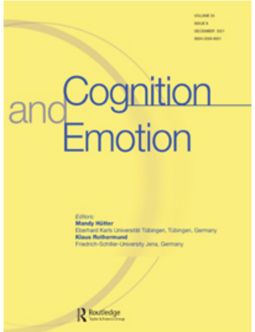

Cognition and Emotion

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/pcem20

Predicting vs. guessing: the role of confidence for pupillometric markers of curiosity and surprise

Maria Theobald, Elena Galeano-Keiner & Garvin Brod

COGNITION AND EMOTION
https://doi.org/10.1080/02699931.2022.2029733

BRIEF REPORT

Predicting vs. guessing: the role of confidence for pupillometric markers of curiosity and surprise

Maria Theobald [a], Elena Galeano-Keiner[a] and Garvin Brod[a,b]

[a]DIPF | Leibniz Institute for Research and Information in Education, Frankfurt, Germany; [b]Department of Psychology, Goethe University, Frankfurt, Germany

ABSTRACT

Asking students to generate a prediction before presenting the correct answer is a popular instructional strategy. This study tested whether a person's degree of confidence in a prediction is related to their curiosity and surprise regarding the answer. For a series of questions about numerical facts, participants ($N = 29$) generated predictions and rated their confidence in the prediction before seeing the correct answer. The increase in pupil size *before* viewing the correct answer was used as a physiological marker of curiosity, and the increase in pupil size *after* viewing the correct answer was used as a physiological marker of surprise. The results revealed that the pupillometric marker of curiosity was most pronounced if students were slightly more confident in their prediction than usual, and it was lower for predictions made with either very high or very low confidence. Furthermore, the results showed that high-confidence prediction errors and low-confidence correct responses yielded a pupillary surprise response, suggesting that highly unexpected results evoke surprise, independent of the correctness of the prediction. Together, results suggest that confidence in a prediction plays an important role in the occurrence of epistemic emotions such as curiosity and surprise.

ARTICLE HISTORY
Received 17 August 2021
Revised 21 December 2021
Accepted 11 January 2022

KEYWORDS
Confidence; prediction error; curiosity; surprise; pupillometry

What is the colour of an orange in tropical regions? When picturing a ripe, juicy orange, most people would say with high confidence that oranges are orange. People who live in tropical regions would likely predict a different colour with high confidence. Still others may suspect that this is a trick question and guess another colour with low confidence. Varying degrees of prior knowledge about oranges and their typical colours thus go hand in hand with varying degrees of confidence in the prediction. After making a prediction, people become curious about whether their answer is correct. Those who were very confident that oranges are orange will be surprised to learn that ripe oranges are green in tropical regions. The answer is not that surprising for those who have already tasted a ripe, green orange. For those who guessed incorrectly, learning that oranges are green is similarly unsurprising. The answer may well be surprising, however, for those who randomly guessed green correctly.

These examples illustrate how generating a prediction before seeing the correct answer can evoke epistemic emotions. Epistemic emotions, such as curiosity and surprise, relate to knowledge and the generation of knowledge, and can be prompted by discrepancies or contradictions between existing knowledge and new information or by knowledge gaps (Vogl et al., 2019). Curiosity is evoked by an information gap, i.e. by a discrepancy between what a person knows and what they want to know (Loewenstein, 1994). Surprise is evoked by feedback that violates their prior expectation, i.e. through prediction errors (e.g. Alexander & Brown, 2019; Reisenzein et al., 2019). Several recent studies have indicated that both curiosity and surprise are enhanced when participants engage in generating predictions before being presented with the correct

CONTACT Maria Theobald theobald@dipf.de DIPF | Leibniz Institute for Research and Information in Education, Frankfurt, Germany
Supplemental data for this article can be accessed at https://doi.org/10.1080/02699931.2022.2029733

answer (Brod et al., 2018; Brod & Breitwieser, 2019; Theobald & Brod, 2021). These results suggest that generating a prediction increases awareness of an existing knowledge gap, which promotes curiosity and, if the prediction turns out to be wrong, may evoke surprise.

Predictions can be made with varying degrees of confidence. Feelings of confidence constitute a metacognitive experience, which is generated after knowledge has been retrieved (Nelson & Narens, 1990). In the context of predicting, feelings of confidence depict a person's subjective certainty about a prediction. Feelings of confidence likely result from pre-existing semantic associations between the question and the correct answer, which allow us to generate a prediction and not merely a guess (c.f. Brod, 2021). One main goal of the current study was to replicate and extend previous findings that found that a person's degree of confidence in a prediction is related to their curiosity and surprise regarding the answer.

