Affective Forecasting Accuracy Across Different Interval Lengths and Affective Valences

Fabienne-Renée Liepelt

s5104491

Department of Psychology, University of Groningen

PSB3E-BT15: Bachelor Thesis

Group number: 26

Supervisor: Fridtjof Petersen

Second evaluator: dr. Oliver Weigelt

In collaboration with: Henrike Nebel, Marieke Boerendonk, Martino Avalle, and Miora Haslacher.

June 30, 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

Background: Affective forecasting, predicting future emotional states, significantly influences decision-making. However, humans tend to be inaccurate in forecasting their future affect because they overestimate the duration and intensity of their future emotions (impact bias). The impact of emotional valence and different interval lengths on affective forecasting accuracy, as well as differences between human and statistical model forecasts have been researched separately leaving unclear how these variables might interact.

Aims: This study investigated whether human affective forecasting accuracy decreases with longer time intervals, whether this effect varies by emotional valence, and how human accuracy compares to Kalman filter predictions across 3-hour and 6-hour intervals.

Methods: We conducted a 14-day experience sampling study in which participants responded to five daily prompts measuring current and predicted positive and negative affect. Absolute prediction errors were analyzed via multilevel models, comparing human and model forecasts.

Results: Human prediction accuracy decreases when the interval size increases. However, this effect is independent of the emotional valence, contradicting previous research. Moreover, there was no significant difference between the Kalman filter and human forecasting accuracy.

Conclusions: While human affective forecasts are more accurate for the near than the distant future, which applies equally to predicting positive and negative affect, the Kalman filter does not perform significantly differently than humans. However, differences in operationalization of interval size as differences in event specificity (specific event vs daily life forecasts) may limit the generalizability of our findings which could be a starting point for future research.

Keywords: affective forecast, experience sampling, statistical model, interval size

Affective Forecasting Accuracy Across Different Interval Lengths and Affective Valences

Major life events such as being accepted for one‘s dream university, or breaking up with a significant other, often do not lead to the long lasting negative nor positive emotion we might have expected. This circumstance is a standard example of affective forecasting which refers to predicting one‘s future emotions. For instance, predicting how happy or sad one will feel after a particular future event (Takano & Ehring, 2024). It involves forecasting whether an anticipated event will produce positive or negative emotions (valence), how strongly these emotions are assumed to be felt (intensity), for how long they will last (duration) and finally which specific emotions are expected (Wilson & Gilbert, 2003).

Understanding affective forecasting is important because we often rely on our anticipated emotions when making decisions. Hoerger et al. (2016) for instance investigated the impact of affective forecasts on women's decisions regarding breast cancer chemoprevention medications. To begin with, women with an increased risk for breast cancer are given the choice to start using chemoprevention medication. While these medications are classified as beneficial since they might lower their chances of developing the disease within the next five years by roughly 50% (Moyer, 2013), most women opt against it (Fragerlin et al., 2011). In their study, the researcher sampled women between the ages of 40 and 71 who were identified as at an elevated risk for developing breast cancer based on previous screening results. After receiving background information about the medication including its side effects, participants were asked to predict how their health-related stress levels might change when undergoing or rejecting treatment. Moreover, they rated their behavioral intentions to actually seek treatment. Three months later, participants reported whether they opted for a specific treatment, against it or whether they are still pondering about it. Their results revealed that the majority of women at risk anticipated the chemopreventive medication to increase their health-related stress levels which in turn was associated with having lower intentions of seeking treatment and being unlikely to eventually decide in favor of chemopreventive medication. In contrast, less than ten percent of the participants expected their health-related stress levels to decrease. These women showed higher intentions for taking medication and a lower likelihood of rejecting the treatment. Importantly, despite the pessimistic forecasts about their health-related stress-levels, research does not support that taking chemopreventive medications actually increases health-related stress (Day et al., 1999). Thus, although chemopreventive medication is a promising intervention and not inevitably associated with more health-related-stress, women might still opt against it based on what they expect to happen. Therefore, this finding suggests that people may not be able to anticipate correctly what is best for their health just as by deciding against chemopreventive medication you might risk developing cancer. Consequently, what we predict to be the best for our health does not lead to the decisions that necessarily benefit our health.

Despite the importance of accurate affective forecasting for decision making, people tend to be inaccurate in predicting their emotions. Researchers have sought to understand the underlying mechanisms that contribute to these inaccuracies. While individuals can usually predict the direction of their emotions rather accurately (i.e. feeling better or worse than at the current moment), their intensity and duration tend to be predicted less accurately. This tendency is called *impact bias* (Wilson & Gilbert, 2003). For example a college graduate might expect to feel frustrated for weeks if his application for a desired job will be rejected again. He accurately forecast that this outcome will make him feel worse than he is currently feeling (direction). Yet, his frustration might not be as intense as anticipated and may only last for a couple of days instead of weeks (duration). Eastwick et al. (2007) for example aimed to examine whether the forecasting errors regarding the dissolution of a romantic relationship occur because people misjudge how strong their initial emotions will be (intensity) or because they overestimate for how long these emotions will last (duration). They conducted a longitudinal study in which US American first year students who were in a relationship completed questionnaires asking them to predict how a breakup of their relationship would affect them emotionally. Those whose relationship actually ended were asked to rate their level of distress when the event was still recent, and additionally, months later. The results showed although participants overestimated how much distress the break-up would cause them in the first weeks (intensity), they did not overestimate how long this event would affect them (duration). Interestingly, the researchers also found individual differences among the participants. Those who anticipated to remain single for a longer period of time before entering a new relationship, or who believed they were less likely than their partner to initiate the breakup, tended to overestimate its emotional impact more severely but not its duration (Eastwick et al., 2007).

The previous study focused on affective forecasting errors about one's emotional reactions after an event has taken place, which raises the question of how forecasting accuracy may differ depending on the temporal distance to an event. That is, whether there are differences in accuracy for the near vs distant future. Braun and Yaniv (1992) were one of the first researchers to investigate general human forecasting (so not affective forecasting) accuracy for different time horizons. Specifically, they examined how well economic experts predicted the likelihood of an economic recession for each of the next five quarters. These estimates were later compared to the actual observed economic outcome. As hypothesized, the authors found that human forecasts were more accurate for short-term than long-term predictions. They reasoned this pattern was to be expected because knowledge about the current state of affairs may be more useful when predicting short-term versus long-term outcomes. Although these findings stem from a different domain (economic expert judgments), they might generalise to affective forecasting. That is, people may similarly be more accurate in forecasting their emotions for near than distant future events or time points.

In addition to the temporal distance to an event, accuracy might be impacted depending on whether the predicted emotion is positive or negative. Finkenauer et al. (2007) found that people tend to make more accurate predictions about positive emotions only when the event was temporally distant (i.e four to five days beforehand). In contrast, predictions of negative emotions were more accurate when the event was temporally close (i.e one to three days beforehand). They reasoned that for events that are far in the future people feel more in control and thus, focus their efforts on achieving a positive outcome. To facilitate realizing this goal, they engage in biased thinking. That is, they downplay how rewarding success may feel and instead exaggerate how dreadful failure would be. As a result people perceive failure to be more threatening than it actually is which in turn helps them to stay motivated to avoid defeat. In terms of forecasting accuracy, people then overestimate the duration of their negative feelings following failure while correctly anticipating that the happiness felt after succeeding will fade rapidly. In contrast, for near events people believe to have less control to influence an outcome. Therefore, they try to shift their efforts to regulating their emotional response. They prepare themselves mentally by de-emphasizing how bad failure might be to weaken the impact of a possible negative outcome beforehand. Simultaneously, they up-regulate their expectation of how good success would feel. Consequently, failure is perceived to be temporary. Thus, they accurately predict that their negative affect will be short-lived but overestimate the duration of happiness in case of a positive outcome. Taken together, for temporally distant events inaccurate forecasts of negative affect following failure may be a motivational source to keep working to achieve one‘s goals. In comparison, for temporally close events, inaccurate forecasts of positive affect may serve as a coping mechanism for possible unpleasant outcomes (Finkenauer et al., 2007, Pennington & Roese, 2003).

