Optimistic amnesia: how online and offline processing shape belief updating and memory biases in immediate and long-term optimism biases

Ziqing Yao,[1] Xuanyi Lin,[1] and Xiaoqing Hu[1,2,3]

[1]Department of Psychology, The University of Hong Kong, Hong Kong S.A.R., China, [2]The State Key Laboratory of Brain and Cognitive Sciences, The University of Hong Kong, Hong Kong S.A.R., China, and [3]HKU, Shenzhen Institute of Research and Innovation, Shenzhen 518057, China

Correspondence should be addressed to Xiaoqing Hu, Department of Psychology, The University of Hong Kong, Pokfulam, Hong Kong S.A.R., China. E-mail: xiaoqinghu@hku.hk.

Abstract

When people are confronted with feedback that counters their prior beliefs, they preferentially rely on desirable rather than undesirable feedback in belief updating, i.e. an optimism bias. In two pre-registered EEG studies employing an adverse life event probability estimation task, we investigated the neurocognitive processes that support the formation and the change of optimism biases in immediate and 24 h delayed tests. We found that optimistic belief updating biases not only emerged immediately but also became significantly larger after 24 h, suggesting an active role of valence-dependent offline consolidation processes in the change of optimism biases. Participants also showed optimistic memory biases: they were less accurate in remembering undesirable than desirable feedback probabilities, with inferior memories of undesirable feedback associated with lower belief updating in the delayed test. Examining event-related brain potentials (ERPs) revealed that desirability of feedback biased initial encoding: desirable feedback elicited larger P300s than undesirable feedback, with larger P300 amplitudes predicting both higher belief updating and memory accuracies. These results suggest that desirability of feedback could bias both online and offline memory-related processes such as encoding and consolidation, with both processes contributing to the formation and change of optimism biases.

Key words: optimism bias; belief updating; offline processing; P300; motivated cognition

Introduction

People view their future via an optimistic lens such that they believe good things are more likely to happen to themselves than bad things (Weinstein, 1980). Optimism biases can be evident in valence-dependent belief updating, whereby people preferentially use desirable over undesirable feedback to update their prior beliefs (Eil and Rao, 2011; Sharot *et al.*, 2011; Sharot and Garrett, 2016; Kuzmanovic and Rigoux, 2017; Kuzmanovic *et al.*, 2018). Despite the accumulating evidence on optimism biases, this research has mostly focused on how individuals update beliefs immediately after receiving feedback. However, belief updating may also require the retrieval of previously

encountered information. To date, much remains unknown regarding how newly encountered feedback will influence long-term judgments in belief updating. In the present study, we aimed to examine how optimism biases change over time, and how electrophysiological brain activities support the formation and change of optimism biases.

Following initial online processing, e.g. encoding recently learned information continues to be processed during offline periods wherein memories are selectively retained or pruned via covert reactivation and consolidation processes (Rasch and Born, 2013; Stickgold and Walker, 2013; Tambini and Davachi, 2019; Hu et al., 2020). Related to optimism bias, feedback desirability plays a key role in subsequent belief updating: neuroimaging findings showed that processing desirable (vs undesirable) feedback engaged brain regions implicated in emotional processing (amygdala) and error tracking (e.g. right inferior frontal gyrus; Sharot et al., 2011). Furthermore, favorable beliefs updating in response to both desirable and undesirable feedback engaged brain regions involved in subjective valuation (e.g. ventral medial prefrontal cortex; Kuzmanovic et al., 2018). Given that motivational salience would bias offline consolidation processes (Payne and Kensinger, 2018), we hypothesized that desirability would continue to bias offline processing of feedback over time and would lead to even larger optimism biases in longer term, delayed tests. It would further be worthwhile to scrutinize whether such changes are due to (i) strengthening of desirable information, (ii) weakening of undesirable information, or (iii) both. To examine initial feedback processing and to link it with long-term belief updating, we examined feedback-related event-related brain potentials (ERPs) during the belief updating task. Aided by ERPs' unparalleled, millisecond temporal resolution, we focused on three well-established ERP components along feedback processing stream: the relatively early frontocentral feedback-related negativity (FRN), the parietal P300 and the late positive potential (LPP).

FRN, also termed as feedback negativity or medial frontal negativity, is a negative-going ERP component with a frontocentral distribution that peaks during 200–400 ms following feedback onset (San Martín, 2012). FRN is typically larger following negative/loss feedback than positive/reward feedback (Yeung and Sanfey, 2004; Hajcak et al., 2006). Sensitivity of FRN to valence-/reward-related processes makes it an ideal candidate to examine valence-dependent belief updating that also engages subjective valuation (e.g. Kuzmanovic et al., 2018, 2016; Sharot and Garrett, 2016). While the FRN could encode unsigned salience prediction errors (e.g. both positive or negative reward prediction errors, see Talmi et al., 2013), accumulating evidence strongly suggests that the FRNs are enhanced following negative than positive prediction errors (Chase et al., 2011; Walsh and Anderson, 2012; Sambrook and Goslin, 2015; Heydari and Holroyd, 2016). Accordingly, if the brain encodes undesirable feedback as negative reward prediction errors, undesirable feedback should elicit larger FRNs than desirable feedback. Moreover, given that FRN has been implicated in behavioral adjustment following prediction errors (Holroyd and Coles, 2002; Hauser et al., 2014), we would also examine whether FRNs predict post-feedback belief updating.