Confidence and curiosity are assumed to be closely related. In particular, it has been suggested that their relation follows an inverted-U-shape (Kang et al., 2009): The authors showed that curiosity ratings are highest for predictions made with an intermediate level of confidence and lowest for low-confidence and high-confidence predictions. In a recent review it has further been suggested that curiosity generally increases with confidence but then sharply decreases when the correct answer is already known (Metcalfe et al., 2020). For instance, Metcalfe et al. (2017) showed that participants were particularly curious when they thought that they almost knew the correct answer. In these cases, participants likely anticipated to be correct. Anticipating a rewarding outcome has been shown to activate brain regions that have also been linked to higher self-reported curiosity (Kang et al., 2009; Schneider et al., 2018). In summary, these results suggest that curiosity is largest for predictions made with intermediate to high levels of confidence while it is lowest for low-confidence ratings and for answers that are already known.

Confidence reflects the strength of expectation and therefore also plays an important role for surprise. Surprise intensity increases with the degree of unexpectedness of an outcome, which depends on learners' confidence in their incorrect belief (Reisenzein et al., 2019). Hence, incorrect guesses that are made with low confidence should not elicit a strong surprise response. In contrast, incorrect predictions made with high confidence should elicit a strong surprise response. For instance, Reisenzein (2000) found that participants reported higher surprise after making an incorrect prediction with high confidence. These surprising events might then receive special attention (see Reisenzein et al., 2019 for an overview). For instance, in Butterfield and Metcalfe (2006), participants' performance in a tone detection task was impaired after they committed a high-confidence error, suggesting that the expectancy-violation feedback took up more attentional resources. Performance in the tone detection task was also impaired after correct predictions that were made with low confidence. In this case, being unexpectedly correct may have induced surprise as well, because the subject did not expect that the low-confidence guess would turn out to be correct. Surprise, however, was not assessed in this study. Hence, it remains unclear whether prediction errors and confidence interactively predict the level of surprise.

Another recent study examined correlations between high-confidence errors and ratings of surprise and curiosity (Vogl et al., 2019). Students answered trivia questions and gave confidence ratings. After receiving feedback, students rated their surprise about the outcome and their curiosity about receiving an explanation for the outcome. Correctness was assessed as a binary outcome. It thus remained unclear whether the size of the error matters for the relation between high and low confidence errors and surprise. Furthermore, surprise and curiosity were assessed simultaneously via retrospective self-reports after seeing the correct outcome. Hence, the temporal dynamics of surprise and curiosity when anticipating the outcome were not examined.

Changes in pupil size capture temporal dynamics and have been shown to be markers of the physiological component of curiosity and surprise. The continuous increase in pupil size *before* an outcome is presented has been shown to be related to self-reported degrees of curiosity (Brod & Breitwieser, 2019; Kang et al., 2009). Likewise, the increase in pupil size within the first two seconds *after* an expectancy-violation outcome is presented constitutes a reliable marker of a violation of expectation (Breitwieser & Brod, 2021; Krüger et al., 2019; Reisenzein et al., 2006; Theobald & Brod, 2021), which has been called the pupillary surprise response (Brod et al., 2018). Taken together, these results suggest that curiosity is associated with an increase in pupil size when anticipating feedback, while surprise is associated with an increase in pupil size after receiving corrective feedback. In summary, while pupil size does not reflect to which degree learners *felt* surprised or curious, it does provide quantitative data on the physiological components of curiosity and surprise and their temporal dynamics.

Task-evoked changes in pupil size offer a fine-grained assessment of changes in arousal during a cognitive task (Beatty, 1982). These changes are driven by the release of norepinephrine in the brainstem's locus coeruleus (Joshi et al., 2016), which in turn receives input from the anterior cingulate cortex (ACC; Aston-Jones & Cohen, 2005). Increased activity in the ACC has been linked to the anticipation of reward and curiosity (Gruber et al., 2014). The ACC is further activated by surprising events that violate prior expectations (Alexander & Brown, 2019). ACC activity has been shown to correspond to an increase in pupil size when anticipating a reward (i.e. indicative of curiosity; Schneider et al., 2018) and when processing conflicting information (i.e. indicative of surprise; Ebitz & Platt, 2015). Hence, the evidence from neurophysiological studies suggests that changes in pupil size correspond to changes in physiological arousal that occur in parallel with curiosity and surprise.

The current study

In the present study, we tested how confidence in a prediction relates to the physiological component of curiosity and surprise, as assessed by pupillometry. We used an experimental paradigm in which university students were asked to predict the correct number in a series of incomplete facts in an "X out of 10" format and to indicate their confidence in their predictions. We assessed changes in pupil size as markers of the physiological component of curiosity and surprise. We define the increase in pupil size when anticipating corrective feedback as a marker of the physiological component of curiosity, and we define the increase in pupil size after seeing an unexpected result as a marker of the physiological component of surprise. These changes in pupil size are assumed to originate from changes in noradrenaline release (Joshi et al., 2016). Noradrenaline release has been linked to the recognition of a violation of expectation (Lawson et al., 2021). Further, an increase in pupil size when anticipating an outcome has been linked to increased activity in brain regions involved in reward anticipation and to higher self-reported curiosity (Kang et al., 2009; Schneider et al., 2018). Hence, the pupillary data provide fine-grained, sensitive markers of trial-to-trial changes in the physiological component of surprise and curiosity in high temporal resolution.

