

**How Does Uncertainty About One's Affective Forecasting Correspond to
Prediction Accuracy?**

Henrike Nebel

S5153603

Department of Psychology, University of Groningen

PSB3E-BT15: Bachelor Thesis
Group 26

Supervisor: Fridtjof Petersen

Second evaluator: Dr. Oliver Weigelt

In collaboration with: Fabienne Liepelt, Marieke Boerendonk, Martino Avalle, Miora Haslacher.

July 02, 2025

A thesis is an aptitude test for students. The approval of the thesis is proof that the student has sufficient research and reporting skills to graduate, but does not guarantee the quality of the research and the results of the research as such, and the thesis is therefore not necessarily suitable to be used as an academic source to refer to. If you would like to know more about the research discussed in this thesis and any publications based on it, to which you could refer, please contact the supervisor mentioned

Abstract

Human affective forecasting - the ability to predict future emotions - has been consistently shown to be flawed. These prediction errors have detrimental consequences for our behavior, affecting one's health and well-being. Past research explaining this phenomenon emphasized the impact bias: the tendency to overestimate intensity and durability of future emotions. A related concept in the human judgment domain is overconfidence: a general tendency to overestimate one's judgement. Given the similarity of both concepts, we hypothesized similar results for our affective forecasting domain: a tendency of individuals to display overconfidence in their predictions. On an individual level, geopolitical forecasting shows that individuals displaying more certainty were more correct in their predictions than uncertain individuals. We hypothesize this trend to transfer to our study: higher certainty is associated with an increase in prediction accuracy. We used the experience sampling method (ESM) - a longitudinal research design implemented in participants daily life. This study newly introduced the interval measurement to assess uncertainty and quantify overconfidence. Throughout a period of 14 days, participants indicated a point and interval prediction five times a day. Results support our hypotheses, indicating general overconfidence in prediction: instead of expected 95%, participants' prediction intervals contained the true value in 64.1% of all cases. Furthermore higher certainty was associated with higher prediction accuracy. Our results introduce overconfidence as a new predictor for affective forecasting accuracy with possible implications for intervention aiming to reduce prediction errors.

Keywords: Affective Forecasting, Overconfidence, Overestimation, Certainty, Prediction, Impact Bias

How Does Uncertainty About One's Affective Forecasting Correspond to Prediction Accuracy?

It might be of no surprise how human life is accompanied by a range of emotions. We may experience happiness thinking about the upcoming vacation, anxiety about an upcoming exam or dreading work because of an annoying coworker. The idea that emotions serve an adaptive purpose has long been explored, dating back to early theories by William James and Carl Lange (Coleman & Snarey, 2011), which emphasized the profound influence emotions have on our behavior. Given how frequently we experience and reflect on emotions, one might assume we are skilled not only at managing our present emotional states but also at anticipating future ones, which is known as affective forecasting (Wilson & Gilbert, 2003). Yet, empirical evidence tells a different story: humans consistently struggle with accurately predicting how they will feel in the future (Pilin, 2020).

This forecasting error has serious implications for people's lives that range from physical health behaviors (e.g., reduced exercise, Loehr & Baldwin, 2014) to their subjective well-being (Bertoni & Corazzini, 2018). In fact, a study by Chapman and Coups (2006) investigated preventive health behaviors in the context of vaccination. Results indicate that individuals' choice to receive a flu vaccination was largely determined by their predicted emotions. Specifically, participants were largely concerned with how much worry and regret they would experience if they were to get sick. Interestingly, they based their decision to a lesser extent on how protective they thought the vaccination would be.

This demonstrates how the prediction error does not solely influence cognition but further impacts decision making. In accordance with this finding, individuals also choose options which provide the largest anticipated satisfaction (Mellers et al., 1999). The influences of our prediction on behavior could be detrimental in that individual's exaggeration of their satisfaction or dissatisfaction towards an event could lead to fearlessness and

disproportionate carefulness, respectively (Mellers & McGraw, 2001). Individuals with social anxiety for example, act upon their predicted negative emotions by avoiding situations all together (Hall et al., 2018). This avoidant behavior may not only impact one's personal life but society as a whole. Research indicates that affective forecasts can contribute to racism. Mallet et al., (2008) found an outgroup bias. That is white individuals predicted to feel worse when engaging with a Black participant than they eventually did. These wrong predictions may lead to a general avoidance of engaging with individuals across groups (Kushlev & Dunn, 2012). Overall, these findings demonstrate the great impact impaired affective forecasting has on our cognition and subsequent behavior. To counteract those negative effects, one needs to understand why and how they occur.

Generally, people can be wrong in their predictions on various dimensions ranging from valence, emotion type, strength to duration (Wilson & Gilbert, 2003). Whereas individuals are typically well-equipped to estimate the valence (i.e., positive, or negative charge) of their future emotions, findings are more inconsistent for estimation of emotion type. Latter dimension of emotion type refers to the exact emotion one can identify (e.g., happiness, anxiety). Wilson and Gilbert (2003) suggest a hypothesis explaining why individuals might have problems predicting these specific emotions. They state how individuals may generally estimate concrete emotions in an oversimplified way. This conflicts with reality which usually incorporates a range of interacting emotions. This complexity requires a more elaborate consideration than only focusing on either positive or negative feelings and more on interactions. However, whereas this explanation is supported by some research (i.e., superficial estimation of emotions for remote future, Liberman et al., 2002) other findings suggest opposing results: demonstrating how individuals are well equipped to consider complex interacting emotions (Wilson et al., 2002, as cited in Wilson & Gilbert, 2003). Hence research is inconclusive about humans' ability to predict specific emotions.

Focusing on the remaining two dimensions, accuracy is also diminished for predicting the strength and length of emotional experience (Wilson & Gilbert, 2005). Although people sometimes underestimate -instead of overestimate- their emotional experience, the susceptibility to overestimate the length and strength of emotions is seen more frequently. This phenomenon is formally known as impact bias (Wilson & Gilbert, 2005). Research on affective forecasting has focused on this bias extensively as a major source of prediction error (e.g., Buehler & McFarland, 2001). Underlying the prediction errors regarding emotional strengths and durability are two biases which may inform us about specific underlying mechanisms.

The first of those two biases is called durability bias and concerns people's tendency to overestimate the time they will experience forecasted emotions (Gilbert et al., 1998). If a person predicts to feel bad because of failing an exam, they tend to overestimate how long this feeling endures. One cause explaining this biased cognition is called immune neglect: a failure to consider one's psychological immune system (Gilbert et al., 1998). That is individuals are generally thought to utilize a defense system which safeguards them from negative experiences (i.e., by using self-serving biases). However, when predicting their emotions, individuals may be unaware of these processes (e.g., cognitive dissonance in Snell et al., 1995), causing them to overestimate the durability of their negative emotions. On the other hand, the second bias reflecting a possible source for erroneous forecasts is called intensity bias. Buehler and McFarland (2001) indicate how participants overestimated their emotional strength to future events. This reaction was associated with a simplistic thinking. Specifically, participants centered their prediction on the predicted event to the extent that they disregarded other potential influences, which led to an increased intensity bias.

Concluding, these processes fall under the umbrella of impact bias. In turn this might suggest a more general cognitive tendency, we call emotional overconfidence. Hence, we

theorize that individuals may be overly certain in their prediction of future affective responses. As for the other biases, overconfidence may occur because individuals disregard their emotional defense systems (i.e., emotional immune system). We investigate this concept, relying on the broader human forecasting and judgment research (e.g., Palmer, 2000; Christensen et al., 2018). Overall, the existing literature might be beneficial in obtaining a comprehensive overview and extending research on forecasting error in the affective forecasting literature.

Overconfidence

In human forecasting literature, someone is overconfident when one's certainty about the accuracy of their prediction goes beyond the actual degree of accuracy of the prediction (Klayman et al., 1999). Moore and Healy (2008) further differentiate between three subtypes of overconfidence: overprecision, overplacement, and overestimation. Overprecision refers to undue certainty in one's judgments, while overplacement describes the tendency to think that one is superior to others. In the following, we will mainly focus on overestimation, which incorporates people's tendency to exaggerate their competencies, accomplishment or mastery of situations. Consequently, overestimation will present as inflated confidence in forecasting performance.

Perhaps not surprisingly, research in the human judgment literature reveals overconfidence to be a frequent problem (e.g., Brenner et al., 1996). Hence, people are said to be poorly calibrated in their predictions. In statistics, the term calibration describes how well one's predicted probabilities align with the observed outcome. Thus, someone who indicates "70%" confidence should be accurate in their predictions in "70%" of all cases (Gneiting et al., 2007). In addition to being well calibrated however, good forecasters should also provide sharp estimates. Sharpness hereby refers to how precise the estimates are. A person saying, they are "70%" confident of a specific outcome indicates a sharper prediction than one

indicating that they are “40%” sure about the outcome. Having stated the prerequisites for good forecasting, past research tried to entangle the possible predictors for good calibration and sharpness.

In the past, a widely cited mechanism explaining overconfidence is called hard - easy effect, demonstrating how people’s confidence tends to grow further above their actual accuracy as tasks become more difficult (Gigerenzer et al., 1991). However, research reveals how this explanation might be overly simplistic, not considering important variables such as measurement errors and response format (Klayman et al., 1999). Indeed, research in the geopolitical forecasting domain shows how as one’s certainty increases, so does the accuracy of their predictions (Moore et al., 2016). This effect has been consistent over a three-year time and might suggest how individuals are generally self-aware when it comes to estimating their abilities.