Nevertheless, it should be noted that they examined forecasts that were made for a very specific event (driver's license). Hence, it is unclear whether this pattern also translates to everyday situations.

The biases found in affective forecasting are also evident in other domains of human judgement (e.g economic forecasting), raising the question whether simple statistical models could offer more accurate predictions and potentially reduce decision-making errors. Such models typically produce forecasts by taking historical data into account and relying on the frequency with which an event occurs rather than on subjective impressions (Braun & Yaniv, 1992). As these models are designed to compare the likelihood of an event to the base rate probability it can be assumed that they would outperform human performance under certain conditions. Braun and Yaniv (1992) for instance, also tested in the aforementioned study whether simple statistical models performed better than human judgement (economic forecast). As expected, the model performed better than humans at far future predictions but not at immediate future prediction, suggesting that the wealth of information available to humans may not be beneficial in the long run as they tend to neglect base rate information, a bias known as base rate fallacy, which contributes to their inaccurate forecasts (Kahneman & Tversky, 1973). Generally, it has become common across different domains to assess how well human judgement compares to algorithmic forecasts (Göndöcs & Dörfler, 2024).

This comparison between human and algorithmic forecasting has also been extended to the domain of affective forecasting. In particular, Takano and Ehring (2024) examined how human affective forecasts compare to those generated by a statistical model known as the Kalman filter. This model works based on adaptive learning processes. That is, it compares affective forecasts to actual experienced affect and uses this comparison to improve the next forecast. Whereas humans are prone to believe that their future emotions will be similar to their current emotions (i.e *projection bias*), the Kalman filter is not influenced by this bias and instead is able to accurately incorporate past forecasts including their errors (Takano & Ehring, 2024). In their study they combined the approach of previous researchers that have explored how affective forecasting accuracy depends on the emotion that is predicted (Finkenauer et al., 2007) and how general human forecasting compares to model forecasts (Braun & Yaniv, 1992). They also addressed prior shortcomings by including a natural setting as opposed to a personally relevant event (e.g Finkenauer et al., 2007). Specifically, in two separate studies they analyzed affective forecasting accuracy for minute (one to two minutes) and hour-long (one to two hours) intervals and looked at differences between positive and negative affect. In addition, they differentiated between the overall size of error (absolute error) and the exact difference between anticipated and actual affect (relative error) which allows for conclusions about patterns of over- and underestimations. Their findings revealed that for minute-long forecasts, human participants only made significantly larger errors than the Kalman filter for negative affect, but no significant differences were found for positive affect. However, they were slightly more likely than the Kalman filter to underestimate positive and negative affect. In comparison, the reverse was true for hour-long forecasts: humans consistently made larger overall errors for positive and negative affect than the model. Yet, there was no significant bias towards over- or underestimation for neither predicted affect. These results are inconclusive regarding conclusions about positive and negative affect as well as short- and long-term forecasts as no clear pattern emerged. Possibly, because they were explored across two studies rather than within one study and because the observed differences between positive and negative affect were not tested for significance. Likewise, differences across short (minute-long) and long (hour-long) intervals were also not tested for significance. This leaves unclear whether the Kalman Filter performs better than humans and whether prior identified differences in positive and negative affect and time frame still hold.

Taken together, the current literature has explored how human and statistical model forecasts differ in accuracy, as well as how forecast accuracy is influenced by temporal distance (near vs far) and emotional valence (positive vs negative). However, these variables have typically been examined in isolation and recent studies show mixed results making meaningful joint conclusions about these variables difficult. Moreover, previous researchers (e.g Takano & Ehring, 2024; Braun & Yaniv 1992) have looked at both model and human forecasts for only one time point in the future, leading to the question whether their findings also hold for time points further in the future. Building on these findings, the present study aims to investigate possible interaction effects; specifically, whether forecast accuracy varies as a function of both the source of the forecast (human vs. model) and the time interval, as well as affective valence (positive vs. negative). In line with Braun and Yaniv's (1992) findings we hypothesise that participants will make more accurate predictions for shorter intervals than longer intervals. However, we expect this tendency to be different for positive and negative affect forecasts. Specifically, consistent with the findings of Finkenauer et al. (2007), we hypothesise participants to be more accurate in predicting positive affect than negative affect when the interval is greater and more accurate predictions for negative affect than positive affect when the interval is shorter. Lastly, we expect the Kalman filter to outperform human forecasts for distant predictions for both negative and positive affect. Yet, for close predictions model and human predictions should be similar (Braun & Yaniv, 1992).

Method

The current research is designed to be a partial replication and extension of the study conducted by Takano and Ehring (2024). Therefore, we also used the experience sampling method (ESM) to investigate affective forecasting. That is, participants predict and report their real-time emotional experiences through their mobile device several times a day (Shiffman et al., 2008). These responses are compared to their actual observations to determine forecasting accuracy (Takano & Ehring, 2024). Likewise, we will also compare human forecasting accuracy to the Kalman filter‘s forecasting accuracy. In addition, we will extend their study by including a two-step ahead prediction rather than only a one-step ahead prediction and also by asking participants to estimate an interval for their current and predicted affect. In this present paper, forecasting accuracy across different interval lengths, affective valences and sources of prediction was examined.

Participants

As we aimed to replicate and extend the study of Takano and Ehring (2024) we thus 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, Mean Age = 19.97 *SD* = 1.83). This deviation in sample size arose due to time limitations. The study was advertised via flyers on campus, social media, and the student research portal. Therefore, we used a convenience sample for our study.

Procedure

Data collection methods

To be eligible for this research project, students must own a smartphone and cannot be diagnosed with depression or anxiety. Participating in research projects like ours constitutes part of their curriculum at their university, and they are rewarded with credits for it. The study was ethically approved (FMG-12534_2025) and consent was received through a Qualtrics questionnaire which also included instructions about installing the m-Path application (Mestdagh et al., 2023). Upon enrollment, participants received a link to the m-Path questionnaire via email. Participants received five prompts per day (9am, 12pm, 3pm, 6pm, 9pm) for 14 consecutive days to complete short questionnaires about current and future affect (three and six hours ahead). For all time points participants were asked to rate and predict their affect by giving point and interval predictions. Each assessment took approximately 5-10 minutes to complete. Moreover, participants had to fill out the questionnaire within 30 minutes and received a reminder notification 15 minutes after the initial beep.

Materials

Participants were asked for their age and gender as a baseline questionnaire. During each of the ESM assessments, like Takano and Ehring (2024) we measured four different emotions: happiness, relaxation, sadness, and anxiety. That is, two positive and two negative emotions. These affect items were selected because they reflect the two axes valence and arousal of emotions. As briefly mentioned in the introduction, valence refers to whether emotions are evaluated as positive or negative. In contrast, arousal describes the intensity level of an emotion. Specifically, happiness and relaxation are positively valenced whereas sadness and anxiety are negatively valenced. Similarly, happiness and anxiety are high in arousal, but relaxation and sadness are low in arousal.

At each ESM prompt, participants rated their current emotional state for each of the four emotions on a 0 (“not at all”) to 100 (“extremely”) Visual Analogue Scale (VAS; Miller & Ferris, 1993) by positioning a slider along a continuous line. Specifically, the item was phrased as follows „*Please rate how much you experience each emotion at the moment*“. Additionally, we also asked participants to provide the one- and two-step ahead affective forecasts which were formulated as follows „*My best guess is that I will be… (emotion)*“. Within each assessment, the current rating was always followed by the two forecast ratings, but the order in which the four emotions appeared was fully randomised to control for order effects. Finally, happiness and relaxation were combined to a positive affect score and sadness and anxiety were combined to a negative affect score.

Model Specification

In addition to asking participants to predict their emotional state, we also used the Kalman filter to generate independent statistical forecasts at each beep. The Kalman filter functions as an optimal learning algorithm that uses participants' actual experienced affect to estimate key parameters for its forecasting model (Takano & Ehring, 2024). A central feature of the Kalman filter is its weighting system, which determines how much weight is placed on the current emotional experience to forecast the next time point, and how much is placed on past forecasts generated by the Kalman filter itself. Unlike human participants, the Kalman filter maintains perfect memory of past (model-generated) forecasts and optimally adjusts these weights based on prediction errors to minimise forecasting errors. Using the Kalman filter enables us to compare human forecasts to a statistical model that tries to imitate the perfect remembering process. (Takano & Ehring, 2024).