We aimed to replicate and extend previous findings on the relation between confidence in a prediction, curiosity, and surprise using a well-tested quiz paradigm (see Reisenzein, 2000). We hypothesised linear and quadratic relations between confidence and the pupillometric marker of curiosity, thereby replicating prior findings (e.g. Brod & Breitwieser, 2019; Kang et al., 2009; Metcalfe et al., 2020). Furthermore, we aimed to replicate and extend previous research that showed an interaction between confidence and prediction errors in determining surprise (see e.g. Reisenzein et al., 2019 for a review; Vogl et al., 2019). We hypothesised that large high-confidence prediction errors would evoke a stronger pupillary surprise response compared to small high-confidence prediction errors. We thereby extend previous findings in two ways. First, we tested whether the proposed interaction between confidence and prediction error depends on the size of the prediction error (i.e. small, medium, or large). Second, we measured the pupillary surprise response compared to self-reported surprise. In addition, we explored whether low-confidence correct predictions likewise evoke a stronger pupillary surprise response compared to high-confidence correct predictions, as suggested by Butterfield and Metcalfe (2006).

Materials and methods

Participants

We tested 29 university students ($M_{Age} = 23.07$, $SD_{Age} = 3.07$, [19; 30], 69% female) to obtain the target sample size of $n = 28$. Data from one participant were discarded because the participant always indicated the same confidence level across all trials, indicating an inadequate use of the confidence scale. Sample size was determined a priori using G*Power with the following settings: 2×2 repeated measures ANOVA, effect size $f = .25$, alpha = .05, beta = .90, correlation among repeated measures = .7. The effect size was derived from a prior study in which we used a similar paradigm to test how unexpected outcomes affect memory and surprise (Brod et al., 2018). Please note that we decided after the data had been collected (but before performing any analyses) that it is more appropriate to perform the analyses using linear mixed models. A major advantage of linear mixed models is that they can account for dependencies among observations originating from the same person by specifying random effects for each person. Furthermore, linear mixed models are suitable for unbalanced designs where the number of observations (e.g. the number of high-confidence errors) varies across participants, as was the case for our study. We therefore conducted an additional power simulation for linear mixed models using a simulated data set based on the current sample and paradigm: 28 participants, 90 trials, fixed effect for interaction term = .05, value for random intercept = .01, value for variance in random intercept = .0002, critical value = 2, 1000 simulation. The input values were derived post-hoc based on the results from a mixed effects regression that tested the interaction between confidence and prediction error as a predictor of the pupillary surprise response. Results revealed a power of 81.10% CI[71.93, 88.16] for a 2-way interaction effect. The power simulation was conducted using the "simr" package in R.

Participants were recruited through advertisements at a large university campus. They gave written informed consent prior to testing and received 10 Euro or course credit for participating. Ethics approval was obtained from the ethics committee of DIPF | Leibniz Institute for Research and Information in Education.

Stimuli

Stimuli comprised 90 numerical facts divided into two subsets of 45 facts each. Each fact was worded in the format "X out of 10", which is equivalent to a percentage estimate. Correct answers ranged between "1" and "9" because "0" and "10" were never correct.

Procedure

Participants completed a computerised, numerical fact prediction task with two blocks of 45 facts each (see Figure 1). The task was very similar to the one used in Brod and Breitwieser (2019) with one exception: we replaced the curiosity rating with a confidence rating. Participants were instructed to try to remember the facts for a subsequent memory test: the data from this test are not part of the current manuscript.

In the study phase, participants were shown a series of incomplete facts in an "X out of 10" format. Participants were asked to predict the correct number by clicking on a 10-point scale. Participants were shown their prediction for 1 s (response presentation) and then rated their confidence in their prediction on a scale from "1" (not confident) to "5" (very confident). After the confidence rating, participants were shown the initial question again for 2 s (anticipation phase) before the correct number was revealed (results phase).

Stimulus presentation and eye-tracking procedures

Stimuli were presented using PsychoPy v1.83.02. Participants were seated about 68 cm from a computer screen in a dimly lit room. The eye-tracking camera (EyeLink 1000, SR Research, Osgoode, Ontario, Canada) was placed below the computer screen and recorded at a frequency of 500 Hz throughout the experiment. Eye tracking was used to record changes in participants' pupil size (right eye) during the anticipation phase and during the results phase when the correct answer was revealed.