P300 is one of the most studied ERP components that reflects a variety of processes including novelty/salience detection, stimuli categorization, context updating, memory encoding, retrieval, etc. (Azizian and Polich, 2007; Polich, 2007, 2012). Specifically, P300s during incidental encoding have been linked with subsequent memories (Paller et al., 1987; Otten and Rugg, 2001; Paller and Wagner, 2002; Polich, 2012), which is of particular interest to the present investigation. That is, P300s may not only capture binary valence information (desirable vs undesirable feedback, Wu and Zhou, 2009) but may also encode specific feedback probabilities to be used in subsequent belief updating and memory retrieval (Marciano et al., 2018). Lastly, we planned to examine LPP, a prolonged ERP activity that usually sustains over longer time windows, e.g. 1000–2000 ms (Schupp et al., 2006). Previous studies suggest that LPPs reflect attentional and emotional responses to motivationally salient stimuli, with greater LPPs reflecting arousal and sustained emotional responses (Hajcak and Foti, 2020). In the context of feedback processing, LPP could reflect relatively late elaboration of the feedback probabilities.

The present research has two primary goals. First, we aimed to investigate how optimism biases change over time from immediate to delayed tests. To capture different aspects of optimism biases, we assessed how desirability of feedback would influence individuals' belief updating, as well as participants' memories of feedback probabilities. Central to our research question regarding how offline processing may contribute to delayed optimism biases, we measured optimistic belief updating and memory biases both immediately and 24 h after receiving feedback. Second, we aimed to investigate ERP activities underlying feedback processing, and how ERPs may influence optimistic belief updating and memory biases. We examined the information processing stream from early, rapid valence coding (200–400 ms frontocentral FRN), to encoding (300–800 ms parietal P300) and to relatively late elaboration processes (800–1600 ms LPP).

Methods

Preregistrations/data/scripts/materials are available at https://osf.io/f2qjv/.

Participants

We preregistered two EEG studies ($n = 20$, 25, respectively) with identical procedures/materials (osf.io/wqjvh and osf.io/c8fyx). Results are reported based on the second preregistered analytical plans with combined samples to increase statistical power (total $n = 45$, 14 males, mean ± s.d., 20.22 ± 1.74 years old). To replicate behavioral findings of the EEG studies, we preregistered a third behavioral study ($n = 30$, 12 males, 19.70 ± 1.39 years old, osf.io/r5uqk) using a modified belief updating task (Kuzmanovic and Rigoux, 2017). Results of the third behavioral study are reported in Supplementary Online Materials (SOM).

Participants were recruited from the University of Hong Kong and received either partial course credit or monetary incentives (60 HKD/h). An additional 19 participants from the EEG studies were excluded (see SOM). Participants with normal/corrected-to-normal vision and without psychiatric/neurological/sleep disorder were included in the experiment. Participants were pre-screened based on Insomnia Severity Index (scores <10; Bastien et al., 2001) and on Beck Depression Inventory-II (BDI-II, scores <28, minimal-to-moderate depressive severity; Beck et al., 1996). The study was approved by the Human Research Ethics Committee of The University of Hong Kong. All participants provided written informed consent before the study.
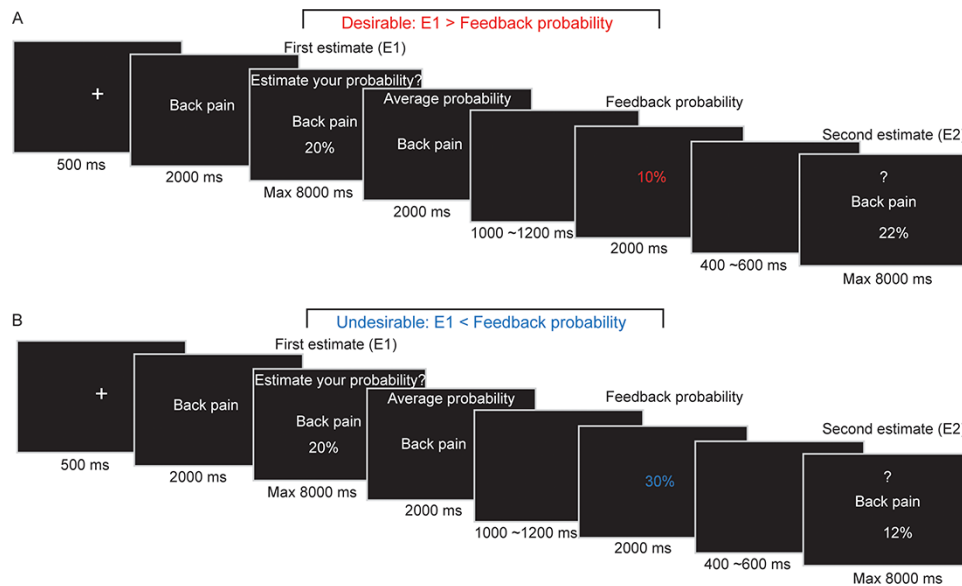
Fig. 1. A trial flow in the life event probability estimation task across two conditions. Each trial began with a 0.5 s fixation on the center of the screen, followed by a life event presented for 2 s. Participants were given 8 s to estimate the probability (E1) of them experiencing the specific event at least once in the future. Feedback probability of the same event happening to people of the same gender/age/socio-cultural background was then presented for 2 s, during which participants' ERPs were analyzed. Following an inter-stimulus interval varying randomly between 0.4 and 0.6 s, participants were instructed to give an estimation of the same event for a second time (E2) within 8 s. (A) Desirable condition. (B) Undesirable condition.