Similarly, Klayman et al. (1999) found results challenging the hard easy effect: the tendency of overconfidence was not influenced by the difficulty of questions within one subject. In their study, two aspects that influenced individuals’ overconfidence were a) the difficulty, not within but across different fields of questions as well as b) response formatting (i.e., binary vs confidence range). Unfortunately, there is no evidence explaining why some fields are associated with an increase in overconfidence and others are less. However, for the area of affective forecasting, subject of the question will remain the same (i.e., asking about emotions) and therefore less important. On the other hand, Klaymen et al. (1999) found results indicating increased overconfidence for questions utilizing confidence range response formats compared to binary choice questions. When using range question formats, individuals are thought to fixate on their prior beliefs about the question. In comparison, a binary format provides participants with different answer possibilities that might make them reflect on their prior beliefs. Imagine someone has a prior belief that about 30.000.000 people lived in

France. However, now they get exposed to the questions: how many inhabitants are living in France: 68.000.000 or 50.000.000. Because they get confronted with different numbers, which are in stark contrast to their prior belief, they might reflect more on their accuracy of their answer. On the other hand, asking them about a possible range that includes the true population value, will only confirm them in their choice (i.e., there is no comparison information) (Klaymen et al., 1999). Overall results indicate that individual calibration depends less on the difficulty of questions and more on reflection processes.

Beyond its direct impact on certainty however, the response format has serious implications on measurement decisions. As stated before, in classic human judgment research, participants can usually choose between two answers (e.g., Moore et al., 2016). The construct of confidence then gets assessed by comparing the frequency of their successes (i.e., accurate responses) to the actual event rates. However, after extensive research, we found that the affective forecasting research based their prediction measures on point estimates. These response formats do not provide any information about the level of confidence on their own and require an additional explicit measure of certainty. This shortcoming encouraged us to ask for interval prediction which can demonstrate individuals' confidence range in relation to their estimate. Taken together, whereas point estimates allow for a direct measure of accuracy (Chatfield, 2001), they do not provide information about an individuals' certainty. To investigate confidence as a predictor, interval predictions yield data demonstrating the degree of calibration and sharpness for continuous outcomes.

While measurements in the form of intervals are not implemented in the affective forecasting field, research in other fields shows that they are comprehensible. Research by Grounds et al. (2017) found that misinterpretations of prediction intervals occurred at similar rates across populations with varying educational backgrounds. Similarly, a study on weather forecasting by Joslyn et al. (2013) demonstrated that nonprofessionals could accurately

understand and interpret graphical representations of prediction intervals. Together, these findings suggest that, despite their complexity, prediction intervals remain accessible and comprehensible to diverse audiences, and might also be useful to directly measure overconfidence in affective forecasting.

Current Study

By now the gravity of the prediction error in affective forecasting should have become apparent. Understanding the mechanisms that underly this prediction error is therefore crucial: False predictions of emotions are not merely another cognitive bias but may negatively influence our behavior, health, and social interaction (Loehr & Baldwin, 2014; Mallet et al., 2008). One could theorize how the impact bias, which has been studied as one of the most prominent causal factors, might reflect overconfidence in one's own ability to predict emotions. However, overconfidence as a predictor of forecasting errors has primarily been studied in the human judgment literature. Our study now aims to fill this research gap, integrate this promising predictor in the affective forecasting literature. More importantly for the first time, our study directly tests overconfidence with respect to calibration using interval measures. We aim to investigate this association using the experience sampling measurement which has been implemented by Takano et al. (2023) to measure emotional predictions in daily life. ESM is a longitudinal research methodology obtaining personal data from participants throughout their day allowing for repeated measures (Myin-Germeys et al., 2009 as cited in Fritz et al., 2024). Using a convenience sample of students we gather data over a period of 14 days, investigating two main hypotheses. Based on the general trend in human judgment research, we first hypothesize that there is a general trend of overconfidence when it comes to making affective predictions. Someone is thought to be overconfident if their level of certainty about the accuracy of their prediction exceeds the actual accuracy of their forecast. Subsequently, we aim to compare the relation between prediction error and interval

width of our participants. (Over)confidence will be assessed through the match between our nominal coverage of intervals and participants' empirical coverage of the actual predicted value. Our exploratory research further investigates potential differences in overconfidence between positive and negative affect. Based on findings in the geopolitical literature focusing on the direct association between certainty and prediction error, we hypothesize that more certainty will be associated with more accurate predictions. Differences in affect variables are of interest for exploratory intentions.

Methods

Our quantitative observational study aimed to conceptually replicate the results of Takano and Ehring (2023) by investigating affective forecasts using ESM data. We extended their work by including a two-step ahead prediction as well as an interval prediction. In this specific paper, we first investigated how well the affective forecasts of participants are calibrated on average (i.e., overconfidence or underconfidence). Lastly, we tested if the different levels of uncertainty (or confidence) are associated with the magnitude of the prediction error in affective forecasts.

Participants

This research project is an experience sampling study about affective forecasting. That is, participants predict and report their real-time emotional experiences through their mobile device. To determine forecasting accuracy, we calculated the prediction error contrasting the forecast of a specific point in time with the corresponding actual value. Based on Takano and Ehring (2023) we aimed for a sample size of 68. The final sample consisted of 30 first year psychology students from the University of Amsterdam (20 women, 10 men) with a mean age of 19.97 ($SD = 1.83$) (Table B1).

Procedure

To be eligible for this research project, students had to own a smartphone, understand the English language and not have been diagnosed with depression or anxiety. The study was advertised via flyers on campus, social media, and the student participant pool. Therefore, we used a convenience sample for our study. Students receive course credits for participating. Before obtaining our data, our study was ethically approved (FMG-12534_2025). Consent was received through a SONA Qualtrics questionnaire, which also included instructions and upon enrollment participants received a link to m-Path questionnaire via email. Data was collected using ESM via the m-Path app (Mestdagh et al., 2023), which allows participants to complete short questionnaires on their phones while going about their daily activities. Participants received five prompts per day over a 14-day period to complete the EMA questionnaire. These were sent at fixed times, starting at 9:00 am, and then every three hours until 9:00 pm. The questionnaire takes around 5-10 minutes to complete. Upon receiving a notification, participants had a 30-minute time window to respond before the prompt expired. The items were presented in the same order for all participants.

Materials

The ESM survey consisted of three repeated items, asking about three different points in time (i.e., present emotion, three hours ahead, six hours ahead). There were four items assessing distinct emotions: two concerning positive affect (happy and relaxed) and two concerning negative affect (sad and anxious). Based on past research on core affect, these emotions vary along the dimensions of reactivity and valence (Russell, 2003). In accordance with ESM research, the participants rated each emotion and prediction on a continuous visual analogue scale ranging from 0 (i.e., = *not at all*) to 100 (i.e., = *extremely*). We based the assessment of each emotion on two classes of items and adjusted them according to the different time points. One type of item asked about the current emotional state ("Please rate how much you are currently experiencing each emotion.") and another concerned the interval

prediction of future emotional strength ("I am 95% sure that my emotion level will be at least /at most..."). We accounted for the different time points by changing the wording accordingly ("Please estimate how much you will experience each emotion at [time of next beep]."). This was done for both measurements of a) point prediction ("My best guess is...") and the b) credible interval prediction ("I am 95% sure that my [emotion] level will be at least [at most]..."). As most ESM questionnaires, our survey items are not validated by research (Wright & Zimmermann, 2019). However, phrasing and implementation is based on the replicated study of Takano and Ehring (2023).

Data Preprocessing

Data collection methods

Data preprocessing was performed using R Statistical software (v4.1.2; R Core Team, 2021), and participants with less than 30% compliance (i.e., 21 beeps) were excluded from the dataset to ensure consistency with the data from Takano and Ehring (2023). The overall completion rate of the questionnaires in our dataset is 73.99% ($SD = 18.42\%$), with an individual compliance range of 31.43% to 95.71%. Moreover, 30 of the total number of participants (74) filled out 21 or more questionnaires. We transformed the raw data to answer our research question. The interval width variable was calculated by subtracting the lower bound value from the upper prediction interval. Subsequently, the prediction error was calculated for both affect variables by subtracting the point prediction of affect from the actual observed affect at that time point. Since we obtained multiple observations per individual, variables were then aggregated by person. For negative and positive affect, we created a variable representing the empirical coverage. First, we created binary variables for successes (= 1) if the point estimate lies within the upper and lower bound of the interval prediction and failures (=0) if it falls outside. Successes added up and transformed into a new

variable for each affect type and finally divided by the number of total trials to obtain a variable representing the empirical coverage for further analysis.