Data Analysis

Data Preprocessing

We excluded data of participants with less than 30% compliance (i.e 21 beeps) in line with the data from Takano and Ehring (2024). The overall completion rate of the questionnaires in our dataset is 72.4% (*SD* = 19.9), which equals 50.7 questionnaires out of 70 with an individual compliance range from 20 to 67 questionnaires. Moreover, of the total number of participants 29 filled out 21 or more questionnaires. Once the missing data was handled, the data was further modified into structured data using R Statistical Software (v. 4.5.0 R Core Team 2025) by calculating the absolute prediction errors for interval length (short vs long), and source of prediction (human vs Kalman).

Statistical Analysis

For all the analyses we used R Statistical Software (v. 4.5.0 R Core Team 2025) to conduct multilevel model analyses. Specifically, we used the lmer Test R package (v3.1.3

Kuznetsova et al., 2017). Because each participant provided repeated emotion ratings over two weeks, measurements within one individual may be more similar (i.e correlated) than measurements between different participants which violates the assumption of independent observations that are central to traditional regression techniques. Multilevel models take that dependency into account by estimating random effects for each participant (Theobald, 2018) .

First, we compared how human prediction accuracy may change for short intervals (three hours) and for long intervals (six hours) and investigated whether this pattern holds true for both positive and negative affect. Interval size and affective valence were included in the model as fixed effects and prediction error as dependent variable. Based on previous literature (Braun & Yaniv, 1992), we expected smaller prediction errors for shorter than longer intervals. In addition, for short intervals we expected that the prediction error would be smaller for forecasting negative than positive affect. In contrast, for longer intervals, prediction errors would be smaller for forecasting positive than negative affect (Finkenauer et al., 2007).

Second, we examined whether there is an interaction effect between the source of prediction (Kalman vs. human) and interval size (short vs. long). Source of prediction and interval size were included as fixed effects. We expected the Kalman filter to make more accurate predictions for longer intervals, but to be similarly accurate as humans for shorter intervals (Braun & Yaniv, 1992).

Results

Assumption Checks

After running the analyses, we checked whether the assumptions were met by using the performance R package (v0.14.0, Lüdecke et al., 2021). Firstly, the linearity assumption seems to be met (see Appendix B Figure 1A) as the reference line is horizontal as required.

Likewise, the variance of the residuals appears to be constant (1B). Similarly, the random effects seem to follow a normal distribution (1D). However, the normality assumption of the residuals is severely violated (2C). That is probably the case because our data shows a right-skewed distribution (see Appendix B Figure 2A). Therefore, a linear mixed model for normally distributed data might not be the best fitting approach to analyze our data. To test the robustness of our original analysis we repeated the analysis using a general linear mixed model (Ng & Cribbie, 2016) by utilizing the lme4 R package (v1.1.37; Bates et al., 2015).

Descriptive Statistics

On average, participants had lower prediction errors in their forecasts than the Kalman filter (M = 12.06, SD = 11.27; M = 22.81, SD= 20.66). Moreover, participants made similar prediction errors when forecasting positive and negative affect (M_{na} = 11.82, SD = 11.49; M_{pa} = 12.30, SD= 11.04). In contrast, the Kalman filter was more accurate in predicting positive than negative affect (M_{pa} = 12.29, SD = 10.41; M_{na} = 33.34, SD = 22.88). Likewise, both participants and the Kalman filter made similar prediction errors for short and long intervals (M_{short} = 11.45, SD= 10.62; M_{long} =12.68, SD = 11.85; $M_{short\ kalman}$ = 22.68, SD = 20.79; $M_{long\ kalman}$ = 22.95, SD = 20.53)

Table 1

Prediction Error of Humans and Kalman filter

	Human		Kalman filter	
	Mean	*SD*	Mean	*SD*
Overall	12.06	11.27	22.81	20.66
Positive Affect	12.30	11.04	12.29	10.41
Negative Affect	11.82	11.49	33.34	22.88
Short Intervals	11.45	10.62	22.68	20.79
Long Intervals	12.68	11.85	22.95	20.53

Human Prediction Accuracy Across Long and Short Intervals and Affect

For our first and second hypothesis we exclusively looked at human forecasts. The analysis revealed that the main effect of interval was significant. Consistent with our hypothesis, forecasts made over a shorter time span ($EMMS_{short}$ = 11.47) were slightly more accurate than those over a longer time span ($EMMS_{long}$ = 12.69; *B* = 1.18, *t*(3735) = 2.33, *p* = 0.0199). Yet, the difference is rather small.

For our second hypothesis, the model revealed that neither the main effect of affect (*B* = 0.43, *t*(3735) = 0.86, *p* = 0.392) nor the interaction effect between interval size and affect was significant (*B* = 0.01, t(3735) = 0.14, *p* = 0.892). That means participants' prediction accuracy did not differ when forecasting positive or negative affect. Additionally, the observed effect of interval length on prediction accuracy was the same for predicting positive and negative affect (see Figure 1). Thus, the data did not support our hypothesis. Although the estimates differ when running the analysis with a general mixed model, the results do not change in terms of significance (see Table 2 and Appendix B Table 1 and Figure 3). That is, both models lead to the same conclusion.

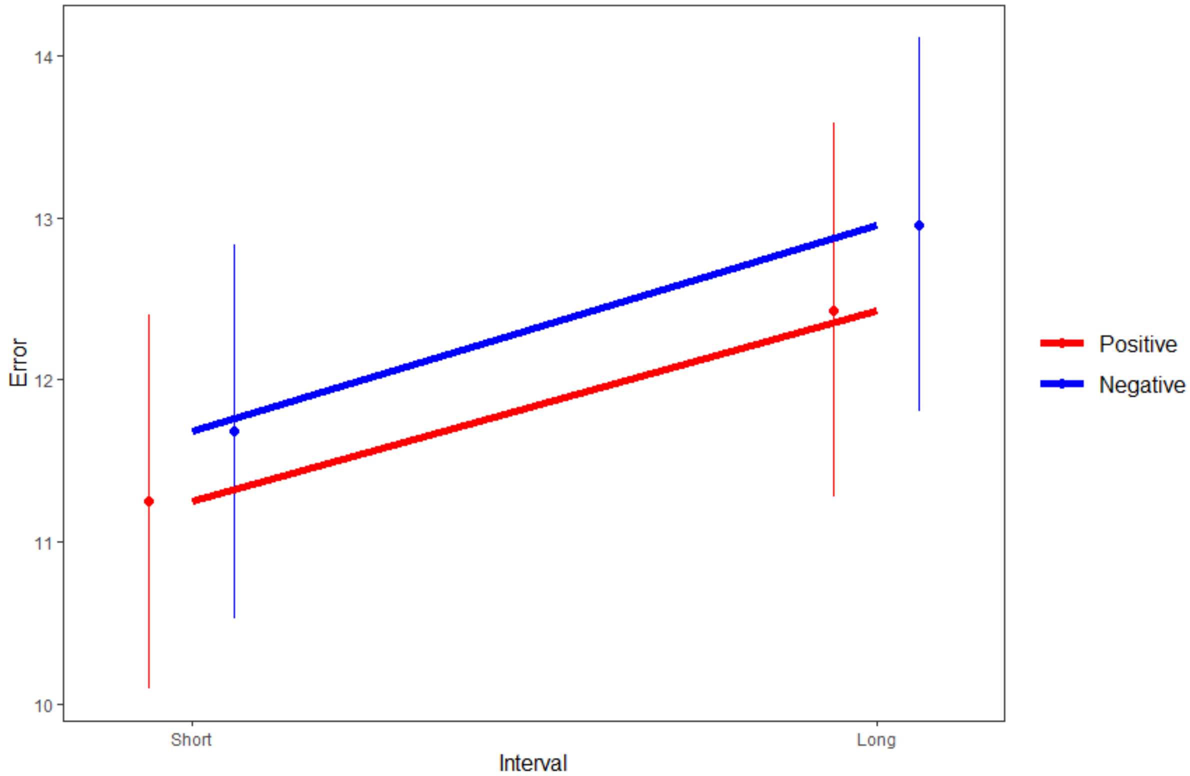
Table 2

Fixed Effects Table: Effect of Interval Size and Affect on Forecasting Error (Linear Mixed Model)