Data analyses

Data were analysed using R and the α level was set to 0.05 throughout all analyses. As a measure of effect size, we report standardised regression weights (β). We used "itrackR" (https://github.com/jashubbard/itrackR) to analyse the pupillary data along with self-developed analysis scripts. We used "lme4" to estimate the linear mixed models, and we used "ggplot2" to generate the plots.

Analysis of the pupillary data

We used itrackR to analyse the pupillary data along with self-developed analysis scripts. We merged the pupillary and behavioural data from the study phase. Then, we removed any blinks and interpolated the missing values using cubic spline interpolation. Next, we derived separate markers for curiosity and surprise in the pupillary data.

To derive a marker of curiosity from the pupillary data, we calculated, for each trial, the average percentage change in pupil diameter from 1.5–3 s after the onset of the anticipation phase, relative to the pupil baseline phase (100 ms before onset of anticipation phase until − 200 ms after onset of anticipation phase). This time window covers 500 ms before the onset of the results phase until 1 s after the onset of the results phase. To derive a marker of surprise from the pupillary data, we calculated, for each trial, the average percentage change in pupil diameter 0.5–2 s

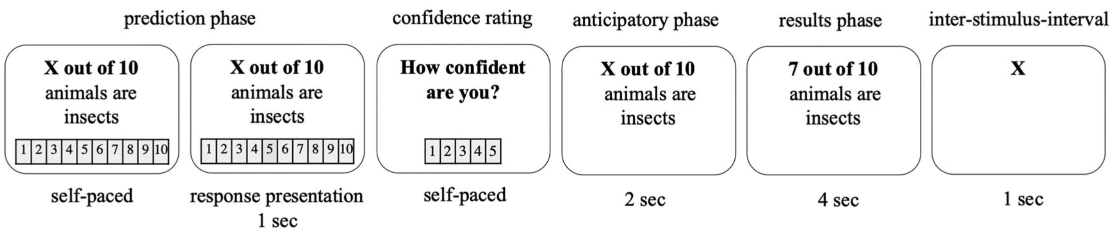

Figure 1. Schematic overview of the study phase. In the study phase, participants predicted the correct value of "X". Then, participants rated their confidence on a 5-point scale ranging from "1" (not confident) to "5" (very confident). Next, participants were presented the initial question again before the correct number was finally presented. The study phase included 90 questions.

after the onset of the results phase relative to the pupil baseline phase (100 ms before onset of results phase until – 200 ms after onset of results phase).

We used the same time windows as in Brod and Breitwieser (2019) to calculate the pupillary marker of surprise. The time window to calculate the pupillary marker of curiosity was similar to but 1 s shorter than the time window used Brod and Breitwieser (2019) because we had a 1 s shorter anticipatory phase in this study. As in Brod and Breitwieser (2019), the time windows to estimate the pupillary markers of curiosity and surprise overlapped for 500 ms during the onset of the results phase. To ensure that the pupillary markers of curiosity and surprise were not confounded, we chose a different pupil baseline phase to calculate the pupillary surprise response and the baseline phase for surprise was inside the time window for curiosity. A within-person correlation analysis revealed a significant but small negative correlation between the pupillary marker of curiosity and surprise (r = -.15, $p < .001$), which provides empirical support that the two measures are distinguishable.

Before data analyses, we conducted outlier analyses for the pupillometric markers of curiosity and surprise. The analysis served to identify trials where the pupil dilation response deviated notably from the rest of the distribution, i.e. more than 3 SD from the average pupil dilation response. We identified 15 outlier trials for the pupillometric marker of curiosity and 55 outlier trials for the pupillometric marker of surprise, which were excluded from further analysis (exclusion of 3% of the data).

Results

Confidence as a predictor of the pupillometric marker of curiosity

We first tested whether confidence predicted an increase in pupil size before the results presentation, thus indicating a higher level of curiosity. Confidence was centred at the person mean to test intra-individual relations. That is, we tested whether being more (or less) confident compared to one's average confidence level predicted curiosity. By centring confidence at the person mean, we also accounted for inter-individual differences in average confidence levels, which may arise from individual response tendencies to generally give high (or low) confidence ratings. We estimated a model with random intercepts only because a model with random slopes for confidence did not provide a better model fit ($\chi^2(5) = 7.84$, $p = .165$). That is, the relation between confidence and the pupillometric marker of curiosity was comparable across participants.

We found evidence for both a linear and a quadratic relation between confidence and pupil size increase (see Figure 2A): while higher confidence ratings (compared to the personal mean) were followed by larger pupil sizes during the anticipation of the result ($\beta = .24$, CI [.21; .27], $p < .001$), the additional quadratic effect indicates that pupil size peaked for confidence ratings that were slightly above the participant's average confidence rating ($\beta = -.10$, CI [-.13; -.07], $p < .001$). Thus, the pupillary data suggest that curiosity peaked when participants were a little more confident in their prediction than usual.