Statistical Analysis

We tested the main hypothesis (i.e., individuals generally tend to be overconfident in their prediction) by comparing the alignment between participants' empirical coverage rates and the nominal 95% coverage rate, using a binomial test in JASP (Version 0.19.3; JASP Team, 2024) and R (v4.1.2; R Core Team 2021). Successes were counted whenever the actual value was included in the predicted interval (i.e., final variable explained in data processing). Subsequently the difference between successes and the nominal coverage of 95% was tested for significance. If the empirical coverage matches the indicated nominal 95% interval, participants are generally well calibrated (i.e., a non-significant result). In cases where the empirical coverage rate a) falls below the nominal 95% rate and b) this difference is statistically significant (i.e., 0.05 threshold), one is considered to be overconfident. If the empirical coverage rate a) is above the nominal 95% rate and b) the difference is statistically significant, one falls into the underconfident category. In addition, we checked for outliers and assumption of a binomial tests (i.e., binary dependent variable, no dependence of observation, fixed n , equal probabilities to obtain same result in Statistics Solutions, 2025). Additionally, we investigated the exploratory hypothesis, that individuals will differ in their certainty depending on the valence of their emotion. The two separate binomial test results for both affect variables were compared to one another.

Due to our repeated measure design, we obtained multiple observations per individual which are not independent of one another (i.e., random effects). To account for these characteristics of our data, we used a linear mixed model (LMM) on JASP and R when testing the hypothesis that greater uncertainty would be associated with an increased prediction error (López et al., 2022). Specifically, this LMM aimed to investigate the association of the

dependent variable prediction error with the independent variables of interval width and affect type width. Random effects were assessed using ParticipantID, accounting for within person variability. Due to severe assumption violations for our LMM, we used an additional model: the generalized linear mixed model (GLMM). The GLMM neither assumes normality nor homoscedasticity and could therefore demonstrate if our findings are robust (Ng & Cribbie, 2016). After checking for assumptions of the GLMM, we ran the model on JASP, using the gamma distribution, predicting the prediction error with the two independent variables of interval width and affect type. We used different R packages for statistical analyses (v4.1.2; R Core Team 2021). To aggregate and manipulate data the tidyverse package (v2.0.0; Wickham et al., 2019) was used. Different visualization packages included ggplot2 (v3.5.2; Wickham, 2016), lattice (v0.22.6; Sarkar, 2008) and sjplot (v2.8.17; Lüdtke, 2024). For inferential statistics among the linear mixed model, lme4 (v1.1.37; Bates et al., 2015), lmerTest (v3.1.3; Kuznetsova et al., 2017) andggeffects (v2.2.1; Lüdtke, 2018) were implemented.

Results

Preliminary Analysis

The subsequent analysis relied on data of 30 participants using three main variables (i.e., interval width, prediction error, affect type) (Table 2). On average individuals had a similar prediction error for both positive ($M = 11.69$, $SD = 10.53$) and negative affect ($M = 11.28$, $SD = 10.90$). The range of prediction error and maximum values were higher for the negative affect variable. This goes along with the interval width variable where participants had wider intervals for negative affect predictions than for positive affect predictions.

Table 2

Descriptive Statistics: Main Variables

M	SD	Min	Max
-----	------	-----	-----

Absolute Prediction				
Error				
PositiveAffect	11.69	10.53	0	72.00
NegativeAffect	11.28	10.90	0	91.50
Interval Width				
PositiveAffect	25.89	11.99	0	79.00
NegativeAffect	27.13	17.13	0	90.50

Assumptions for binomial tests were checked and met. That is the output variable is binary (i.e., creation of empirical coverage variable) and there were equal chances for each of those outcomes to occur. After aggregation of our data, observations for each individual were independent and we worked with a predetermined sample size (McClenaghan, 2024).

Following, assumptions for the mixed model analysis were checked (Ushakova & Waterston, 2019). Namely these were linearity of observation, normality of residuals, homoscedasticity of variance as well as multicollinearity.

To check if our variables are linearly related to one another, we used a scatter plot (i.e., residuals vs fitted values) to investigate the pattern of observations. Figure B1B displays a slight cone shaped pattern of our data, indicating violation from our linearity assumption of our model. We checked for normally distributed residuals, using a Q-Q plot (i.e., residuals vs quantiles of standard normal distribution). The residuals failed to align with the theoretical normally distributed quantiles, showing an upward curved pattern and violating our normality assumption (Figure B1D). To check for equal variances of residuals, we used a scatterplot (i.e., residuals vs fitted values), inspecting the scatter of our observations. Our plot showed a cone shaped pattern, indicating different values for our variance: our assumption of homoscedasticity is violated (Figure B1C). Further analysis of outliers indicates slight

concerns for various participants with Cook's distance greater than 1 (Figure B2). Results of the variance inflation factor show no concern for multicollinearity regarding variables of affect type and interval width but slight inflation of the interaction variable ($VIF > 5$) which requires caution (Table B3).

Due to the overall assumption violations, we used a generalized linear mixed model in addition to investigate the robustness. Assumptions of homogeneity of variance, uniformity of residuals as well as normality of random effects were met (Figure B3). Specifically, observations in the residual plot checking for homogeneity showed no clear pattern or systematic shape as with the LMM: indicating homoscedasticity of variance for our data (Figure B3B).

Main Analysis

We investigated if there is a general overconfidence trend using a binomial test: testing whether the empirical coverage matches the nominal 95% interval. The proportion of participants successfully predicting their emotion within the indicated interval differed significantly ($P = 0.641$; $p < .001$, $CI [0.62, .66]$) from the expected coverage ($P = 0.95$). Participants' intervals contained the true value in just 64.1% of trials, far below the nominal 95%, demonstrating overconfidence (i.e., for a graphical representation of the individual coverage in relation to the nominal coverage see Appendix B Figures 4 and 5). On average, individuals were more certain about their predictions, compared to the actual lower accuracy. Specifically, their interval estimation was only accurate in 64.1 % of cases compared to required 95%. We also conducted an exploratory analysis to determine if there was a difference in overconfidence between positive and negative effects. This involved using two additional binomial tests. Results show significant differences between the empirical coverage of negative affect ($P = .64$; $p < .001$, $CI [.61, .67]$.) and expected coverage as well as for positive affect ($P = .64$; $p < .001$, $CI [.61, .67]$). There is no difference between the proportion

of successes between both affect variables. Consequently, individuals display the same degree of overconfidence when predicting positive and negative emotions.

In addition, we investigated the impact of certainty on prediction accuracy: we tested if more uncertainty was associated with larger prediction error using a linear mixed model. Whereas interval width ($b = 0.09$, $t(888) = 5.18$, $p < .001$) emerged as a significant predictor of prediction error, affect type ($b = -.84$, $t(2300) = -1.78$, $p = .076$) did not. The positive value for the slope of the interval width indicates that more uncertainty is associated with less accuracy in affective forecasting. In contrast to their single contribution, the interaction between the two main effects (i.e. interval width x affect type) was not found to be significant ($b = .02$, $t(2252) = 1.12$, $p = .264$). Investigating the robustness of our data with the generalized mixed model, we can detect some differences. The general significant effect on interval width on prediction error remains constant. However, both affect type ($b = -.05$, $t = -2.69$, $p = .007$) and the interaction of affect type with interval width ($b = .04$, $t = 2.01$, $p = .045$) show statistically significant values for the GLMM. A plot demonstrates this interaction effect for our GLMM: For predictions of negative affect type, smaller intervals were associated with lower prediction error than for positive affect type (Figure B6). This difference disappears as uncertainty increases (i.e., wider intervals). These results might be due to a better model fit for the data. However, differences of significant levels for different variables between both models need to be treated with caution, especially considering the relatively high p values.

Discussion

We conducted our ESM study to investigate the underlying factors contributing to individuals' diminished ability to accurately predict future emotions. Based on past research of cognitive biases (i.e., impact bias), we investigated a) whether participants displayed a tendency to be under or overconfident when forecasting their emotions and b) whether the

associated prediction uncertainty has an impact on the prediction error. Results of our analysis support our hypothesis of an overall overconfidence regarding affective forecasting. Our overall sample displayed a general impaired prediction ability: Participants frequently overestimated their prediction accuracy. Interestingly, our results suggested that on an individual level, participants who were more certain were also more accurate in their predictions. This suggests that our overall self-judgment might be more accurate overall.

Results of our exploratory analysis, focusing on the difference between positive and negative affect for prediction accuracy, is less consistent. There was no difference between the general tendency of overconfidence for predicting negative or positive affect. However, we did find significant results indicating how certainty might predict forecasting error differently for both affect variables: When predicting negative affect, participants who were more certain had a decreased prediction error compared to when they predicted positive emotions. However, as uncertainty increased this difference of prediction accuracy between positive and negative affect vanished. This effect should be treated as a careful pattern rather than a definite effect since statistical significance was inconsistent across both of our models and provided relatively large significant statistics. Generally, however, we might suspect significant results of the GLMM to be more reliable, as our data had a better fit with this model than the LMM (i.e., assumption violation).

Although there is a lack of research for investigating overconfidence in the affective forecasting literature, most research in the field of human judgment and forecasting is in line with our results (see; Brenner et al., 1996; Feiler & Tong, 2021). In the human judgment field Pallier et al. (2002) present three possible models explaining the general tendency of overconfidence. Among those, a) heuristic and biased cognition and b) individual differences provide explanation in the affective forecasting area. First participants' personality characteristics might have influenced their overconfidence and consequently prediction

accuracy. Specific personality traits may lead to a higher risk of overconfidence. Research demonstrated that extraversion (Schaefer et al., 2003) as well as narcissism (Campbell et al., 2004) were correlated with increased overconfidence for human judgment.