	Estimate	Std. Error	*df*	*t*	*p*
Intercept	11.25	.59	56.98	19.12	<.001
Interval	1.18	.51	3735	2.33	.0199
Affect	.43	.51	3735	.86	.392
Interval*Affect	.01	.72	3735	.14	.892

Figure 1

Change in Forecasting Error as a Function of Interval Size and Affective Valence (Linear Mixed Model)



Interaction between Interval Size and Source of Prediction

For our last hypothesis, we looked at both human and model forecasts. The analysis showed that only the main effect of interval size significantly affected prediction error (B = 1.23, t(7497.78) = 3.59, p < .001). However, neither the main effect of source of prediction (B = -0.39, t(7497.78) = -1.15, p = .250) nor the interaction between source and interval were significant (B = -0.28, t(7497.78) = -0.58, p = .564). Put differently, while both humans and the model made smaller errors when forecasting three-hour intervals compared to six-hour intervals, they did not significantly differ in how accurate their predictions were. Moreover, their accuracy did neither improve nor did it deteriorate depending on whether the predictions were made for short or long intervals (see Figure 2). Similar to the previous analysis, the outcomes of both statistical models allow for an identical inference (see Table 4 and Appendix B Table 2 and Figure 4).

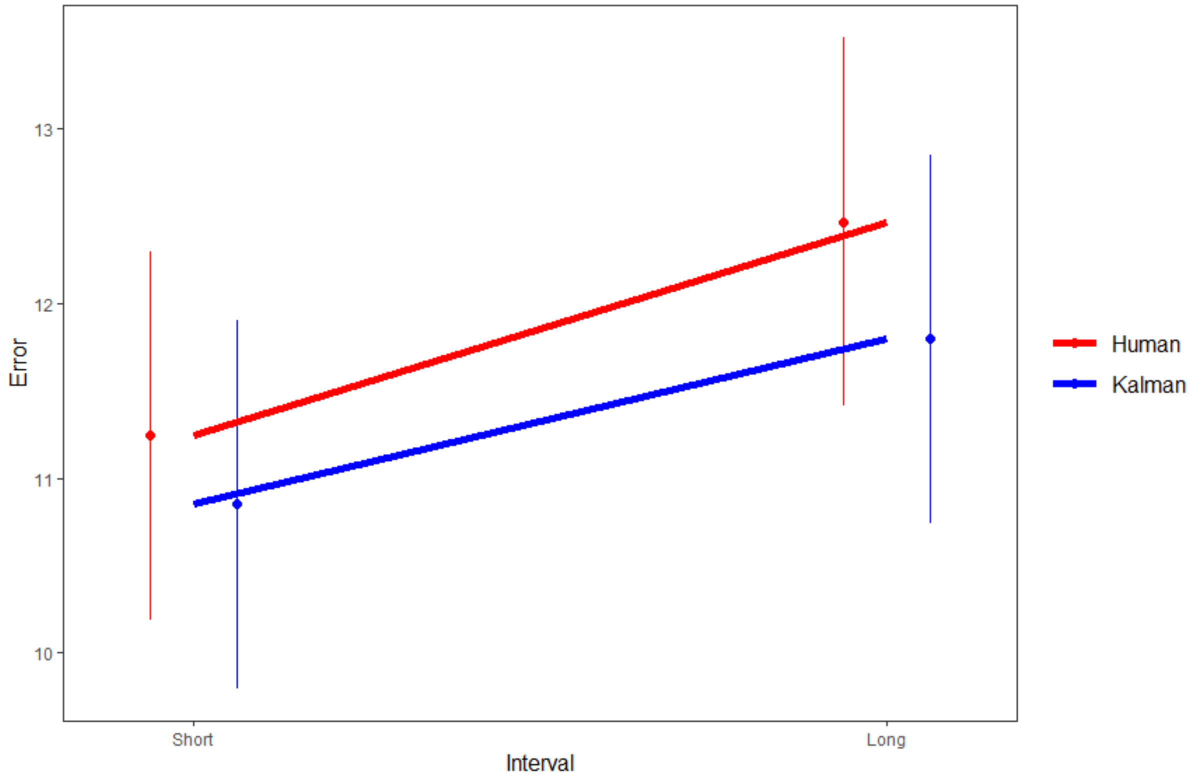
Table 4

Fixed Effects Table: Effect of Interval Size and Source of Prediction on Forecasting Error (Linear Mixed Model)

	Estimate	Std. Error	*df*	*t*	*p*
Intercept	11.24	.54	40.95	20.87	<.001
Interval	1.23	.34	7497.78	3.59	<.001
Source	- .39	.34	7497.78	-1.15	.250
Interval*Source	- .28	.48	7497.78	- .58	.564

Figure 2

Change in Forecasting Error as a Function of Interval and Source of Prediction (Linear Mixed Model)



Discussion

Affective forecasts influence decision making in a way that does not necessarily lead to decisions that are optimal for our well-being. That is because people tend to misjudge their future emotions and use these erroneous forecasts in part to guide their decision (Hoerger et al., 2016). Therefore, this paper aimed to examine which factors impact affective forecast accuracy and how they might interact. Specifically, whether human forecast errors are a function of different time horizons and affective valence, and how a statistical model compares to human prediction accuracy. The first research question explored how prediction accuracy may change for time points further in the future. The lower prediction errors for short intervals compared to long intervals were significant and aligned with our hypothesis suggesting that people may be better at predicting how they feel in the near than in the distant future. Our second research question aimed at investigating whether the observed change in forecasting accuracy for different intervals depends on the affective valence that is predicted. However, we found no significant evidence supporting our hypothesis as participants consistently predicted their emotions more accurately for short than long time spans regardless of the valence they predicted. Lastly, our third research question examined how forecasting accuracy may differ for humans and the Kalman filter across different time horizons. Surprisingly, we found no significant differences between the Kalman filter and human participants also irrespective of the interval length. Hence, we did not observe the hypothesized interaction effect.

The results from our first research question are consistent with previous literature that suggests that accuracy is higher for short intervals due to informational wealth (Braun & Yaniv, 1992). That is, as the interval size increases uncertainty also increases which makes it more difficult to correctly anticipate one's emotional state. Furthermore, a potential explanation for why we did not detect an interaction between interval size and valence may lie in methodological differences between the present study and the study we used as a basis for our hypothesis. First, Finkenauer et al. (2007) used greater interval lengths in their research (one to five days) while we looked into much shorter interval lengths (three to six hours). Second, and perhaps more important, they explored affective forecasting accuracy regarding taking their driver’s license which is a personally relevant event where they can either fail or succeed. Therefore, participants might have felt a greater need to engage in motivational biases or emotional regulation strategies (Finkenauer et al., 2007). In contrast, participants in our research project engaged in affective forecasts that concerned their typical daily live routines without asking them about any particular events. Given that we asked them more generally about their emotional affect, it seems likely that emotionally salient events were less present (Takano & Ehring, 2024). Consequently, as Finkenauer et al. (2007) reasoned, the valence-related biases such as coping by being overly optimistic about the outcomes of a temporally close event may not have come into play. Thus, the nonsignificant result in our study might indicate that this interaction between interval size and valence may only be present for very specific events with an actual outcome.

Additionally, a potential reason for the Kalman filter's poor performance could be that compared to humans, it does not have access to anticipatory knowledge and instead exclusively updates its predictions based on previous patterns of prediction error. However, humans also take information about what is about to happen into account when making affective forecasts (Kahneman & Snell, 1990 as cited in Colombo et al., 2020). For instance, knowing that you will meet a good friend later on, you may expect that to uplift your mood. Therefore, although the Kalman filter is theoretically able to adjust for prediction errors it cannot foresee changes in emotional states that occur because of an individual's personal agenda. Nevertheless, lack of anticipatory knowledge does not always lead to poor prediction accuracy as Takano and Ehring (2024) found that for hour-long intervals, the Kalman filter was still more accurate (absolute error) than humans despite not having had the advantage of anticipatory knowledge. In comparison, for minute-long intervals they did not find a significant difference between human and model forecasts (absolute error). This pattern suggests that knowledge of upcoming events does not necessarily translate into more accurate affective forecasts made by humans or the model and that it might not be needed for very short intervals. Possibly, because in that case relying on the most recent affect might be more informative (projection bias) (Takano & Ehring, 2024).