Prediction error and confidence as predictors of the pupillary surprise response

Next, we tested prediction error and confidence (person-mean centred), as well as the interaction between prediction error and confidence as predictors of the pupillary surprise response. To quantify the prediction error, we calculated the absolute difference between the predicted and the correct number. A value of "0" indicated correct predictions and higher values indicated a larger prediction error. The

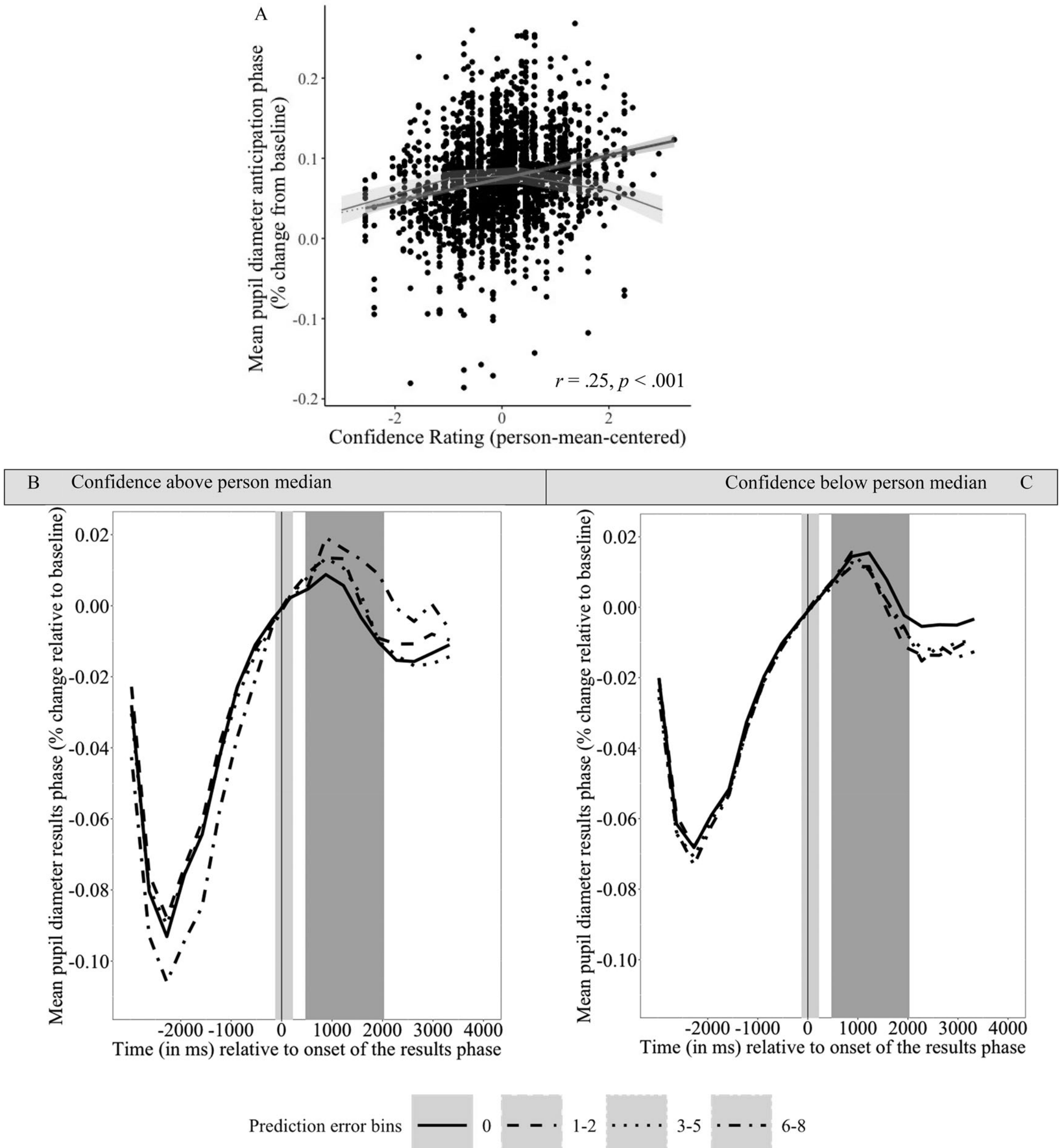

Figure 2. Relation between confidence and pupillometric markers of curiosity and surprise. Figure 2A shows the relation between the person-mean centred confidence and pupillometric markers of curiosity. Each dot represents one trial. Higher confidence predicts an increase in pupil size before results presentation (linear effect: blue line). Pupil size is smaller for predictions made with very high and very low confidence relative to the personal average confidence rating (quadratic effect: red line). Figure 2 B and C show the pupillary surprise response aggregated across prediction error bins and participants for above median confidence (left side) and for below median confidence (right part). The pupillary surprise response depicts the change in pupil size relative to the pupil baseline (light grey area). The pupillary surprise response was calculated for the time interval from 500 ms to 2000 ms after the onset of the results phase (dark grey area). For above median confidence, the pupillary surprise response was larger for large prediction errors compared to small to medium prediction errors and correct predictions. For below median confidence, the pupillary surprise response was larger for correct predictions compared to erroneous predictions.