However, although these might predict when overconfidence occurs, they do not explain why we tend to overestimate our emotions. When focusing on biased cognitive processes, the affective state during decision making was important in predicting overconfidence: positive feelings increased likelihood of individuals to overestimate (Koellinger & Treffers, 2015). This effect can be explained by the affect-as-information hypothesis, whereby individuals base their decisions on their current emotional state (Clore & Huntsinger, 2007). Overconfidence occurs when they are not aware of what caused their current positive emotional state. This finding corresponds to the earlier mentioned impact bias (i.e., overestimation of strength and durability of feelings), whereby people believe future events to be more influential for their emotions than they actually are (Wilson & Gilbert, 2003). For both biases individuals do not consider or overestimate relevant cues important to predict emotions. Indeed, Koellinger and Treffers (2015) suggests if individuals are conscious about a) their feelings and 2) their causes, their reflection will help them make better judgements. By providing their current emotional estimate, participants in our study might have been made overly aware of their positive emotions. This positive emotion in turn might have led to overconfidence (i.e., providing narrower intervals) for subsequent predictions: they attributed their salient positive emotions to their forecasting ability (i.e., whereas the positive emotion should have been attributed to the task at hand). However further research needs to investigate how the influence of current emotions on future emotions might be explained by overconfidence. Specifically, regarding our ESM study design, it might be interesting to compare the participants' current emotional value to the interval width given at this time point.

However, our study also demonstrated how increased certainty predicted more accurate forecasting. Research in the field of ability assessment confirms that students' self-evaluation might generally correlate highly with their actual performance and show high reliability (Darrow et al., 2002). Another study showed how participants are generally better calibrated when their accuracy increases. However, this relationship changed at a specific cutoff of 80% accuracy, where good accuracy was associated with more uncertainty (i.e., underconfidence) (Lichtenstein & Fischhoff, 1977). Although these studies were not directly focusing on affective forecasting, it might offer potential clues to explain our results (i.e., accurate self-evaluation). Indeed, their study might explain our interaction effect. Although we did not find underconfidence, we also observed a cutoff point, when the improved prediction accuracy predicted by more certainty was no longer more beneficial for negative than positive affect. Whereas Lichtenstein and Fischhoff (1977) and our study may show a similar trend, further research needs to gather empirical evidence to draw optimal conclusions. It might be interesting to explore if overconfidence develops in a specific pattern instead of simply fitting a predefined model. Having investigated the cutoff point, the general interaction between positive and negative prediction accuracy might be explained by the negativity bias: whereby individuals focus on negative aspects more than on positive (Lazarus, 2021). This distorted view might be hypothesized to shift awareness and make individuals more attentive to negative cues, leading to more certainty and eventual accuracy.

Taken together, these results suggest a new predictor of affective forecasting accuracy: forecasting uncertainty. This might inform research investigating the various negative consequences that occur due to forecasting errors. Interventions might benefit from focusing on reducing people's uncertainty to help reduce the overall prediction error. Since research suggests how affective forecasts influence cognition (i.e., intention) and subsequently behavior (Rizeq, 2024), these interventions might directly challenge decision making. On a

more practical level, this new predictor might be especially important in the medical field. When relying on their false forecasts, patients might follow catastrophic thinking or ignore possible risks (Rhodes & Strain, 2007). Indeed, a metaanalysis of found promising results of past medical interventions targeting affective forecasting. Those interventions were associated with an increase in the behavior of interest. Such interventions who in the past focused among other aspects on anticipated regret or emotional intelligence (e.g., Abraham & Sheeran, 2004; Peter & Brinberg, 2012), might now target individuals' prediction uncertainty. Rhodes and Strain (2007) suggest how patients might overestimate (i.e., be overly confident towards) the extent and length of medical consequences and dismiss their coping resources. Decreasing overconfidence or guiding them towards better calibration might facilitate them developing more accurate beliefs about health behavior.

Strengths and Limitations

One limitation of our study concerns cultural variability. Our participants most likely belong to the WEIRD (i.e., western, educated, industrialized, rich, democratic, Henrich et al., 2010) population. Generally, research suggests a difference in overconfidence (i.e., in human judgment literature) across cultures, whereby Chinese cultures display the most pronounced overconfidence (Yates et al., 1998). Although other studies show less consistent evidence for such differences between cultures, these findings suggest special caution. Future research should focus on why these differences arise and if they are a result of measurement problems. In this case, measurement for different populations needs to be adapted or interventions adjusted.

Research suggests that individuals are generally well adjusted to understand the use of interval questions (Joslyn et al., 2013). However special consideration needs to be considered for the phrasing of interval questions. Providing an interval of 95% coverage might be quite artificial for some participants. Indeed, Bolger and Önköl-Atay (2004) shows that people

showed less over or under confidence when given feedback about their current calibration. This enhancement might have been due to participants simply not understanding how confidence relates to interval range. That is instead of being less overconfident, participants only stated larger intervals while maintaining or increasing their initial confidence level: leading to an increased success rate. We thereby need to ask ourselves, if the indicated interval rate was an accurate measure of certainty and represents real calibration. It might therefore be useful to obtain an additional point estimation of participants' certainty, which can be compared to the interval construct (i.e., success rate of including true value in interval).

Despite these limitations however, our study had two major strengths. First, we aimed to close a gap in research investigating overconfidence in a new field of psychology: affective forecasting. As one of the firsts in the affective forecasting field, we also measure overconfidence using interval predictions. This allowed for more sensitive estimates of our construct and might inform future research about the benefits compared to point predictions (Chatfield, 2001). Using ESM, we not only obtained better ecological validity but also temporal precedence (Shiffman et al., 2007). Our data is not based on data assessed in an artificial laboratory setting but followed participants in their daily life taking in the context. Results might therefore be more generalizable for practical use which might be especially important for clinical interventions. Furthermore, due to the repeated measurement, errors might be less influential which increases validity and reliability of assessed profiles (Stone et al., 2007 as cited in Verhagen et al., 2016). Lastly, we obtained data throughout a wider time period. Assessing data using this longitudinal design allows us to see changes in responding depending on time (Shiffman et al., 2007).

Conclusion

Overall, our study extended the affective forecasting literature regarding the influence of individual differences. Specifically, we replicated findings showing a general tendency for

individuals to be overconfident when predicting emotions and translated it to the field of affective forecasting. Interestingly, however, people are generally well calibrated on an individual level: more certainty in their forecast is also associated with an increased accuracy. Whereas general overconfidence might be explained by cognitive biases (i.e., impact bias), good calibration might support the premise that humans are generally well self-aware of their abilities which has been found in human forecasting. Lastly, we found that negative emotions are predicted with less error than positive emotions when confidence is high, however this effect disappears as uncertainty increases. Whereas our study found promising results about how overconfidence relates to affective forecasting, this research area lacks replicability. We hope to encourage other researchers to expand on this novel finding.

References

- Abraham, C., & Sheeran, P. (2004). Deciding to exercise: The role of anticipated regret. *British Journal of Health Psychology*, 9(2), 269–278.
<https://doi.org/10.1348/135910704773891096>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects models using lme4. *Journal of Statistical Software*, 67(1). <https://doi.org/10.18637/jss.v067.i01>
- Bertoni, M., & Corazzini, L. (2018). Asymmetric affective forecasting errors and their correlation with subjective well-being. *PLoS ONE*, 13(3), e0192941.
<https://doi.org/10.1371/journal.pone.0192941>
- Brenner, L. A., Koehler, D. J., Liberman, V., & Tversky, A. (1996). Overconfidence in Probability and Frequency Judgments: A Critical Examination. *Organizational Behavior and Human Decision Processes*, 65(3), 212–219.
<https://doi.org/10.1006/obhd.1996.0021>
- Bolger, F., & Önköl-Atay, D. (2004). The effects of feedback on judgmental interval predictions. *International Journal of Forecasting*, 20(1), 29–39.
[https://doi.org/10.1016/s0169-2070\(03\)00009-8](https://doi.org/10.1016/s0169-2070(03)00009-8)
- Buehler, R., & McFarland, C. (2001). Intensity Bias in Affective Forecasting: The role of Temporal focus. *Personality and Social Psychology Bulletin*, 27(11), 1480–1493.
<https://doi.org/10.1177/01461672012711009>
- Campbell, W. K., Goodie, A. S., & Foster, J. D. (2004). Narcissism, confidence, and risk attitude. *Journal of Behavioral Decision Making*, 17(4), 297–311.
<https://doi.org/10.1002/bdm.475>
- Chapman, G. B., & Coups, E. J. (2006). Emotions and preventive health behavior: Worry, regret, and influenza vaccination. *Health Psychology*, 25(1), 82–90.
<https://doi.org/10.1037/0278-6133.25.1.82>