Limitations and Future Directions

Still, our findings should be considered in light of the study‘s strengths and limitations. The major strength of the present study is the inclusion of two future time points within a single study in an ESM context. While Finkenauer et al., (2007) examined different time distances within a single individual, they did so for a very specific event raising the question of how affective forecasts may unfold in a day-to-day context. Furthermore, Takano and Ehring (2024) also employed an ESM design but they did not test for the effect of different time distances on accuracy because they conducted two separate studies. We, however, investigated forecasting accuracy for different time distances within the same individual using ESM which allows for drawing conclusions about forecasting accuracy for different intervals in people's daily lives thereby enhancing ecological validity (Shiffman et al., 2008).

Despite this strength, our study also has some limitations. First, while we investigated short and long intervals (three and six hours), those still occur only in one day and thus could be considered rather short altogether. Although the chosen time frame is in line with Takano and Ehring‘ approach (2024), past research has operationalised long intervals over days or even months (e.g Buehler & McFarland, 2001) which might limit the validity of statements about long-term affective forecasts.

Second, like Takano and Ehring (2024), we did not randomize the order of rating and predicting one's affect. Since they always rated their current affect before they predicted their future affect, we might have inadvertently induced them to rely on their current emotions when engaging in the forecast (projection bias). Consequently, their current and predicted affect scores might have become rather similar but since we did not examine this relationship it cannot be determined to what extent this bias was indeed present. However, as they also reasoned, our study included repeated measurements so they would learn at some point that they will always have to give estimates for both current and future affect, making it nearly impossible to not make them aware of their current affect while engaging in the forecast (Takano & Ehring, 2024).

Lastly, in comparison to Takano and Ehring (2024), we used a significantly smaller sample size. It could be that the small sample size undermined the statistical power to detect our hypothesised interaction effects (Button et al., 2013). Nonetheless, whereas they averaged the data points across individuals to use a t-test -despite repeated measurements and a skewed distribution- we accommodated for the multilevel structure of our data by using the Gamma model. Thus, although it remains unclear whether a lack of power might have obscured potential interaction effects between interval size, source of prediction, and emotional valence our data analysis may offer a more appropriate approach to the data's structure.

Given the aforementioned limitations and alternative explanations, future research could aim at comparing human and model affective forecasts for an event with a particular outcome using different operationalisations of interval lengths in a longitudinal within-person design. Implementing such a design, directly addresses the limited validity of statements about long-term affective forecasts and enables drawing conclusions about how model performance compares to that of humans for situations in which valence-related biases are likely to be present (Finkennauer et al., 2007). Specifically, near events could be defined as taking place within the same or next day whereas distant events as taking place after seven to eight weeks. That interval could be suitable for investigating forecasts regarding upcoming exams as one block of the academic year in the Netherlands typically aligns with that time interval. Therefore, participants and a statistical model could be asked to predict positive and negative affect concerning their course grade at the beginning of the block and one day before the exam. Immediately after receiving their results, they then would be asked to rate their actual affect. In addition, measures of perceived controllability and actions taken to prepare for the exam should be collected throughout the block. This allows for exploring how potential erroneous forecasts together with perceived controllability might influence decisions made about preparation. For instance, students anticipating devastation upon failure and feeling low control over the outcome might be especially likely to avoid dealing with the subject (e.g Milgram & Toubiana, 1999) increasing the chances of failure. If findings reveal model forecasts to be more accurate, study advisors could make use of them to help such students setting more realistic expectations which in turn could help at making better preparation choices.

Conclusions

Taken together, this present study offers supporting evidence that prediction accuracy suffers when individuals make forecasts about the distant future. However this generalization is limited to predictions made within the same day. Yet, it seems to be a general trend that does not differ for the emotional valence that is predicted. Lastly, the Kalman filter's forecasting accuracy did not significantly differ from that of humans overall nor depending on forecasting affect for short or long intervals. However, these results should be interpreted with caution given the study's limitations.

References

Bates D., Maechler M., Bolker B., Walker S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.

Braun, P. A. & Yaniv, I. (1992). A case study of expert judgment: Economists' probabilities versus base-rate model forecasts. Journal Of Behavioral Decision Making, 5(3), 217–231. https://doi.org/10.1002/bdm.3960050306

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

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. Nature reviews. Neuroscience, 14(5), 365–376. https://doi.org/10.1038/nrn3475

Colombo, D., Fernández-Álvarez, J., Suso-Ribera, C., Cipresso, P., García-Palacios, A., Riva, G., & Botella, C. (2020). Biased affective forecasting: a potential mechanism that enhances resilience and Well-Being. Frontiers in Psychology, 11. https://doi.org/10.3389/fpsyg.2020.01333

Day, R., Ganz, P. A., Costantino, J. P., Cronin, W. M., Wickerham, D. L., & Fisher, B. (1999). Health-Related Quality of life and tamoxifen in breast cancer prevention: A report from the National Surgical Adjuvant Breast and Bowel Project P-1 study. Journal of Clinical Oncology, 17(9), 2659. https://doi.org/10.1200/jco.1999.17.9.2659

Eastwick, P. W., Finkel, E. J., Krishnamurti, T. & Loewenstein, G. (2007). Mispredicting distress following romantic breakup: Revealing the time course of the affective forecasting error. Journal Of Experimental Social Psychology, 44(3), 800–807. https://doi.org/10.1016/j.jesp.2007.07.001

Fagerlin, A., Dillard, A. J., Smith, D. M., Zikmund-Fisher, B. J., Pitsch, R., McClure, J. B., Greene, S., Alford, S. H., Nair, V., Hayes, D. F., Wiese, C., & Ubel, P. A. (2011). Women's interest in taking tamoxifen and raloxifene for breast cancer prevention: response to a tailored decision aid. Breast Cancer Research and Treatment, 127(3), 681–688. https://doi.org/10.1007/s10549-011-1450-1

Finkenauer, C., Gallucci, M., Van Dijk, W. W. & Pollmann, M. (2007). Investigating the Role of Time in Affective Forecasting: Temporal Influences on Forecasting Accuracy. Personality And Social Psychology Bulletin, 33(8), 1152–1166. https://doi.org/10.1177/0146167207303021

Göndöcs, D. & Dörfler, V. (2024). AI in medical diagnosis: AI prediction & human judgment. Artificial Intelligence in Medicine, 149, 102769. https://doi.org/10.1016/j.artmed.2024.102769

Hoerger, M., Scherer, L. D., & Fagerlin, A. (2016). Affective forecasting and medication decision making in breast-cancer prevention. Health Psychology, 35(6), 594–603. https://doi.org/10.1037/hea0000324

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. Psychological Review, 80(4), 237–251. https://doi.org/10.1037/h0034747

Kuznetsova A., Brockhoff P.B., Christensen R.H.B. (2017). LmerTest Package: Tests in Linear Mixed Effects Models. Journal of Statistical Software, 82(13), 1-26. doi:10.18637/jss.v082.i13<https://doi.org/10.18637/jss.v082.i13>.