interaction term was used to test the relation between high-confidence prediction errors and the pupillary surprise response. We estimated a model with random intercepts only because a model with random slopes for prediction error and confidence did not provide a better model fit ($\chi^2(5) = 5.27$, $p = .384$). Thus, the relation between prediction error, confidence, and the pupillary surprise response was comparable across participants. In line with our hypothesis, prediction error and confidence interactively predicted the pupillary surprise response (β = .05, CI [.01; .08], $p = .013$, see supplementary Table 1).

To interpret and illustrate the interaction between prediction error size and confidence, we divided prediction error into four bins: Correct responses (prediction error equals "0"), small prediction errors ("1–2"), medium prediction errors ("3–5"), and large prediction errors ("6–8"). Small prediction errors (on average 52% of trials) and medium prediction errors (on average 28% of trials) occurred more frequently than correct responses (on average 16% of trials) and large prediction errors (on average 4% of trials).

Figure 2 shows the pupillary surprise response for the four prediction error bins, separated out for confidence ratings that were above and below the personal median confidence rating. For high-confidence ratings, the pupillary surprise response was more pronounced after large prediction errors when compared to correct responses and small to medium prediction errors (see Figure 2B). For low-confidence ratings, the pupillary surprise response was more pronounced after correct responses and small prediction errors compared to medium to large prediction errors (see Figure 2C).

We then conducted a post-hoc contrast analysis to test the statistical significance of these descriptive differences. For high-confidence predictions, large prediction errors were associated with a more pronounced pupillary surprise response compared to correct responses (the reference category) (β = .09, CI[.01; .17], $p = .023$, see supplementary Table 2). The pupillary surprise response for high-confidence small and high-confidence medium prediction errors was comparable to that of correct predictions. These results suggest that large, high-confidence prediction errors induce a pupillary surprise response which is more pronounced than that produced by high-confidence correct responses.

For low-confidence predictions (i.e. guesses), correct predictions (β = .51, CI[.07; .96], $p = .022$) and small prediction errors (β = .21, CI[.01; .42], $p = .043$, see supplementary Table 3) were associated with a more pronounced pupillary surprise response than large prediction errors (the reference category). These results suggest that low-confidence correct and almost correct predictions evoke a pupillary surprise response. The pupillary surprise response for low-confidence medium prediction errors was comparable to that of large prediction errors.

In line with our hypotheses, we found that the degree of prediction error and confidence interactively predicted surprise. Larger pupillary surprise responses occurred after large, high-confidence prediction errors and after low-confidence correct and almost correct guesses. These results support the hypothesis that events which strongly violate prior expectations evoke a pupillary surprise response.

Discussion

This study used pupillometry to test how people's confidence in a prediction is related to both their curiosity as well as their surprise in cases of prediction errors. Confidence in a prediction was positively related to pupil size before results presentation, indicating that a higher level of confidence tends to correspond with a higher level of curiosity. An additional quadratic trend indicated that curiosity was largest for confidence ratings that were slightly above the person's average confidence rating. Furthermore, high-confidence prediction errors and low-confidence correct predictions were followed by a pupillary surprise response. This pattern indicates that highly unexpected outcomes induce surprise, independent of the correctness of the prediction. Together, these findings suggest that confidence in a prediction plays an important role in both curiosity and surprise.

The first important finding in this study is that there is a close and intricate relation between confidence and curiosity. We found both a linear and a quadratic relation between confidence and the pupillometric marker of curiosity. The linear increase in the pupillometric marker of curiosity depending on confidence provides evidence for the hypothesis that curiosity increases with confidence (Metcalfe et al., 2020). However, the additional quadratic relation between confidence and the pupillometric marker of curiosity speaks for an inverted u-shaped relation between curiosity and confidence. Our results thus corroborate and extend previous findings by Kang et al. (2009) who found a quadratic relation between confidence and self-reported curiosity. The pupil dilation suggests an increase in physiological arousal when anticipating a rewarding outcome (Schneider et al., 2018), which has been linked to self-reported curiosity (Brod & Breitwieser, 2019; Kang et al., 2009). Together, our results suggest that, in general, curiosity increases with confidence while peaking at confidence ratings slightly above the participant's average confidence rating.