Design and procedure

Tasks were programed using E-Prime® 3.0 (Psychology Software Tools, Inc., Sharpsburg, Pennsylvania, USA). Each participant completed two experimental sessions, separated by a 24 h delay.

Session 1: Following EEG setup, participants completed the Stanford Sleepiness Scale (SSS, Hoddes *et al.*, 1973) to measure their alertness levels, and the Positive and Negative Affect Schedule (PANAS). Participants next completed a life event probability estimation task (Figure 1), during which they estimated the probability of experiencing adverse life events in the future. Eighty negative life events were randomly assigned to the desirable and undesirable feedback conditions, anew for each participant. Six additional negative life events were used as practice trials before the task.

Upon reading each negative life event, participants were given 8 s to enter their first estimation of its probability within the range of 1% and 99% (E1). Trials were excluded from analyses if no response was registered within this window. Feedback (1% to 90%) was then presented in red/blue color to indicate its desirability (i.e. desirable when Feedback < E1; undesirable when Feedback > E1), with color assignments counterbalanced across participants. Feedback was presented in white when Feedback = E1. Unbeknownst to participants, feedback probability was manipulated via subtracting or adding a random number to the initial estimation (see SOM). Feedback was presented for 2 s, followed by prompts that required participants to give a second probability estimate of the same negative life event (E2) within 8 s. Estimation and feedback were presented within the same trial to reduce memory load, see Kuzmanovic and Rigoux (2017).

Participants then completed a surprise cued recall task, during which they were presented with the same negative life events, and were asked to recall the feedback probability presented earlier. After the recall task, participants finished the Life Orientation Test to assess trait optimism (Scheier *et al.*, 1994),

the BDI-II (Beck *et al.*, 1996), the State-Trait Anxiety Inventory (STAI, Spielberger, 1983) and the Rumination on Sadness Scale (Conway *et al.*, 2000).

Session 2: After 24 h, participants returned to the lab for session 2. Participants first finished the SSS, PANAS and the STAI state subscale. They then estimated the probabilities of the same 80 negative life events (i.e. E3), followed by the same cued recall task as in session 1. Participants then rated each life event along dimensions such as vividness, familiarity, etc., on a 6-point scale. Finally, they completed the Pittsburgh Sleep Quality Scale (Buysse *et al.*, 1989), the five-item Morningness–Eveningness Questionnaire (Adan and Almirall, 1991) and answered three questions to probe suspicion regarding the feedback manipulation (see SOM).

EEG acquisition and pre-processing

Continuous EEGs were recorded using a 64-channel Waveguard cap connected to an eego amplifier (ANT Neuro, Enschede, Netherlands), with electrodes mounted and fixed according to the 10–20 System (ANT Neuro, Enschede, Netherlands). The online sampling rate was 500 Hz, with electrode AFz as the ground and CPz as the online reference. Horizontal electroocu-lograms (EOGs) were recorded from an electrode placed 1.5 cm next to the left canthus. Impedance of all electrodes was kept below 20 kΩ during the recording.

Raw EEGs were processed using the EEGLAB toolbox (Delorme and Makeig, 2004) and ERPLAB toolbox (Lopez-Calderon and Luck, 2014) in Matlab (Mathworks Inc, Natick, MA, USA). EEG data from EOG, M1 and M2 electrodes were first removed from further analyses and were down-sampled to 250 Hz, then band-pass filtered between 0.05 and 30 Hz using the default IIR Butterworth filter and notch filtered at 50 Hz implemented in ERPLAB. Bad channels were replaced using interpolation. EEG data were then re-referenced to the whole brain average, after

which interpolated channels were removed. Continuous EEG data were segmented into −200 to 2000 ms epochs relative to the onset of feedback, before being subjected to Independent Component Analysis (ICA). Eye-/muscle-movement artifacts were identified and corrected using visual inspection of ICA components in ICLabel toolbox (Pion-Tonachini et al., 2019). Epochs were excluded if amplitudes exceeded a threshold of ±75 μV, leaving 34.76 (s.d. = 4.80) and 37.22 (s.d. = 4.53) trials for desirable and undesirable conditions, respectively.

For ERP quantifications, the −200 to 2000 ms EEG epochs were baseline corrected using the mean amplitude of the 200 ms pre-stimulus baselines. We examined three a priori defined ERP components as preregistered: the FRN, P300 and LPP. The FRN was calculated as the mean amplitudes within the 200–350 ms/250–400 ms time windows, also taking the mean of the most negative 50-ms segment within these two time windows (i.e. adaptive means) at the frontocentral cluster (Fz, FCz, F1/2 and FC1/2). The P300 was calculated as the mean amplitude of the 300–800 ms time window, and as the mean of the most positive 100 ms segment within this time window at the left- (P1/3, CP1/3), the midline (Cz/Pz) and the right-parietal clusters (P2/4, CP2/4). LPP was calculated as the mean amplitude within the 800–1600 ms time window, also taking the mean of the most positive 100 ms segment within this time window. We extracted LPP amplitudes from both frontocentral and parietal clusters as in the FRN and P300 quantifications.

Preregistered behavioral analyses

Definitions of behavioral variables and their definitions are presented in Table 1.