Lüdecke D., Ben-Shachar M.S., Patil I., Waggoner P., Makowski D. (2021). Performance: An R Package for Assessment, Comparison and Testing of Statistical Models. Journal of Open Source Software, 6(60), 3139. https://doi.org/10.21105/joss.03139

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

Milgram, N., & Toubiana, Y. (1999). Academic anxiety, academic procrastination, and parental involvement in students and their parents. British Journal of Educational Psychology, 69(3), 345–361. https://doi.org/10.1348/000709999157761

Miller, M. D. & Ferris, D. G. (1993). Measurement of subjective phenomena in primary care research: the Visual Analogue Scale. PubMed, 13(1), 15–24. https://pubmed.ncbi.nlm.nih.gov/8484338

Moyer, V. A. (2013). Medications for risk reduction of primary breast cancer in women: U.S. Preventive Services Task Force Recommendation Statement. Annals of Internal Medicine. https://doi.org/10.7326/0003-4819-159-10-201311190-00718

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

Pennington, G. L., & Roese, N. J. (2003). Regulatory focus and temporal distance. *Journal of Experimental Social Psychology*, *39*(6), 563–576. https://doi.org/10.1016/S0022-1031(03)00058-1

R Core Team (2025). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. Annual review of clinical psychology, 4, 1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

Takano, K. & Ehring, T. (2024). Affective forecasting as an adaptive learning process. Emotion, 24(3), 795–807. https://doi.org/10.1037/emo0001303

Theobald, E. (2018). Students Are Rarely Independent: When, Why, and How to Use Random Effects in Discipline-Based Education Research. *CBE—Life Sciences Education*, *17*(3), rm2. https://doi.org/10.1187/cbe.17-12-0280

Wilson, T. D., & Gilbert, D. T. (2003). Affective forecasting. Advances in experimental social psychology, 35(35), 345-411.

Appendix A

Acknowledgement

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

I acknowledge the use of ChatGPT-4o (https://chatgpt.com/) to receive feedback on the clarity of my writing that is included in my work in a modified version.

I acknowledge the use of ChatGPT-4o (https://chatgpt.com/) to generate materials for self-study in the drafting of this assessment.

The following prompts were input into ChatGPT-4o (https://chatgpt.com/):

(1)

"Can you point out passages in the following paragraph that can be shortened?:
First, we wanted to compare how prediction accuracy may change for short intervals (three hours) and for long intervals (six hours) and investigate whether this pattern holds true for both positive and negative affect. Therefore, the dependent variable is prediction error, and the independent variables are the size of the interval and affective valence. To determine whether prediction error significantly differed between short and long intervals, a multilevel model was used with interval length and affect as fixed effects and a random effect for the intercept to account for individual differences in prediction error. Based on previous literature (Braun & Yaniv (1992) we expected smaller prediction errors for shorter than longer intervals. In addition, we expected that for short intervals, the prediction error would be smaller for forecasting negative than positive affect. In contrast, for longer intervals, prediction errors would be smaller for forecasting positive than negative affect(Finkenauer et al., 2007). Second, we investigated whether there is an interaction effect between the source of prediction (Kalman vs. human) and interval size (short vs. long) We expected the Kalman

Filter to make more accurate predictions for longer intervals, but to be similarly accurate as humans for shorter intervals (Braun & Yaniv, 1992)."

The output obtained was:

Here's a version with **highlighted suggestions for shortening or clarifying** without changing meaning:

Suggestion:

First, ***we compared*** *prediction accuracy across short (three hours) and long (six hours) intervals, and examined whether this pattern held for both positive and negative affect.*

(Shortens by combining and simplifying phrasing.)

Suggestion:

A multilevel model tested whether *prediction error differed by interval length and affect, with a random intercept to account for individual differences.*

(Makes the sentence more concise and active.)

The output was changed by me in the following ways: I cut out unnecessary phrasings in my initial version like "to determine whether prediction error significantly differed between short and long intervals, a multilevel model was used."and instead only wrote "For all the analyses we conducted multilevel analyses." Likewise instead of writing "we wanted to compare…" I reduced it to "We compared…". However, besides that I sticked to my own phrasings

Final output:

For all the analyses we used R Statistical Software […] **to conduct multilevel model analyses** First, **we compared** how human prediction accuracy may change for short intervals (three hours) and for long intervals (six hours) and investigated whether this pattern holds true for both positive and negative affect. Interval size and affective valence were included in the model as fixed effects and prediction error as dependent variable. Based on previous literature (Braun & Yaniv (1992), we expected smaller prediction errors for shorter than longer intervals. In addition, for short intervals we expected that the prediction error would be smaller for forecasting negative than positive affect. In contrast, for longer intervals, prediction errors would be smaller for forecasting positive than negative affect (Finkenauer et al., 2007). [...]

(2)

"Can you make suggestions on how to express the following idea in simpler words: Additionally, the observed effect of interval length on prediction accuracy was independent of predicting positive or negative affect."

The output obtained was:

Of course! Here are a few simpler rephrasings:

1. **"Also, the effect of interval length on prediction accuracy did not depend on whether people were predicting positive or negative feelings."**

2. **"In addition, the impact of how long the interval was on prediction accuracy stayed the same for both positive and negative emotions."**

3. **"The influence of interval length on how accurate the predictions were was the same whether the feelings were positive or negative."**

4. **"Interval length affected prediction accuracy in the same way for both positive and negative emotions."**

The output was changed by me in the following ways: Based on the second output I replaced “independent” with “was the same” to visualize it more clearly to the reader but kept everything else like in my original writing.

Final output: Additionally, the observed effect of interval length on prediction accuracy **was the same** for predicting positive and negative affect

(3)

“Can you make suggestions on how to write the following paragraph more clearly?: Due to their notions of how emotions affect performance people tend to minimize how rewarding success may feel and instead exaggerate how painful failure may be. As a result, people perceive failure to be worse than it actually might be which may fuel their motivation to work consistently on pursuing their goal. In line with that, people overestimate the negative impact and the duration of failure on their well-being while correctly anticipating that happiness upon success will fade rapidly. In contrast, for near events people have less control over an outcome because they cannot change the course of events anymore as they could for distant events. Therefore, they prepare themselves mentally by increasing how good they expect to feel in case they do succeed and by downplaying how bad they might feel to weaken the impact of a possible negative outcome beforehand.”

The output obtained was:

Here’s a **clearer and smoother revision** of your paragraph:

"Finkenauer et al. (2007) suggest that when an event is still far in the future, people perceive themselves as having more control over its outcome. Because of this, they tend to think strategically: they downplay how rewarding success will feel and exaggerate how painful failure might be. This biased way of thinking may serve a motivational function—it makes the prospect of failure seem like a significant threat, which could encourage people to invest more effort in order to prevent that outcome. As a result, people overestimate how badly failure will affect their well-being and how long this negative feeling will last, while at the same time anticipating (accurately) that any happiness from success will be brief. In contrast, when the event is close in time, people generally perceive that they have little ability left to influence its outcome. Instead of motivating action, their focus shifts to managing their emotional response: they boost their expectations of how good success will feel and reduce their expectations of how bad failure might be, in order to soften the blow of a possible negative outcome."

The output was changed by me in the following ways: In my paragraph I used different phrases to express the same idea: For instance, instead of "while at the same time anticipating any happiness from success will be brief ", I wrote "while correctly anticipating that the happiness felt after succeeding will fade rapidly" or instead of "Because of this, they tend to think strategically" I wrote " To facilitate realizing this goal, they engage in biased thinking."

Final Output:

They reasoned that for events that are far in the future people feel more in control and thus, focus their efforts on achieving a positive outcome. To facilitate realizing this goal, they engage in biased thinking. That is, they downplay how rewarding success may feel and instead exaggerate how dreadful failure would be. As a result people perceive failure to be more threatening than it actually is which in turn helps them to stay motivated to avoid defeat. In terms of forecasting accuracy, people then overestimate the duration of their negative feelings following failure while correctly anticipating that the happiness felt after succeeding will fade rapidly. In contrast, for near events people believe to have less control to influence an outcome. Therefore, they try to shift their efforts to regulating their emotional response. They prepare themselves mentally by de-emphasizing how bad failure might be to weaken the impact of a possible negative outcome beforehand. Simultaneously, they up-regulate their expectation of how good success would feel. Consequently, failure is perceived to be temporary. Thus, they accurately predict that their negative affect will be short-lived but overestimate the duration of happiness in case of a positive outcome. Taken together, for temporally distant events inaccurate forecasts of negative affect following failure may be a motivational source to keep working to achieve one‘s goals. In comparison, for temporally close events, inaccurate forecasts of positive affect may serve as a coping mechanism for possible unpleasant outcomes.