This study's second important finding is that the relation between confidence and surprise differs significantly between correct and incorrect predictions. For prediction errors, *high* confidence was associated with a larger pupil dilation upon seeing the correct result. For correct or almost correct predictions, *low* confidence was associated with a larger pupil dilation response upon seeing the correct result. These results are in line with the definition of surprise as the emotional reaction to a violation of expectations (Itti & Baldi, 2009; Reisenzein et al., 2019). In the case of high-confidence errors, subjects did not expect to be wrong, while in the case of low-confidence correct responses, they did not expect to be correct. Highly unexpected outcomes evoke a pupil dilation response, which indicates enhanced physiological arousal (e.g. Ebitz & Platt, 2015).

The current findings further underline the importance of metacognitive judgements for the creation of curiosity and surprise in the learning process. By making a confidence judgement, learners evaluate their certainty in a prediction (Nelson & Narens, 1990). In doing so, learners often become more aware of their knowledge gap, which boosts curiosity (Loewenstein, 1994). By making confidence judgements, learners also evaluate the strength of their expectations. In line with Bayesian accounts of surprise (Itti & Baldi, 2009), the strength of the prior expectation plays an important role in surprise: Surprise about an unexpected outcome increases with the prior level of confidence in the prediction (Reisenzein, 2000). Hence, confidence may boost curiosity and surprise because it makes knowledge gaps more salient.

The current study has several limitations which offer avenues for future research. First, we assessed changes in pupil size as a marker of the physiological component of curiosity and surprise. Previous studies found evidence for a link between changes in pupil size and various measures of curiosity and surprise (Brod & Breitwieser, 2019; Kang et al., 2009; Reisenzein et al., 2006), although the magnitude and reliability of this correlation is currently unknown. It is unclear, however, whether or not brief increases in physiological arousal are sufficient to infer the feeling of curiosity or surprise. Self-reports may offer insights into the subjectively perceived level of surprise and curiosity. However, self-report data are inherently limited in pinpointing small-scale temporal dynamics. Future research should combine physiological and self-report measures to assess both surprise and curiosity, which have, as of yet, only been conducted separately (Brod & Breitwieser, 2019; Kang et al., 2009; Reisenzein et al., 2006).

Second, we did not experimentally manipulate the size of prediction error nor confidence. Our results are thus correlative in nature. Further experimental studies are required for investigating the causal link between, e.g. high-confidence errors and surprise. In an experimental setting, whether or not the surprising effect of high-confidence errors diminishes over time could also be tested.

Third, future studies could investigate the role of surprise in more complex tasks that involve stronger prior beliefs. For instance, one fruitful area for further research would be investigating the role of surprise about high-confidence errors in conceptual change in the learning process. Surprise about high-confidence incorrect predictions could guide learners' attention to the unexpected information (Theobald & Brod, 2021). Learners may then begin to search for an explanation for the unexpected information (Vogl et al., 2019). Surprise about high-confidence errors may facilitate conceptual change by encouraging learners to pay attention to and search for alternative explanations for the unexpected information.

Conclusion

Our results highlight the importance of feelings of confidence in the occurrence of the emotions of curiosity and surprise. Returning to the example at the beginning, those who knew the answer were probably neither curious about the right answer nor surprised when they heard that oranges are green in tropical regions. The correct answer was, however, awaited with more curiosity and was more surprising for those who were quite sure that oranges are orange, and for those whose random low-confidence guess that oranges are green in tropical regions turned out to be correct. In other words, by making a confidence judgement, people quantify their certainty in their prediction and evaluate the strength of their expectation. Thereby, confidence judgements may increase the salience of a knowledge gap, which boosts curiosity, and may enhance surprise when expectations are violated.

Author contributions

G.B. designed the study. Testing and data collection were overseen by G.B. M.T. performed the data analysis and interpretation with support from G.B and E.G-

K. M.T. drafted the paper. G.B and E.G-K provided revisions. All authors approved the final version of the paper for submission.

Acknowledgements

This research was funded by a grant of the German Research Foundation (DFG, BR 5736/2-1) to GB. GB was supported by a Jacobs Foundation Research Fellowship.

Disclosure statement

No potential conflict of interest was reported by the authors.

Open practices statement

The experiment, the data, and the script that was used to analyse the data are available via the Open Science Framework and can be accessed at https://osf.io/h92gb/.

Funding

This work was supported by Deutsche Forschungsgemeinschaft [Grant Number BR 5736/2-1]; Jacobs Foundation [Jacobs Foundation Research Fellowship].