- Chatfield, C. (2001). Prediction Intervals for Time-Series Forecasting. In: Armstrong, J.S. (eds) *Principles of Forecasting*. International Series in Operations Research & Management Science, vol 30. Springer, Boston, MA. https://doi.org/10.1007/978-0-306-47630-3_21
- Christensen, P., Gillingham, K., & Nordhaus, W. (2018). Uncertainty in forecasts of long-run economic growth. *Proceedings of the National Academy of Sciences*, 115(21), 5409–5414. <https://doi.org/10.1073/pnas.1713628115>
- Clore, G. L., & Huntsinger, J. R. (2007). How emotions inform judgment and regulate thought. *Trends in Cognitive Sciences*, 11(9), 393–399. <https://doi.org/10.1016/j.tics.2007.08.005>
- Coleman, A. E., & Snarey, J. (2011). James-Lange Theory of Emotion. In *Springer eBooks* (pp. 844–846). https://doi.org/10.1007/978-0-387-79061-9_3146
- Darrow, A., Johnson, C. M., Miller, A. M., & Williamson, P. (2002). Can students Accurately assess themselves? Predictive Validity of Student Self-Reports. *Update Applications of Research in Music Education*, 20(2), 8–11. <https://doi.org/10.1177/875512330202000203>
- Feiler, D., & Tong, J. (2021). From noise to bias: Overconfidence in new product forecasting. *Management Science*, 68(6), 4685–4702. <https://doi.org/10.1287/mnsc.2021.4102>
- Fritz, J., Piccirillo, M. L., Cohen, Z. D., Frumkin, M., Kirtley, O., Moeller, J., Neubauer, A. B., Norris, L. A., Schuurman, N. K., Snippe, E., & Bringmann, L. F. (2024). So You Want to Do ESM? 10 Essential Topics for Implementing the Experience-Sampling Method. *Advances in Methods and Practices in Psychological Science*, 7(3). <https://doi.org/10.1177/25152459241267912>

- Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, 98(4), 506–528.
<https://doi.org/10.1037/0033-295x.98.4.506>
- Gilbert, D. T., Pinel, E. C., Wilson, T. D., Blumberg, S. J., & Wheatley, T. P. (1998). Immune neglect: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology*, 75(3), 617–638. <https://doi.org/10.1037/0022-3514.75.3.617>
- Gneiting, T., Balabdaoui, F., & Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, 69(2), 243–268. <https://doi.org/10.1111/j.1467-9868.2007.00587.x>
- Grounds, M. A., Joslyn, S., & Otsuka, K. (2017). Probabilistic Interval Forecasts: An individual differences approach to understanding forecast communication. *Advances in Meteorology*, 2017, 1–18. <https://doi.org/10.1155/2017/3932565>
- Hall, K. a. A., Joormann, J., Siemer, M., & Timpano, K. R. (2018). The impact bias in self and others: Affective and empathic forecasting in individuals with social anxiety. *Behaviour Research and Therapy*, 106, 37–46.
<https://doi.org/10.1016/j.brat.2018.05.001>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2–3), 61–83.
<https://doi.org/10.1017/s0140525x0999152x>
- Joslyn, S., Nemec, L., & Savelli, S. (2013). The benefits and challenges of predictive interval forecasts and verification graphics for end users. *Weather Climate and Society*, 5(2), 133–147. <https://doi.org/10.1175/wcas-d-12-00007.1>
- Klayman, J., Soll, J. B., González-Vallejo, C., & Barlas, S. (1999). Overconfidence: It depends on how, what, and whom you ask. *Organizational Behavior and Human Decision Processes*, 79(3), 216–247. <https://doi.org/10.1006/obhd.1999.2847>

- Koellinger, P., & Treffers, T. (2015). Joy leads to overconfidence, and a simple countermeasure. *PLoS ONE*, 10(12), e0143263.
<https://doi.org/10.1371/journal.pone.0143263>
- Kushlev, K., & Dunn, E. W. (2012). Affective forecasting: Knowing how we will feel in the future. In S. Vazire & T.D. Wilson (Eds.), *Handbook of self-knowledge* (pp. 277–292). Guilford Press.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). LMERTest Package: Tests in Linear Mixed Effects models. *Journal of Statistical Software*, 82(13).
<https://doi.org/10.18637/jss.v082.i13>
- Lazarus, J. (2021). Negativity bias: An evolutionary hypothesis and an empirical programme. *Learning and Motivation*, 75, 101731. <https://doi.org/10.1016/j.lmot.2021.101731>
- Liberman, N., Sagristano, M. D., & Trope, Y. (2002b). The effect of temporal distance on level of mental construal. *Journal of Experimental Social Psychology*, 38(6), 523–534.
[https://doi.org/10.1016/s0022-1031\(02\)00535-8](https://doi.org/10.1016/s0022-1031(02)00535-8)
- Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance*, 20(2), 159–183. [https://doi.org/10.1016/0030-5073\(77\)90001-0](https://doi.org/10.1016/0030-5073(77)90001-0)
- Loehr, V. G., & Baldwin, A. S. (2014). Affective forecasting error in exercise: Differences between physically active and inactive individuals. *Sport Exercise and Performance Psychology*, 3(3), 177–183. <https://doi.org/10.1037/spy0000006>
- López, O. a. M., López, A. M., & Crossa, J. (2022). Linear Mixed models. In *Springer eBooks* (pp. 141–170). https://doi.org/10.1007/978-3-030-89010-0_5
- Lüdtke, D. (2018). ggeffects: Tidy Data Frames of Marginal Effects from Regression Models. *The Journal of Open Source Software*, 3(26), 772.
<https://doi.org/10.21105/joss.00772>

- Lüdecke, D. (2024, November 29). *Data Visualization for Statistics in Social Science [R package sjPlot version 2.8.17]*. <https://cran.r-project.org/package=sjPlot>
- Mallett, R. K., Wilson, T. D., & Gilbert, D. T. (2008). Expect the unexpected: Failure to anticipate similarities leads to an intergroup forecasting error. *Journal of Personality and Social Psychology*, 94(2), 265–277. <https://doi.org/10.1037/0022-3514.94.2.94.2.265>
- McClenaghan, E. (2024, March 26). *The binomial test*. Informatics From Technology Networks. <https://www.technologynetworks.com/informatics/articles/the-binomial-test-366022#D4>
- Mellers, B. A., & McGraw, A. P. (2001). Anticipated emotions as guides to choice. *Current Directions in Psychological Science*, 10(6), 210–214. <https://doi.org/10.1111/1467-8721.00151>
- Mellers, B., Schwartz, A., & Ritov, I. (1999). Emotion-based choice. *Journal of Experimental Psychology General*, 128(3), 332–345. <https://doi.org/10.1037/0096-3445.128.3.332>
- Mestdagh, M., Verdonck, S., Piot, M., Niemeijer, K., Kilani, G., Tuerlinckx, F., Kuppens, P., & Dejonckheere, E. (2023). m-Path: an easy-to-use and highly tailorable platform for ecological momentary assessment and intervention in behavioral research and clinical practice. *Frontiers in Digital Health*, 5. <https://doi.org/10.3389/fdgth.2023.1182175>
- Moore, D. A., Dev, A. S., & Goncharova, E. Y. (2018). Overconfidence across cultures. *Collabra Psychology*, 4(1). <https://doi.org/10.1525/collabra.153>
- Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, 115(2), 502–517. <https://doi.org/10.1037/0033-295X.115.2.502>
- Moore, D. A., Swift, S. A., Minster, A., Mellers, B., Ungar, L., Tetlock, P., Yang, H. H. J., & Tenney, E. R. (2016). Confidence calibration in a multiyear geopolitical forecasting

- competition. *Management Science*, 63(11), 3552–3565.
<https://doi.org/10.1287/mnsc.2016.2525>
- Ng, V. K., & Cribbie, R. A. (2016). Using the Gamma Generalized linear model for modeling continuous, skewed and heteroscedastic outcomes in psychology. *Current Psychology*, 36(2), 225–235. <https://doi.org/10.1007/s12144-015-9404-0>
- Pallier, G., Wilkinson, R., Danthiir, V., Kleitman, S., Knezevic, G., Stankov, L., & Roberts, R. D. (2002). The role of individual differences in the accuracy of confidence judgments. *The Journal of General Psychology*, 129(3), 257–299.
<https://doi.org/10.1080/00221300209602099>
- Palmer, T. N. (2000). Predicting uncertainty in forecasts of weather and climate. *Reports on Progress in Physics*, 63(2), 71–116. <https://doi.org/10.1088/0034-4885/63/2/201>
- Peter, P. C., & Brinberg, D. (2012). Learning Emotional Intelligence: An exploratory study in the domain of health. *Journal of Applied Social Psychology*, 42(6), 1394–1414.
<https://doi.org/10.1111/j.1559-1816.2012.00904.x>
- Pilin, M. A. (2020). The past of predicting the future: A review of the multidisciplinary history of affective forecasting. *History of the Human Sciences*, 34(3–4), 290–306.
<https://doi.org/10.1177/0952695120976330>
- Rhodes, R., & Strain, J. J. (2007). Affective forecasting and its implications for medical ethics. *Cambridge Quarterly of Healthcare Ethics*, 17(01).
<https://doi.org/10.1017/s0963180108080067>
- Rizeq, J. (2024). Affective forecasting and psychopathology: A scoping review. *Clinical Psychology Review*, 108, 102392. <https://doi.org/10.1016/j.cpr.2024.102392>
- Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, 110(1), 145–172. <https://doi.org/10.1037/0033-295X.110.1.145>
- Sarkar, D. (2008). Lattice. In *Springer eBooks*. <https://doi.org/10.1007/978-0-387-75969-2>

Schaefer, P. S., Williams, C. C., Goodie, A. S., & Campbell, W. (2003). Overconfidence and the big five. *Journal of Research in Personality*, 38(5), 473–480.