Appendix B

Robustness Check with General Mixed Model

Figure 1

Assumption Check Linear Mixed Model

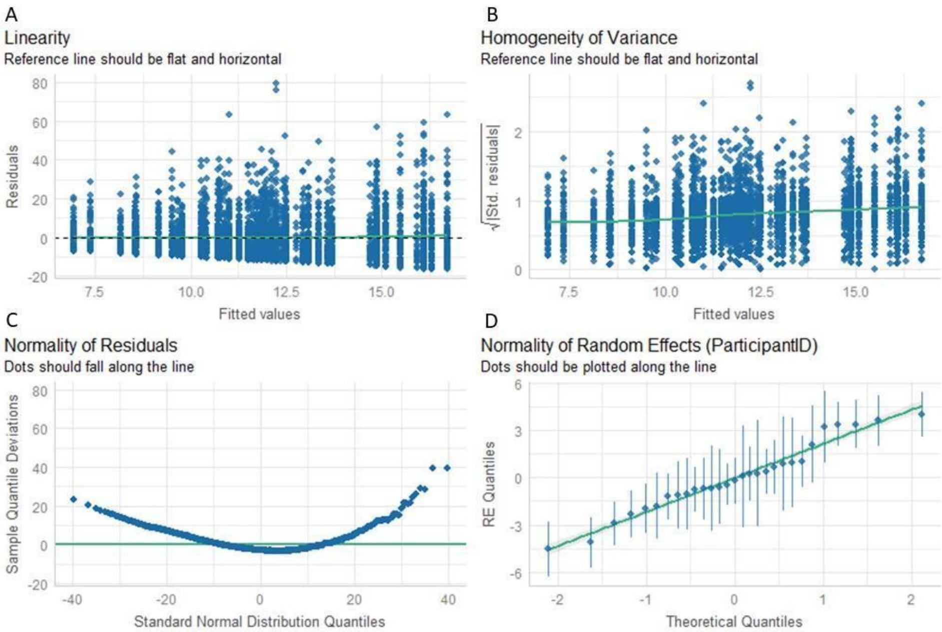


Figure 2

Assumption Check General Mixed Model

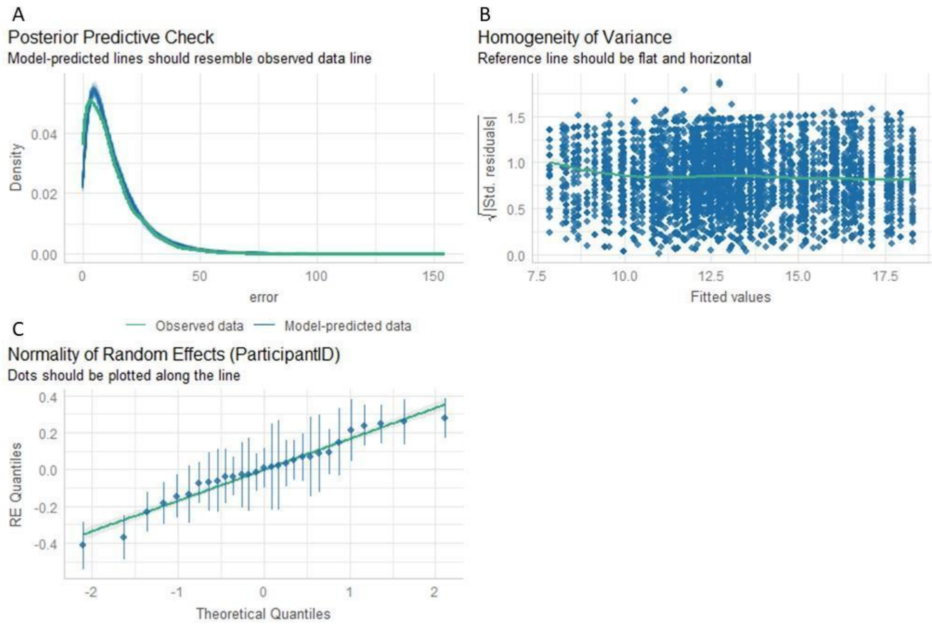


Table 1

Fixed Effects Table: Effect of Interval Size and Affect on Forecasting Error (General Mixed Model)

	Estimate	Std. Error	*t*	*p*
Intercept	2.477	.047	53.108	<.001
Interval	.084	.038	2.214	.0268
Affect	.056	.038	1.473	.141
Interval*Affe ct	.010	.054	.189	.850

Figure 3

Change in Forecasting Error as a Function of Interval Length and Affective Valence (General Mixed Model)