Linear mixed-effects models (LMMs) were conducted with trial-level data to account for trial-by-trial variances and covariates (e.g. E1, E2, event ratings, Baayen et al., 2008; Meteyard and Davies, 2020; Schad et al., 2020). LMMs were conducted using the lme4 package (Bates et al., 2015), with P-values computed using Satterthwaite's approximation in the lmertest package (Kuznetsova et al., 2017) in R (Team, 2019). Post-hoc pairwise comparisons were conducted using the emmeans package with Tukey corrections (Lenth et al., 2018). The optimal model was selected via the R package buildmer (Voeten, 2020) based on the Akaike information criterion value using backward stepwise elimination. Both time (i.e. immediate vs delayed) and desirability (i.e. desirable vs undesirable) were coded using a sum contrast with 0.5 for immediate/desirable and −0.5 for delayed/undesirable to test main effects and interactions. E1 and event ratings were z-scored within each participant and were included as covariates. Event ID, the trial number difference between desirable and undesirable conditions, estimation errors (EE) and participant ID were coded as random effects. The inclusion of intercept (and/or slopes) over random effects was determined by the optimal model (see SOM for the maximal model). The statistical significance criterion was set at 0.05.

The optimal updating LMM that provided the best fit of the data was specified as follows: updating ∼1 + desirability * time + E1 + personal relevance + (1 + desirability | EE) + (1 + desirability + time | subject)

The optimal memory error LMM that provided the best fit of the data was specified as follows: memory error ∼1 + desirability * time + E1 + personal relevance + prior experience + negativity + familiarity + (1 + desirability | subject) + (1 + desirability | EE) + (1 + desirability | event ID) + (1 | trial number difference).

Correlational analyses and results from analysis of variance (ANOVA) and analysis of covariance (ANCOVA) using participant-level data are reported in SOM.

Preregistered ERP analyses

Paired sample t-tests comparing desirable vs undesirable conditions were conducted on FRN, P300 and LPP amplitudes.

Table 1. Descriptive statistics of behavioral variables

Variables	Mean (s.d.)		P	dz	Formula
	Desirable	Undesirable			
Numbers of trials	36.93 (4.73)	40.02 (3.33)	0.008	0.41	NA
Feedback (FB)	23.36 (17.41)	32.67 (16.71)	<0.001	1.27	Algorithms see SOM
1st estimation (E1)	30.35 (19.2)	24.22 (15.63)	<0.001	0.81	NA
2nd estimation (E2)	24.79 (17.91)	28.72 (17.42)	0.001	0.52	NA
3rd estimation (E3)	24.92 (18.43)	22.76 (16.79)	0.01	0.4	NA
Estimation error (EE)	6.99 (2.36)	8.45 (1.78)	<0.001	1.37	EE = |FB − E1|
Immediate updating	5.56 (2.64)	4.5 (2.78)	0.008	0.41	D: E1–E2; UD: E2–E1
Delayed updating	5.43 (3.93)	−1.45 (4.46)	<0.001	0.99	D: E1–E3; UD: E3–E1
Immediate memory error	10.05 (4.25)	12.23 (4.22)	<0.001	0.73	|PresentedFB − RecallFB|
Delayed memory error	10.8 (4.51)	14.3 (5.24)	<0.001	0.94	|PresentedFB − RecallFB|
RT E1 in ms	2774 (951)	2593 (953)	0.004	0.46	NA
RT E2 in ms	1781 (679)	1773 (732)	0.868	0.02	NA
RT E3 in ms	2087 (805)	2063 (899)	0.629	0.07	NA
RT M1 in ms	3693 (818)	3771 (876)	0.155	0.22	NA
RT M2 in ms	3618 (898)	3594 (907)	0.652	0.07	NA
Emotional arousal	3.12 (0.75)	3 (0.78)	0.038	0.32	NA
Familiarity	3.38 (0.69)	3.22 (0.67)	0.003	0.47	NA
Negativity	3.86 (0.86)	3.83 (0.94)	0.667	0.06	NA
Prior experience	1.48 (0.45)	1.33 (0.27)	0.001	0.51	NA
Vividness	3.5 (0.82)	3.21 (0.83)	<0.001	0.72	NA
Personal relevance	3.74 (1.05)	3.54 (1.11)	0.001	0.54	NA

P-values and Cohen's dz were calculated based on desirable vs undesirable, paired sample t-tests. D, desirable updating; UD, undesirable updating.

Bayesian factors favoring the null over alternative hypothesis (BF_{01}) are reported to provide strength of evidence supporting non-significant results with 0.707 as Cauchy prior.