ORCID

Maria Theobald http://orcid.org/0000-0003-0766-1680

References

Alexander, W. H., & Brown, J. W. (2019). The role of the anterior cingulate cortex in prediction error and signaling surprise. *Topics in Cognitive Science*, *11*(1), 119–135. https://doi.org/10.1111/tops.12307

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, *28*(1), 403–450. https://doi.org/10.1146/annurev.neuro.28.061604.135709

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*(2), 276–292. https://doi.org/10.1037/0033-2909.91.2.276

Breitwieser, J., & Brod, G. (2021). Cognitive prerequisites for generative learning: Why some learning strategies are more effective than others. *Child Development*, *92*(1), 258–272. https://doi.org/10.1111/cdev.13393

Brod, G. (2021). Predicting as a learning strategy. *Psychonomic Bulletin and Review, 28*, 1839–1847. https://doi.org/10.3758/s13423-021-01904-1

Brod, G., & Breitwieser, J. (2019). Lighting the wick in the candle of learning: Generating a prediction stimulates curiosity. *Npj Science of Learning*, *4*(1), 1–7. https://doi.org/10.1038/s41539-019-0056-y

Brod, G., Hasselhorn, M., & Bunge, S. A. (2018). When generating a prediction boosts learning: The element of surprise. *Learning and Instruction*, *55*, 22–31. https://doi.org/10.1016/j.learninstruc.2018.01.013

Butterfield, B., & Metcalfe, J. (2006). The correction of errors committed with high confidence. *Metacognition and Learning*, *1*(1), 69–84. https://doi.org/10.1007/s11409-006-6894-z

Ebitz, R. B., & Platt, M. L. (2015). Neuronal activity in primate dorsal anterior cingulate cortex signals task conflict and predicts adjustments in pupil-linked arousal. *Neuron*, *85*(3), 628–640. https://doi.org/10.1016/j.neuron.2014.12.053

Gruber, M. J., Gelman, B. D., & Ranganath, C. (2014). States of curiosity modulate hippocampus-dependent learning via the dopaminergic circuit. *Neuron*, *84*(2), 486–496. https://doi.org/10.1016/j.neuron.2014.08.060

Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research*, *49*(10), 1295–1306. https://doi.org/10.1016/j.visres.2008.09.007

Joshi, S., Li, Y., Kalwani, R. M., & Gold, J. I. (2016). Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron*, *89*(1), 221–234. https://doi.org/10.1016/j.neuron.2015.11.028

Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T., & Camerer, C. F. (2009). The wick in the candle of learning. *Psychological Science*, *20*(8), 963–973. https://doi.org/10.1111/j.1467-9280.2009.02402.x

Krüger, M., Bartels, W., & Krist, H. (2019). Illuminating the dark ages: Pupil dilation as a measure of expectancy violation across the life spanilluminating the dark ages: Pupil dilation as a measure of expectancy violation across the life span. *Child Development*. *Advance on*, 1–16. https://doi.org/10.1111/cdev.13354

Lawson, R. P., Bisby, J., Nord, C. L., Burgess, N., & Rees, G. (2021). The computational, pharmacological, and physiological determinants of sensory learning under uncertainty. *Current Biology*, *31*, 163–172. https://doi.org/10.1016/j.cub.2020.10.043

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin*, *116*(1), 75–98. https://doi.org/10.1037/0033-2909.116.1.75

Metcalfe, J., Schwartz, B. L., & Bloom, P. A. (2017). The tip-of-the-tongue state and curiosity. *Cognitive Research: Principles and Implications*, *2*(1), 1–8. https://doi.org/10.1186/s41235-017-0065-4

Metcalfe, J., Schwartz, B. L., & Eich, T. S. (2020). Epistemic curiosity and the region of proximal learning. *Current Opinion in Behavioral Sciences*, *35*, 40–47. https://doi.org/10.1016/j.cobeha.2020.06.007

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *In The psychology of learning and motivation*, Vol. *26*(Issue C), 125–173. https://doi.org/10.1016/S0079-7421(08)60053-5

Reisenzein, R. (2000). Exploring the strength of association between the components of emotion syndromes: The case of surprise. *Cognition and Emotion*, *14*(1), 1–38. https://doi.org/10.1080/026999300378978

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*(2), 295–315. https://doi.org/10.1037/0022-3514.91.2.295

Reisenzein, R., Horstmann, G., & Schützwohl, A. (2019). The cognitive-evolutionary model of surprise: A review of the evidence. *Topics in Cognitive Science*, *11*(1), 50–74. https://doi.org/10.1111/tops.12292

Schneider, M., Leuchs, L., Czisch, M., Sämann, P. G., & Spoormaker, V. I. (2018). Disentangling reward anticipation with simultaneous pupillometry / fMRI. *NeuroImage*, *178*, 11–22. https://doi.org/10.1016/j.neuroimage.2018.04.078

Theobald, M., & Brod, G. (2021). Tackling scientific misconceptions: The element of surprise. *Child Development*, *00*, cdev.13582. https://doi.org/10.1111/cdev.13582

Vogl, E., Pekrun, R., Murayama, K., Loderer, K., & Schubert, S. (2019, November). Surprise, curiosity, and confusion promote knowledge exploration: Evidence for robust effects of epistemic emotions. *Frontiers in Psychology*, *10*, 2474. https://doi.org/10.3389/fpsyg.2019.02474