<https://doi.org/10.1016/j.jrp.2003.09.010>

Shiffman, S., Stone, A. A., & Hufford, M. R. (2007). Ecological Momentary assessment. *Annual Review of Clinical Psychology*, 4(1), 1–32.

<https://doi.org/10.1146/annurev.clinpsy.3.022806.091415>

Snell, J., Gibbs, B. J., & Varey, C. (1995). Intuitive hedonics: Consumer beliefs about the dynamics of liking. *Journal of Consumer Psychology*, 4(1), 33–60.

https://doi.org/10.1207/s15327663jcp0401_02

Statistics Solutions. (2025, May 12). *Assumptions of Logistic Regression - Statistics Solutions*. <https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/assumptions-of-logistic-regression/>

Takano, K., & Ehring, T. (2023). Affective forecasting as an adaptive learning process. *Emotion*, 24(3), 795–807. <https://doi.org/10.1037/emo0001303>

Ushakova, A., & Waterston, E. (2019). *PPLS PhD Training Workshop: Statistics and R*. bookdown. https://bookdown.org/animestina/phd_july_19/

Verhagen, S. J. W., Hasmi, L., Drukker, M., Van Os, J., & Delespaul, P. a. E. G. (2016). Use of the experience sampling method in the context of clinical trials. *Evidence-Based Mental Health*, 19(3), 86–89. <https://doi.org/10.1136/ebmental-2016-102418>

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Golemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., . . . Yutani, H. (2019). Welcome to the Tidyverse. *The Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>

- Wilson, T. D., & Gilbert, D. T. (2003). Affective forecasting. In *Advances in experimental social psychology* (pp. 345–411). [https://doi.org/10.1016/s0065-2601\(03\)01006-2](https://doi.org/10.1016/s0065-2601(03)01006-2)
- Wilson, T. D., & Gilbert, D. T. (2005). Affective forecasting. *Current Directions in Psychological Science*, 14(3), 131–134. <https://doi.org/10.1111/j.0963-7214.2005.00355.x>
- Wright, A. G. C., & Zimmermann, J. (2019). Applied ambulatory assessment: Integrating idiographic and nomothetic principles of measurement. *Psychological Assessment*, 31(12), 1467–1480. <https://doi.org/10.31234/osf.io/6qc5x>
- Yates, J. F., Lee, J.-W., Shinotsuka, H., Patalano, A. L., & Sieck, W. R. (1998). Cross-cultural variations in probability judgment accuracy: Beyond general knowledge overconfidence? *Organizational Behavior and Human Decision Processes*, 74(2), 89–117. <https://doi.org/10.1006/obhd.1998.2771>

Appendix A

AI Acknowledgment

No content generated by AI technologies has been presented as my own work.

I acknowledge the use of ChatGPT-4o (<https://openai.com/de-DE/index/chatgpt/>) to generate materials for background research and self-study in the drafting of this assessment.

The following prompts were input into ChatGPT-4o:

(1) “Give me a synonym for tendency”

The output obtained was:

Here are some suitable synonyms for tendency, depending on context:

- Inclination
- Propensity
- Disposition
- Trend
- Pattern
- Predisposition
- Habit
- Proclivity
- Drift (in specific contexts)
- Lean(ing)

Let me know the sentence you're using it in if you want the most natural replacement.

The output was changed by me in the following ways:

I replaced the often used word “tendency” with “trend” i.e., “We investigated if there is a general overconfidence trend using a binomial test”

(2) “If I have a dataset, how do I read it in r”

The output obtained was:

To read a dataset into R, you can use different functions depending on the file type (e.g. CSV, Excel, SPSS, etc.). Here's how to read the most common types:

1. CSV File

If your dataset is a .csv file:

r

```
data <- read.csv("path/to/your/file.csv")
```

Example:

r

```
data <- read.csv("C:/Users/YourName/Documents/data.csv")
```

If it's separated by semicolons (common in some countries), use:

r

```
data <- read.csv2("path/to/your/file.csv")
```

2. Excel File (.xlsx or .xls)

First, install and load the readxl package:

r

```
install.packages("readxl")
```

```
library(readxl)
```

Then:

r

```
data <- read_excel("path/to/your/file.xlsx")
```

3. SPSS, Stata, or SAS Files

Use the haven package:

r

```
install.packages("haven")
```

```
library(haven)
```

```
# For SPSS
```

```
data <- read_sav("file.sav")
```

```
# For Stata
```

```
data <- read_dta("file.dta")
```

```
# For SAS
```

```
data <- read_sas("file.sas7bdat")
```

TIP: Use RStudio's built-in GUI

If you're using RStudio, go to:

File > Import Dataset to import a file interactively.

Let me know the file format if you want help with a specific dataset.

The output was changed by me in the following ways:

I used the instruction to get familiar with r and understand basic functions

(3) “my r says: error in use method arrange. no applicable method for the arrange applied to an object of class function”

The output obtained was:

The error message you're seeing:

```
pgsql
```

Error in UseMethod("arrange"): no applicable method for 'arrange' applied to an object of class "function" means that R thinks arrange is a **function object**, not the arrange() function from the dplyr package.

Likely Cause:

You probably **typed or created a variable named** `arrange` somewhere earlier in your code, which **overwrote the** `dplyr::arrange()` **function**.

How to Fix It:

1. **Restart your R session** (clears everything from memory):

In RStudio: Session > Restart R

Or press: Ctrl + Shift + F10 (Windows) or Cmd + Shift + F10 (Mac)

2. **Reload dplyr** after restarting:

```
r
```

```
library(dplyr)
```

Then try your `arrange()` code again:

```
r
```

```
my_data <- arrange(my_data, time)
```

 Optional: Check for name conflict

Before restarting, you can confirm the conflict with:

```
r
```

```
arrange
```

If it prints something like `<function>` or a value you didn't expect, you've overwritten the `arrange()` function.

Let me know if you want help rewriting your specific line or debugging further — happy to help!

The output was changed by me in the following ways:

I resolved this error message by starting a new `r` session and beginning again.

These prompts are exemplary of my use of ChatGPT when learning to code in R (v4.1.2; R Core Team 2021).

Appendix B

Table 1

Sample Characteristics (N = 30)

Age in years, <i>M</i> (<i>SD</i>)	19.97 (1.83)
Sex, <i>n</i> (%)	
Female	20 (66.7)
Male	10 (33.3)

Table 2

Variance Inflation Factors for Predictors

Predictor	VIF
Interval_Width	3.05
Affect_Type	4.58
Interval_Width:Affect_Type	6.85

Table 4

Binomial Test

Variable	Level	Counts	Total	Proportion	<i>p</i>	95% CI for Proportion	
						Lower	Upper
PA_Interval	0	419	1168	.36	<.001	.33	.39
	1	749	1168	.64	<.001	.61	.67
NA_Interval	0	419	1168	.36	<.001	.33	.39
	1	749	1168	.64	<.001	.61	.67

Note. Proportions tested against value: .95.

Table 5

Linear Mixed Model: Fixed Effects Estimate

Term	Estimate	SE	df	t	p
Intercept	8.77	.66	100	13.21	<.001
Affect	-.84	.47	2300	-1.78	.076
Interval_Width	.09	.02	888	5.18	<.001
Affect X	0.02	.02	2252	1.12	.264
Interval					

Table 6

Generalized Mixed Model: Fixed Effects Estimates

Term	Estimate	SE	t	p
Intercept	2.48	.04	66.92	<.001
Affect	-.05	.02	-2.69	.007
Interval_Width	.13	.02	5.81	<.001
Affect X	.04	.02	2.01	.045
Interval_Width				