Preregistered ERP-behavioral analyses

LMM analyses were conducted to examine how single-trial ERPs predicted belief updating and memory errors, with centered mean amplitudes of ERPs (FRN, P300 and LPP) and desirability as fixed effects. Desirability and time were coded using a sum contrast with 0.5 for desirable/immediate and −0.5 for undesirable/delayed for testing main effects and interactions. E1 and event ratings were z-scored within each participant and were included as covariates. Event ID, trial number difference between desirable and undesirable conditions, trial-level EE and participant ID were coded as random effects. The inclusion of intercept (and/or slopes) over random effects was determined by the optimal model (see SOM for the maximal model). The optimal model that could be applied to all ERPs is:

1) Immediate/delayed updating $\sim 1 + $ E1 $+$ prior experience $+$ personal relevance $+$ vividness $+$ familiarity $+$ desirability $+$ ERPs $+$ (1 | EE) $+$ (1 $+$ desirability | subject).
2) Immediate/delayed memory error $\sim 1 + $ E1 $+$ prior experience $+$ personal relevance $+$ desirability * ERPs $+$ (1 | subject) $+$ (1 $+$ desirability | EE) $+$ (1 | event ID) $+$ (1 | trial number difference)

Results

Preregistered behavioral analyses

Descriptive of variables are presented in Table 1. For belief updating, the LMM results indicated significant main effects of desirability and time: participants showed reduced updating following undesirable *vs* desirable feedback, $\beta = 7.10$, $SE = 0.89$, $P < 0.001$ and reduced updating in the delayed *vs* immediate test, $\beta = 3.07$, $SE = 0.39$ and $P < 0.001$. Critically, the desirability × time interaction was significant: $\beta = -5.63$, $SE = 0.55$ and $P < 0.001$. Post-hoc comparisons showed that belief updating significantly declined from immediate to delayed test in the undesirable, $\beta = 5.88$, $SE = 0.47$ and $P < 0.001$, but not in the desirable condition, $\beta = 0.25$, $SE = 0.48$ and $P = 0.601$. When focusing on how updating biases (i.e. desirable minus undesirable belief updating) changed across time, we found that updating biases became significantly larger in the delayed than in the immediate test: $\beta = -2.81$, $SE = 0.28$ and $P < 0.001$ (Figure 2A).

For memory errors, the LMM showed significant desirability and time effects: participants were more accurate in remembering desirable *vs* undesirable feedback, $\beta = -5.08$, $SE = 0.83$ and $P < 0.001$ and were more accurate in the immediate than in the delayed test, $\beta = -1.40$, $SE = 0.28$ and $P < 0.001$. Critically, the desirability × time interaction was significant, $\beta = 1.25$, $SE = 0.56$ and $P = 0.026$. Post-hoc comparisons revealed that memories for feedback significantly declined from immediate to delayed test in the undesirable condition, $\beta = -2.02$, $SE = 0.39$ and $P < 0.001$, but only declined marginally in the desirable condition, $\beta = -0.78$, $SE = 0.40$ and $P = 0.055$. When focusing on how memory biases (i.e. desirable minus undesirable memory errors) changed across time, results showed that memory biases became significantly larger from immediate to delayed test: $\beta = 0.62$, $SE = 0.28$ and $P = 0.026$ (Figure 2B). Similar results were obtained when using ANOVA/ANCOVA with participant-level data (see SOM for details).

Non-preregistered behavioral analyses

Instead of recalling feedback probabilities, participants may recall their second estimation (E2) in the cued recall tasks and in the delayed probability estimation task. When repeating the delayed belief updating/memory error LMM with E2 as a covariate, the delayed optimistic belief updating biases remained significant ($\beta = 6.70$, $S.E. = 0.92$ and $P < 0.001$). Similarly, in predicting memory errors, main effects of desirability and time (*ps* < 0.001) and their interaction were significant ($\beta = 1.25$, $S.E. = 0.56$, $P = 0.026$). Therefore, controlling for E2 did not influence valence-dependent updating and memory biases.

To test whether updating biases were still evident when differences of memory errors were controlled for, we added z-scored memory errors as a covariate in the abovementioned LMM. Results confirmed the significant main effects of desirability and time (*ps* < 0.001), Critically, the desirability × time interaction remained significant: $\beta = -5.46$, $SE = 0.55$ and $P < 0.001$: optimistic updating biases became significantly larger in the delayed than in the immediate test: $\beta = -2.73$, $SE = 0.27$ and $P < 0.001$). Thus, optimistic belief updating effects remained significant when memory errors were statistically controlled.
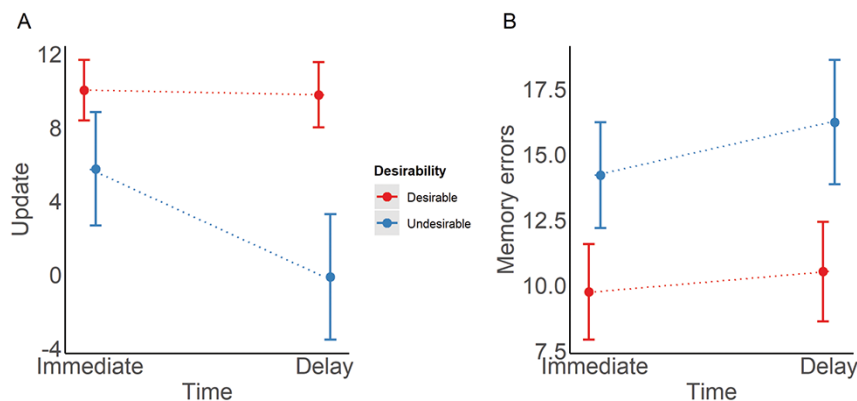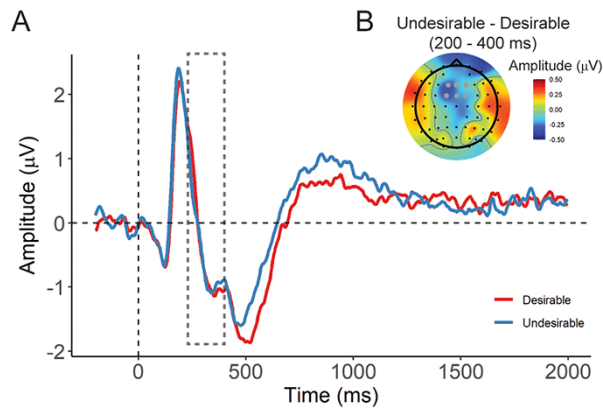

Fig. 2. Predicted values of belief updating and memory errors. (A). Participants showed significant optimistic belief updating, which became larger after a 24 h delay. (B). Participants showed worse memories (higher memory errors) for undesirable than desirable feedback, i.e. an optimistic amnesia effect. Error bars indicate 95% confidence intervals (CIs).