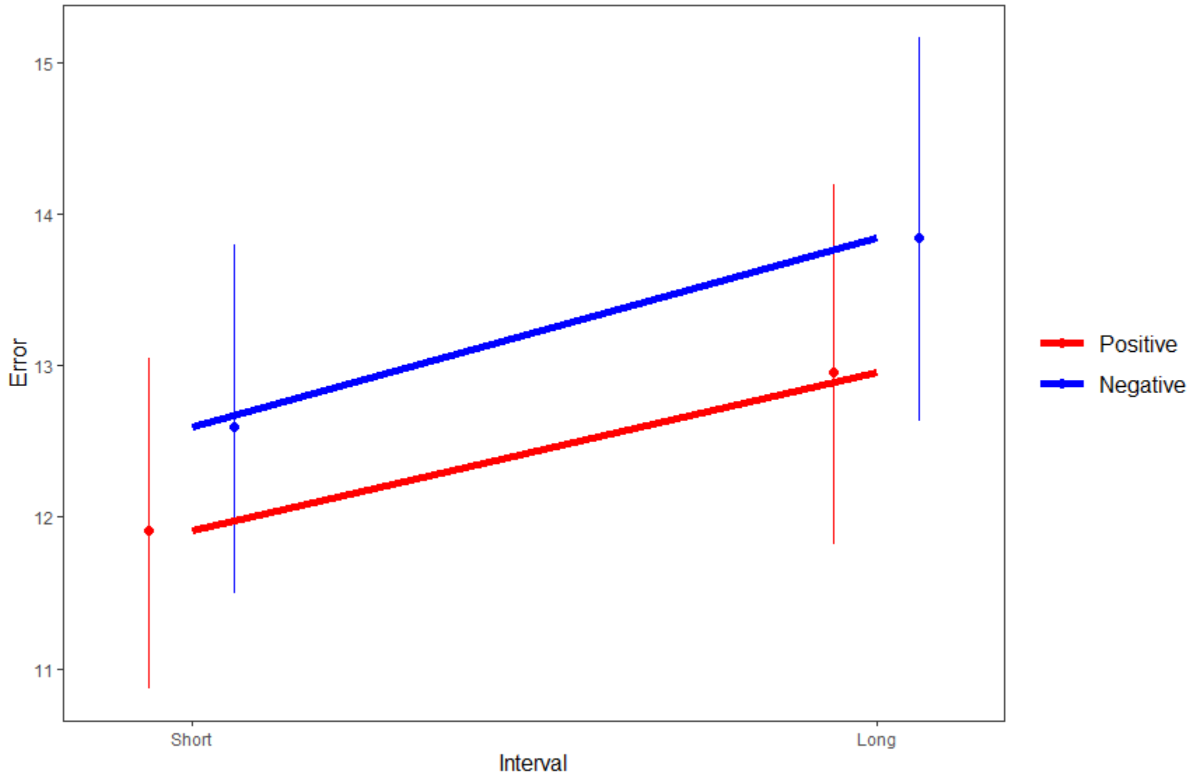


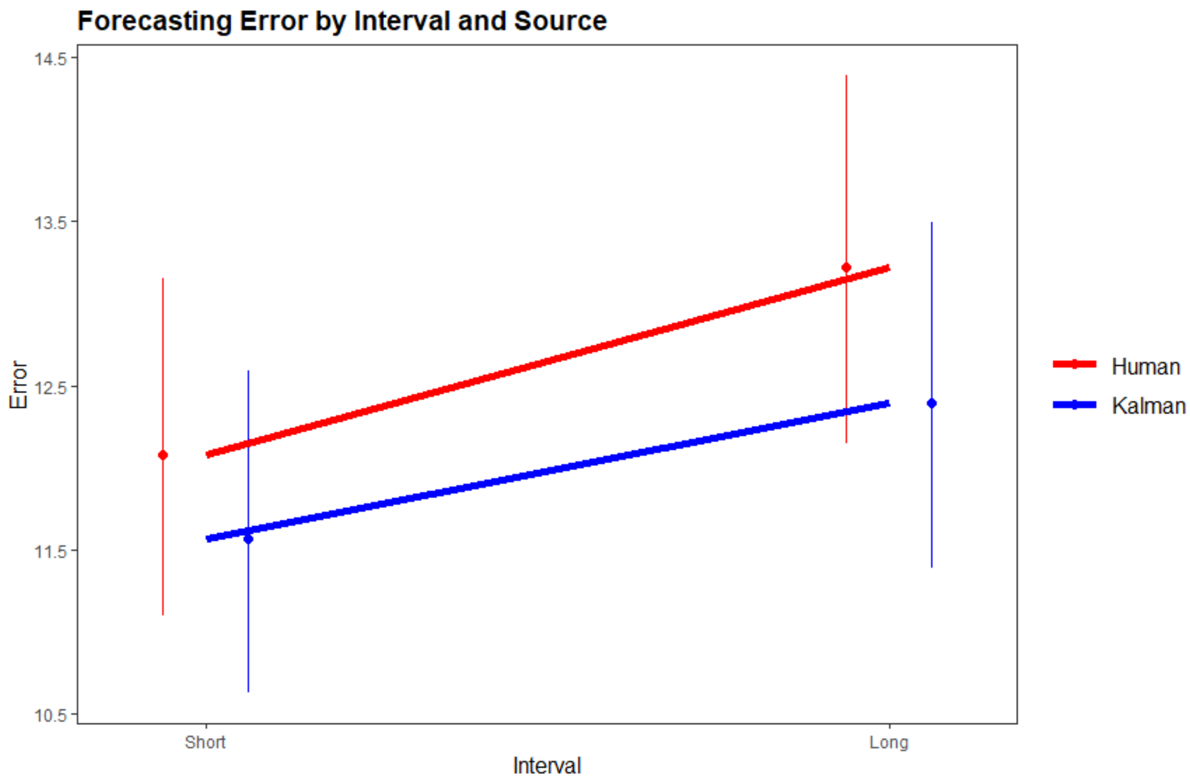
Table 2

Fixed Effects Table: Effect of Interval Size and Source of Prediction on Forecasting Error(General Mixed Model)

	Estimate	Std. Error	*t*	*p*
Intercept	2.492	.043	57.714	<.001
Interval	.090	.026	3.462	<.001
Source	- .043	.026	-1.666	.096
Interval*Sou rce	- .021	.037	- .574	.566

Figure 4

Change in Forecasting Error as a Function of Interval Length and Source of Prediction Error (General Mixed Model)



Appendix C

R Code Used for the Analysis

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

# creating error variables for the Kalman filter.

data <- data %>%
  # important otherwise it gets carried forward from one person to the next
  group_by(ParticipantID) %>%
  mutate(PA_kalman_two_step = lag(PA_kalman_one_step),
```

```
    NA_kalman_two_step = lag(NA_kalman_one_step))

# check here that the prediction is the same but one step carried forward
head(data[,c("ParticipantID","day_n","time_n","PA_now_point",
"PA_kalman_one_step","PA_kalman_two_step")])

#creating interval variables
# Short interval (3 hours)
data$PA_error_human_short <- abs(data$PA_now_point - data$PA_one_step_point_match)
data$NA_error_human_short <- abs(data$NA_now_point - data$NA_one_step_point_match)
data$PA_error_kalman_short <- abs(data$PA_now_point - data$PA_kalman_one_step)
data$NA_error_kalman_short <- abs(data$NA_now_point - data$NA_kalman_one_step)

# Long interval (6 hours)
data$PA_error_human_long <- abs(data$PA_now_point - data$PA_two_step_point_match)
data$NA_error_human_long <- abs(data$NA_now_point - data$NA_two_step_point_match)

data$PA_error_kalman_long <- abs(data$PA_now_point - data$PA_kalman_two_step)
data$NA_error_kalman_long <- abs(data$NA_now_point - data$NA_kalman_two_step)

long_data <- data %>%
  #add filter
  filter(!is.na(PA_now_point) & !is.na(PA_one_step_point_match) &
!is.na(PA_two_step_point_match)) %>%
  select(ParticipantID, day_n, time_n,
       PA_error_human_short, PA_error_human_long,
       NA_error_human_short, NA_error_human_long,
       PA_error_kalman_short, NA_error_kalman_short,
       PA_error_kalman_long,NA_error_kalman_long) %>%

  pivot_longer(
    cols = -c(ParticipantID, day_n, time_n),
    names_to = c("affect", "source", "interval"),
    names_pattern = "(PA|NA)_error_(human|kalman)_(short|long)?",
    values_to = "error"
  ) %>%

  mutate(
    affect = recode(affect, "PA" = "positive", "NA" = "negative"),
    interval = factor(interval, levels = c("short", "long")),
    interval = factor(interval, levels = c("short", "long")),
    source = factor(source, levels = c("human", "kalman")),
    ParticipantID = as.factor(ParticipantID)
  )
```

```
# check whether data matches

data %>% filter(ParticipantID ==1 & day_n ==1 &time_n == 3) %>%
  select(ParticipantID, time_n,
        PA_error_human_short, PA_error_human_long,
        NA_error_human_short, NA_error_human_long,
        PA_error_kalman_short, NA_error_kalman_short,
        PA_error_kalman_long,NA_error_kalman_long)


long_data %>%  filter(ParticipantID ==1 & day_n ==1 &time_n == 3)

write.csv(long_data_clean, "code_students/dat_fabienne.csv", row.names = FALSE)

library(lmerTest)
library(lattice) ## for dotplot
library(sjPlot)

# First filter to only have the predictions of the humans for the first and second hypothesis

long_dat_human <- long_data %>% filter(source == "human")


mod_affect_interval <- lmerTest::lmer( error ~ interval*affect  + (1 | ParticipantID),
                      data = long_dat_human)

#assumption check
summary(mod_affect_interval)
performance::check_model(mod_affect_interval, panel = T,check =
c("reqq","linearity","qq","homogeneity"))
tab_model(mod_affect_interval)

library(ggeffects)

ggemmeans(mod_affect_interval, terms = c("interval","affect")) %>%
  plot(log_y = F)+
  geom_line (size = 2) +theme_bw()

#Analysis with General Linear Mixed Model

library(lme4)

mod_affect_interval_glm <- glmer( error +1  ~ interval*affect  + (1 | ParticipantID),
```

```
                                  data = long_dat_human, family = Gamma(link = "log"))

#assumption check
summary(mod_affect_interval_glm)
performance::check_model(mod_affect_interval_glm, panel = T,check =
c("reqq","pp_check","qq","homogeneity"))
tab_model(mod_affect_interval_glm)

ggemmeans(mod_affect_interval_glm, terms = c("interval","affect")) %>%
  plot(log_y = F)+
  geom_line (size = 2) +theme_bw()


# third hypothesis

mod_source_interval <- lmerTest::lmer( error ~ interval*source  + (1 | ParticipantID),
                                  data = long_data)

summary(mod_source_interval)
performance::check_model(mod_source_interval, panel = T,check =
c("reqq","linearity","qq","homogeneity"))

#Analysis with General Linear Mixed Model

mod_source_interval_glm <- glmer( error + 1 ~ interval*source  + (1 | ParticipantID),
                                  data = long_data, family = Gamma(link = "log"))

summary(mod_source_interval_glm)
performance::check_model(mod_source_interval_glm, panel = T,check =
c("reqq","pp_check","qq","homogeneity"))
tab_model(mod_source_interval_glm)

ggemmeans(mod_source_interval_glm, terms = c("interval","source")) %>%
  plot(log_y = F)+
  geom_line (size = 2) +theme_bw()
```

---

[1] Copied from the AI guide used at the University College London

https://www.ucl.ac.uk/students/exams-and-assessments/assessment-success-guide/engaging-ai-your-education-and-assessment