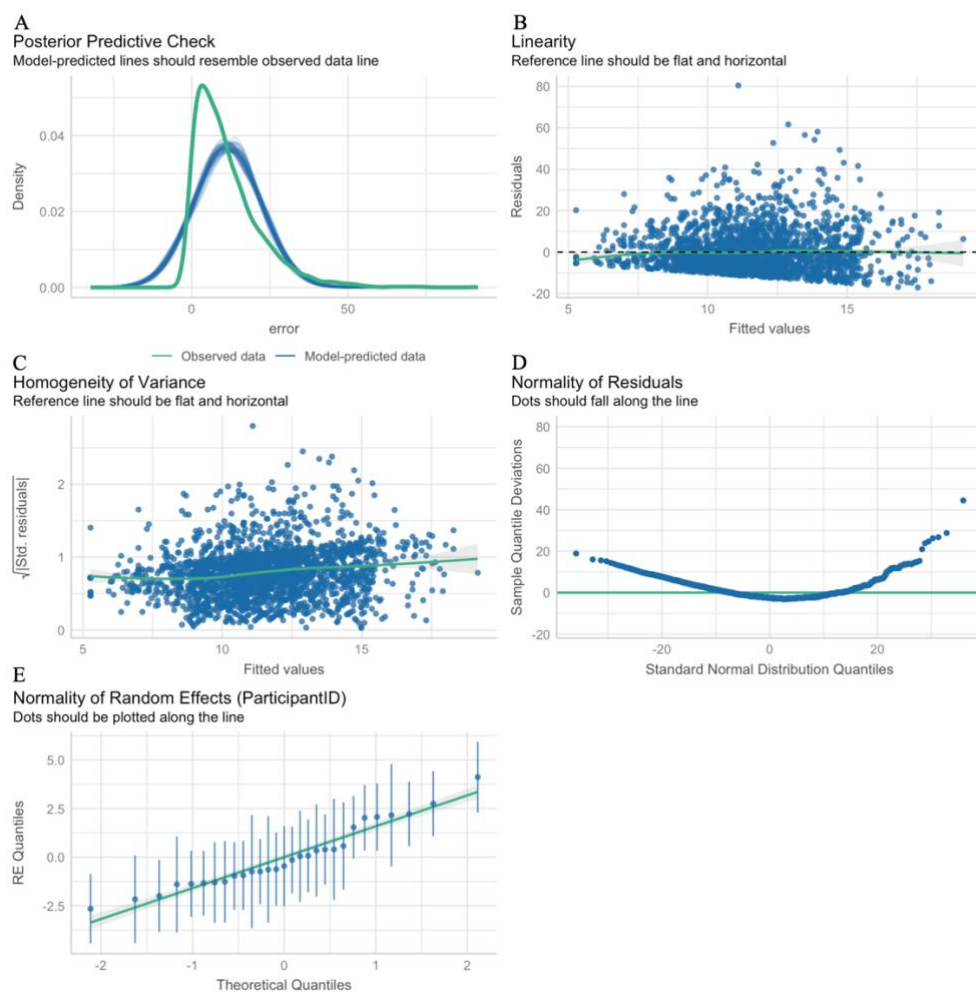
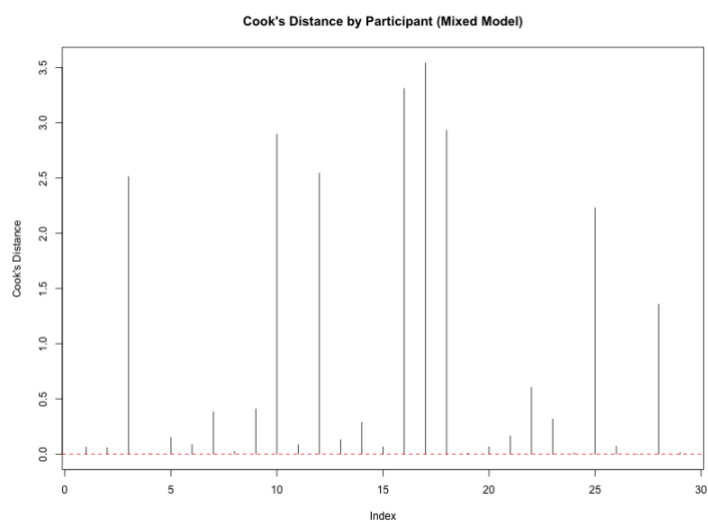
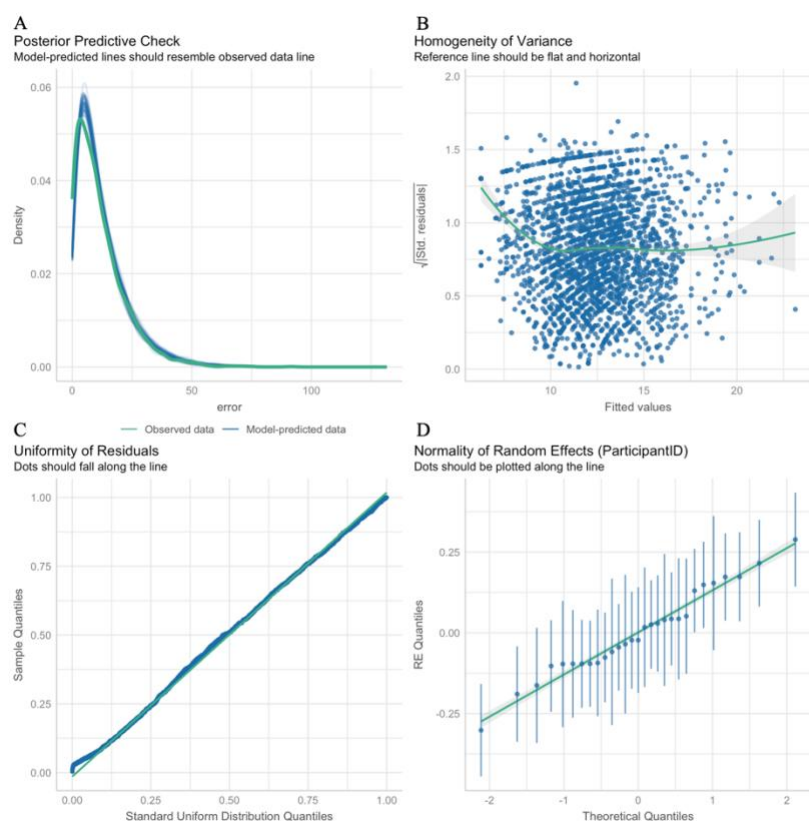
Figure 1*Assumption Checks: Linear Mixed Model***Figure 2***Cooks Distance: Linear Mixed Model*

Figure 3

Assumption Checks: Generalized Linear Mixed Model

**Figure 4**

Person Specific Proportions Within Interval: Positive Affect

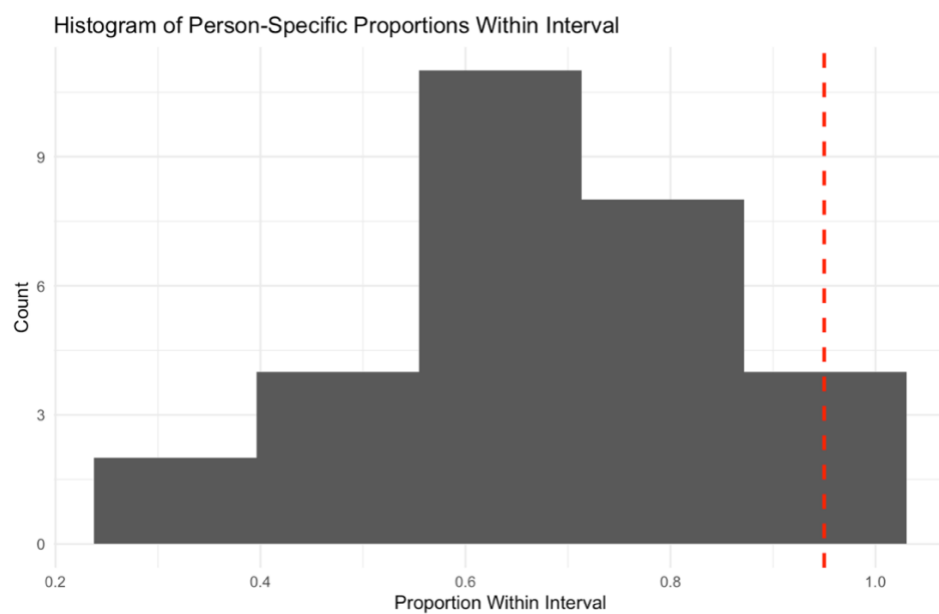
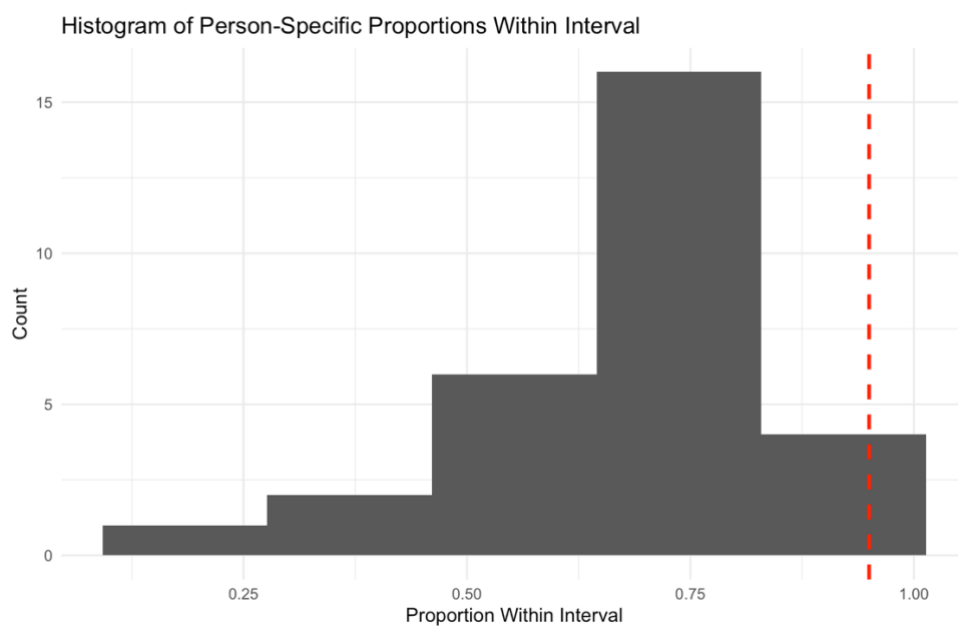
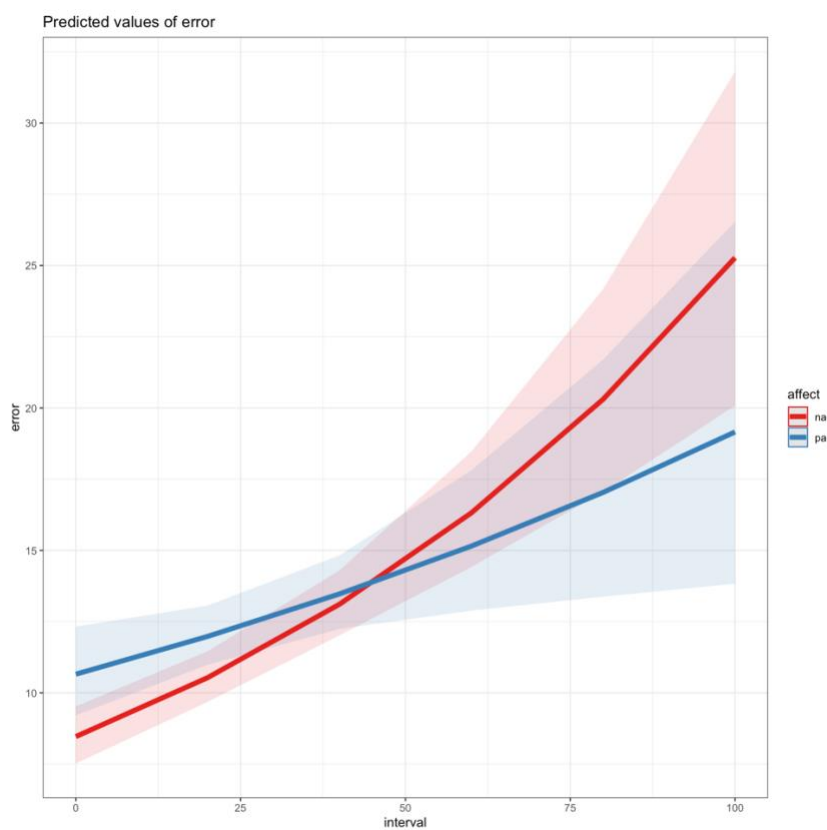