Fig. 3. FRN over the frontocentral electrodes. (A). Grand averaged ERPs over the frontocentral cluster (Fz, FCz, F1/2 and FC1/2). There were no significant FRN differences between desirable and undesirable conditions. (B). The topography of mean amplitudes of undesirable–desirable ERP difference waves between 200 and 400 ms.

Regarding the relationship between the size of EE and belief updating in the LMM, we found that larger EEs led to higher belief updating. Importantly, this relationship was significantly stronger in desirable than in undesirable condition for both immediate and delayed tests ($ps < 0.001$). Thus, participants were more likely to update their beliefs in response to larger estimations in the desirable (vs undesirable) condition. Full results are reported in SOM.

Preregistered ERP analyses

For feedback-locked FRN, paired sample t-tests revealed no significant ERP between desirable vs undesirable feedback (all $ps > 0.503$, all $BF_{01} > 4$, Figure 3A,B).

For P300s, we found that desirable feedback elicited significantly larger P300 amplitudes compared to undesirable feedback across all three clusters (300–800 ms mean amplitudes: left-parietal, $t(44) = 2.36$, $P = 0.023$ and $dz = 0.35$; midline,

$t(44) = 2.08$, $P = 0.043$, $dz = 0.31$; and right-parietal, $t(44) = 2.21$, $P = 0.033$ and $dz = 0.33$, Figure 4A–C). Results were similar when we used the most positive 100 ms segment of the left-parietal, $t(44) = 2.27$, $P = 0.028$ and $dz = 0.34$, and the midline P300, $t(44) = 2.12$, $P = 0.040$ and $dz = 0.32$, but not with right parietal cluster, $t(44) = 1.48$, $P = 0.146$, $dz = 0.22$ and $BF_{01} = 2.26$.

For LPPs, we found significantly larger LPP amplitudes in desirable vs undesirable condition when using the most positive 100 ms segment of the left-parietal cluster, $t(44) = 2.03$, $P = 0.048$ and $dz = 0.30$, but neither with the frontocentral, midline, or right-parietal clusters nor the mean amplitudes (800–1600 ms) across four clusters (all $ps > 0.104$, all $BF_{01} > 1$).

Preregistered ERP-behavioral analyses

Regarding belief updating, the LMM showed that enhanced left-parietal P300s and LPPs significantly predicted increased immediate belief updating in both desirable and undesirable conditions (P300: $\beta = 0.05$, $SE = 0.02$ and $P = 0.016$, LPP: $\beta = 0.04$, $SE = 0.02$ and $P = 0.024$, Figure 5A,B). We also found that enhanced right-parietal LPPs significantly predicted increased delayed belief updating, $\beta = 0.11$, $SE = 0.05$ and $P = 0.039$, Figure 5C. No other significant associations were found for immediate or delayed updating (all $ps > 0.196$).

