and more smiles in the positive as compared to the negative target condition.

General Discussion

Responses to a surprising stimulus are dynamic and unfold from initial interruption to cognitive mastering of the event (Meyer et al., 1991, 1997; Noordewier & Breugelmans, 2013; Noordewier et al., 2015). To study surprise and distinguish it from the state that follows it, we tested the temporal unfolding of facial expression in response to a surprising stimulus. Results of two repetition-change studies showed that initial expressions after positive surprises are more negative than later expressions. Moreover, expressions after a positive and negative surprise are initially similar and only after some time start to differentiate, depending on the valence of the event. Finally, irrespective of the valence of the surprise, participants showed initially equal number of frowns (Experiment 3.2), which only later turned to smiles when the outcome was positive (Experiments 3.1 and 3.2). Taken together, these results confirmed the notion that responses to surprising stimuli unfold from responses to the unexpectedness of the event to the valence of the event (see also Meyer et al., 1997; Noordewier & Breugelmans, 2013; Noordewier et al., 2015).

Interestingly, when we compare the two studies in terms of speed of unfolding, it becomes clear that expressions unfolded much faster in the study with the surprising faces than the study with the surprising puppy (see Figures 3.1 and 3.4). The relation between expectancy and surprise is a plausible explanation for this difference. The surprising puppy in Experiment 3.1 was categorically different from the preceding repetition trials (buildings), whereas the surprising positive/negative faces in Experiment 3.2 were categorically similar to the preceding repetition trials (neutral faces). Categorical similarity of surprise to the preceding context may make the surprise easier to categorize, 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.

Besides showing unfolding of responses after a surprising stimulus, we also aimed to get more insight into what the expression after a surprise looks like. Previous research already showed that the typical surprise expression with raised eyebrows, eye-widening, and jaw drop is rare (Reisenzein et al., 2006) and that people might initially frown (Topolinksi & Strack, 2015). We find some (marginal) evidence for initial frowning and later eye-widening (Experiment 3.2). Importantly, in line with our unfolding logic, frowns were initially equally strong for both the positive and negative surprises and also the eye-widening was independent of the valence of the surprising target. In addition, in response to positive surprises, smiles were never observed right after the surprising stimuli and only occurred after some time passed (Experiments 3.1 and 3.2). So, also this more detailed expression-coding supports the view that initially, people respond to the unexpectedness of the outcome and only later, after sense-making, respond to the valence of the outcome. Moreover, a tentative conclusion is that frowning may be part of the initial response to a surprising stimulus[8].

If frowning regularly occurs in response to surprising events, the question remains how this should be interpreted. Corrugator activity might just reflect orientation (Van Dillen, Harris, Van Dijk, & Rotteveel, 2015; Yartz & Hawk, 2002), but it has also been related to mental effort (e.g., Van Boxtel & Jessurun, 1993) and negative affect (Topolinski, Likowski, Wyers, & Strack, 2009; Topolinski & Strack 2015). These latter connections would fit the notion that surprise reflects a negative state as a result of inconsistency and lack of meaning (Abelson et al., 1968; Gawronski & Strack, 2012; Kay et al., 2009; Miceli & Castelfranchi, 2015; Noordewier & Breugelmans, 2013; Proulx et al., 2012; Rutjens et al., 2013).

Finally, while we did not predict to find "typical" surprise expressions (based on Reisenzein et al., 2006; Reisenzein & Schützwohl, 2012), it is still intriguing that people think they show this

[8] Note that we do not find this in Experiment 3.1, where expressions were overall less intense than in Experiment 3.2. As such, it remains possible that the frowns were too subtle to detect without fEMG.

expression (see Reisenzein et al., 2006) and recognize the expression as surprise in others (e.g., Ekman et al., 1987). One possibility that is still untested is that this expression occurs when intending to *communicate* surprise. So, rather than a direct consequence of an internal state, people would raise their eye-brows, widen their eyes and drop their jaw to inform someone else that they are surprised. This would mean that a social context alone is not enough to induce the "typical" surprise expression (as indeed found in Reisenzein et al., 2006; Schützwohl & Reisenzein, 2012), but instead people should directly interact with each other. Future research could test this possibility.

Conclusion

When people are surprised, they initially respond to the unexpectedness of the event and later to the valence of the event. Two repetition-change studies supported this unfolding notion and showed that 1) after positive surprises, initial expressions are more negative than later expressions and 2) expressions to positive and negative surprises are initially similar and only start to differentiate depending on the valence of the event after some time. Finally, initial frowning was independent of the valence of the surprising target and this only later turned to smiles in the case of positive surprises. Taken together, these studies show that to study surprise, it is key to take its temporal dynamics into account and to distinguish surprise from the state that follows it.