Figure 5*Person Specific Proportions Within Interval: Negative Affect***Figure 6***Interaction Effect Generalized Mixed Model*

Appendix C

```

library(tidyverse)

data <- read.csv("results/full_predictions.csv", na="NA")

data_clean <- data %>%

  filter(!is.na(PA_now_point) & !is.na(PA_one_step_point_match) ) %>%

  group_by(ParticipantID) %>%

  select(ParticipantID:time_n,

         PA_now_point,PA_one_step_point_match, PA_one_step_lower_match,PA_one_step_upper_match,

         NA_now_point,NA_one_step_point_match,

         NA_one_step_lower_match,NA_one_step_upper_match) %>%

  mutate(

    #as.numeric transforms it to 0 if not in interval and to 1 if in interval

    PA_interval_included = as.numeric(PA_now_point >= PA_one_step_lower_match & PA_now_point <=

    PA_one_step_upper_match),

    NA_interval_included = as.numeric(NA_now_point >= NA_one_step_lower_match & NA_now_point

    <= NA_one_step_upper_match),

    PA_interval_width = abs(PA_one_step_lower_match - PA_one_step_upper_match),

    NA_interval_width = abs(NA_one_step_lower_match - NA_one_step_upper_match),

    PA_error = abs(PA_now_point - PA_one_step_point_match),

    NA_error = abs(NA_now_point - NA_one_step_point_match

    ) %>% select(ParticipantID:time_n,starts_with("PA"),starts_with("NA"))

  # view for visual inspection

  View(data_clean)

write.csv(data_clean,"dat_henrike.csv")

# now group per person

data_clean_agg <- data_clean %>% group_by(ParticipantID) %>%

  summarise(PA_interval_included_prop = mean(PA_interval_included),

            NA_interval_included_prop = mean(NA_interval_included))

library(ggplot2)

```

```

# Create histogram with vertical line at 0.95

ggplot(data_clean_agg, aes(x = PA_interval_included_prop)) +

  geom_histogram(bins = 5) +

  geom_vline(xintercept = 0.95, color = "red", linetype = "dashed", size = 1) +

  labs(

    title = "Histogram of Person-Specific Proportions Within Interval",

    x = "Proportion Within Interval",

    y = "Count"

  ) +

  theme_minimal()

# Create histogram with vertical line at 0.95

ggplot(data_clean_agg, aes(x = NA_interval_included_prop)) +

  geom_histogram(bins = 5) +

  geom_vline(xintercept = 0.95, color = "red", linetype = "dashed", size = 1) +

  labs(

    title = "Histogram of Person-Specific Proportions Within Interval",

    x = "Proportion Within Interval",

    y = "Count"

  ) +

  theme_minimal()

total_successes_pa <- sum(data_clean$PA_interval_included)

total_successes_na <- sum(data_clean$NA_interval_included)

# Total number of values

total_trials <- nrow(data_clean)

# Run binomial test

binom.test(total_successes_pa, total_trials, p = 0.95)

library(lme4)

model_pa <- glmer(PA_interval_included ~ 1 + (1 | ParticipantID),

  data = data_clean,

```

```

    family = binomial)

summary(model_pa)

confint_model <- confint(model_pa, parm = "beta_", method = "Wald")

# Point estimate (log-odds intercept)

logodds_estimate <- fixef(model_pa)[1]

# CI in log-odds scale

ci_logodds <- confint_model[1, ]

# Transform to probability scale

prob_estimate <- plogis(logodds_estimate)

prob_ci <- plogis(ci_logodds)

cat("Estimated probability:", round(prob_estimate, 4), "\n")

cat("95% CI:", round(prob_ci[1], 4), "-", round(prob_ci[2], 4), "\n")

model_na <- glmer(NA_interval_included ~ 1 + (1 | ParticipantID),
    data = data_clean,
    family = binomial)

summary(model_na)

confint_model <- confint(model_na, parm = "beta_", method = "Wald")

# Point estimate (log-odds intercept)

logodds_estimate <- fixef(model_na)[1]

# CI in log-odds scale

ci_logodds <- confint_model[1, ]

# Transform to probability scale

prob_estimate <- plogis(logodds_estimate)

prob_ci <- plogis(ci_logodds)

cat("Estimated probability:", round(prob_estimate, 4), "\n")

cat("95% CI:", round(prob_ci[1], 4), "-", round(prob_ci[2], 4), "\n")

# merge person specific proportion with individual error

# now we use the average proportion per participant of confidence as covariate

data_all_merge <- data_clean %>% left_join(data_clean_agg)

```

```

library(lmerTest)

library(lattice) ## for dotplot

mod_pa <- lmerTest::lmer( PA_error ~ PA_interval_included_prop + (1 | ParticipantID),
                        data = data_all_merge)

#visualise random effects

dotplot(ranef(mod_pa))

summary(mod_pa)

mod_na <- lmerTest::lmer( NA_error ~ NA_interval_included_prop + (1 | ParticipantID),
                        data = data_all_merge)

#visualise random effects

dotplot(ranef(mod_na))

summary(mod_na)

# put into long format to have one column for error, one for affect type and one for affect type

dat_long <- data_clean %>%

  select(ParticipantID,day_n,time_n,PA_error,NA_error,PA_interval_width,NA_interval_width) %>%

  pivot_longer(

    cols = -c(ParticipantID,day_n,time_n),

    names_to = c("affect", "measure"),

    names_sep = "_",

    values_to = "value"

  ) %>%

  pivot_wider(

    names_from = measure,

    values_from = value,

    names_prefix = ""

  ) %>%

  mutate(affect = tolower(affect))

dat_long %>% group_by(affect) %>%

  summarise(

```

```

mean_error = mean(error),

sd_error = sd(error),

mean_interval = mean(interval),

sd_interval = sd(interval),

n = n()
)

mod_interval_affect <- lmerTest::lmer(error ~ interval*affect + (1 | ParticipantID),

data = dat_long)

summary(mod_interval_affect)

# different effects !

ggemmeans(mod_interval_affect, terms = c("interval","affect")) %>%

plot(log_y = F)+

geom_line (size = 2) +theme_bw()

performance::check_model(mod_interval_affect, panel = T,check =

c("pp_check","reqq","linearity","qq","homogeneity"))

# assumptions are violated so lets double check with another model that is appropriate for these types of

data

mod_interval_affect_glm <- glmer(error + 1 ~ scale(interval)*affect + (1 | ParticipantID),

data = dat_long,family = Gamma(link="log"))

summary(mod_interval_affect_glm)

performance::check_model(mod_interval_affect_glm, panel = T,check =

c("reqq","pp_check","qq","homogeneity"))

library(ggeffects)

ggemmeans(mod_interval_affect_glm, terms = c("interval","affect")) %>%

plot(log_y = F)+

geom_line (size = 2) +theme_bw()

library(sjPlot)

# https://cran.r-project.org/web/packages/sjPlot/vignettes/tab\_mixed.html

tab_model(mod_interval_affect_glm)

```

```
write.csv(dat_long,'results_students/dat_henrike_new.csv')
```