In predicting memory errors, LMM analyses revealed a significant interaction between desirability and left-parietal P300 ($\beta = 0.24$, $SE = 0.10$ and $P = 0.017$, Figure 6A) and LPP ($\beta = 0.17$, $SE = 0.09$ and $P = 0.042$, Figure 6D): increased P300/LPP amplitudes were associated with decreased immediate memory errors in the undesirable condition (P300: $\beta = -0.19$, $SE = 0.07$, $P = 0.009$; LPP: $\beta = -0.15$, $SE = 0.06$ and $P = 0.011$) but not in the desirable condition ($ps > 0.499$). At midline cluster, no significant results were found on P300s ($P = 0.084$, Figure 6B); whereas enhanced LPP predicted reduced immediate memory errors in both desirable and undesirable conditions, $\beta = -0.08$, $SE = 0.04$ and $P = 0.036$, Figure 6E. Lastly, enhanced right-parietal P300s/LPPs significantly predicted reduced immediate memory errors in both conditions (P300: $\beta = -0.10$, $SE = 0.05$ and $P = 0.048$; LPP: $\beta = -0.09$, $SE = 0.04$ and $P = 0.018$, Figure 6C,F). No other interactions or main effects were significant (all $ps > 0.075$).
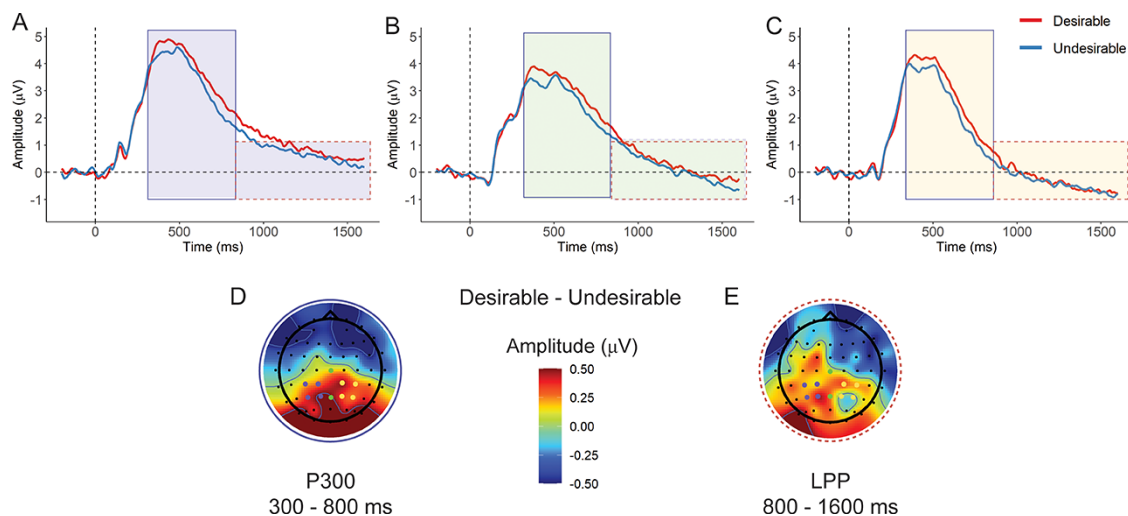
Fig. 4. Grand averaged P300 (300–800 ms) and LPP (800–1600 ms) over the parietal electrodes. (A) left-parietal cluster (P1/3, CP1/3), (B) midline cluster (Cz/Pz) and (C) right-parietal cluster (P2/4, CP2/4). Desirable feedback elicited significantly higher P300s than undesirable feedback across left-/right-parietal and midline clusters. The topography of the mean amplitude of desirable–undesirable difference waves for (D) P300 and (E) LPP.
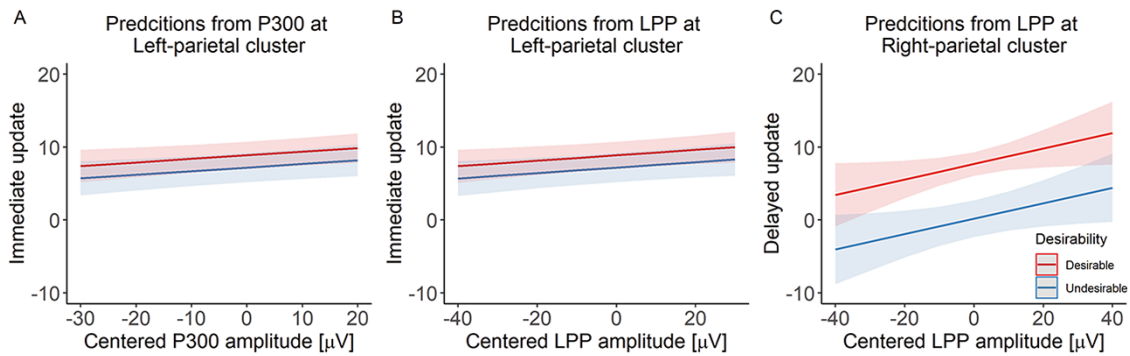
Fig. 5. Single-trial P300 and LPP predicted belief updating in both desirable and undesirable conditions. Enhanced (A) P300s and (B) LPPs over the left-parietal cluster predicted higher immediate belief updating. (C) Enhanced LPPs over the right-parietal cluster predicted higher delayed belief updating. The shaded areas represent 95% CIs. For visualization purposes, predictions from ERPs are plotted separately for both desirable and undesirable conditions.
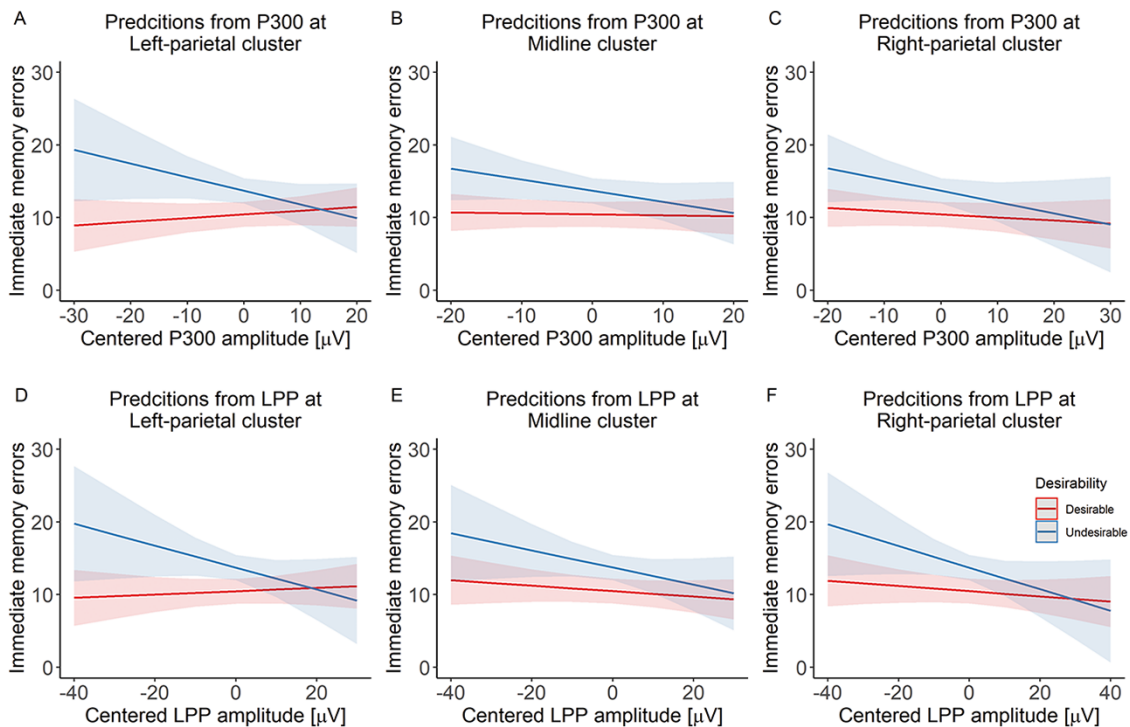

Fig. 6. Single-trial P300s/LPPs predicted immediate memory errors. (A), (D) Enhanced left-parietal P300/LPP amplitudes predicted reduced immediate memory errors only to undesirable feedback. (B), (E) While midline P300 effects were not significant, enhanced midline LPP amplitudes significantly predicted reduced immediate memory errors in both desirable and undesirable feedback. (C), (F) Enhanced right-parietal P300/LPP amplitudes predicted reduced immediate memory errors to both desirable and undesirable feedback. Shaded areas represent 95% CIs.

We next explored whether the size of EEs may interact with desirability in influencing ERPs in LMM models. Results confirmed a significant desirability effect on P300 amplitudes as reported in the paired sample *t*-tests (all *p*s < 0.018). Importantly, larger size of EEs predicted reduced midline P300 amplitudes only to desirable feedback. Given the well-established, inverse relationship between P300 amplitudes and cognitive efforts (Polich, 2007), this result suggested that in the desirable condition, participants would devote greater cognitive effort in updating their beliefs in response to larger EEs. Full results are reported in SOM.

Discussion

Encountering feedback that challenges one's prior beliefs, people preferentially rely on desirable than undesirable feedback to guide belief updating, i.e. the optimism bias (Sharot *et al.*, 2011; Sharot and Garrett, 2016; Dricu *et al.*, 2020). Here, we provide novel evidence that optimism biases are partially driven by shallower encoding and inferior memories of undesirable *vs* desirable feedback, i.e. an optimistic amnesia effect. Moreover, we observed that optimistic updating biases became larger over time, with preferential retention of updating in the desirable

condition and declined updating in the undesirable condition. Desirability of feedback consistently modulated parietal P300 brain activities that may indicate encoding depth, with larger P300s for desirable than undesirable probability feedback.

The present research provides the first evidence that optimism biases become larger over 24 h. This finding is noteworthy because it suggests that the desirability of feedback not only influences online attention/encoding-related processes but also biases offline consolidation processes. A closer inspection of our data suggested that over time, belief updating and memories of desirable feedback were largely preserved, whereas updating and memories significantly declined for undesirable feedback. These findings contribute to a growing literature suggesting that motivation (e.g. valence and reward) could bias offline consolidation processes and then influence long-term judgments (Rasch and Born, 2013; Stickgold and Walker, 2013; Payne and Kensinger, 2018).

Our findings that belief updating and memories changed more significantly in the undesirable but not in the desirable condition provide additional evidence that optimism bias is primarily driven by insufficient updating when receiving undesirable feedback (Eil and Rao, 2011; Sharot et al., 2011). Prior research found that self-related undesirable updating was not only lower than self-related desirable updating but also lower than other-related updating in general (Kuzmanovic et al., 2016). Moreover, aging participants showed reduced belief updating following undesirable feedback compared to young adults, leading to larger optimism biases (Chowdhury et al., 2014). In contrast, patients with major depressive disorder or individuals with high functioning autism showed enhanced belief updating toward undesirable feedback relative to healthy controls, leading to smaller optimism biases (Garrett et al., 2014; Korn et al., 2014; Kuzmanovic et al., 2019). These findings, together with our novel results on delayed belief updating and memory biases, consistently suggest that insufficient updating in response to undesirable feedback is a fundamental mechanism that drives immediate and long-term optimism biases.

Tracking ERPs allows us to investigate how the desirability of feedback biases information processing along millisecond temporal scale. We found that the desirability of feedback significantly modulated P300, and to a less extent, the LPP but not the FRN. As one of the most investigated ERP components, P300 has been associated with a range of cognitive processes, including context updating, motivational salience, evaluation and categorization, encoding depth, etc. (Azizian and Polich, 2007; Polich, 2007, 2012). In the present study, enhanced parietal P300s to desirable *vs* undesirable feedback suggested that participants preferentially encoded desirable feedback, which then exerted a greater impact on subsequent belief updating and memory performance. Regarding the LPP effect, prior research suggested that LPP may reflect in-depth elaboration of motivationally salient stimuli (Hajcak and Foti, 2020). Indeed, multi-level analyses with trial-level data showed that enhanced P300 and LPPs to feedback predicted larger belief updating and more accurate memories of feedback probabilities, substantiating the putative role of the P300/LPP in encoding and elaboration processes (Otten and Donchin, 2000; Otten and Rugg, 2001; Kamp et al., 2015; Rigney et al., 2020). These ERPs results also suggest that differential processing of desirable and undesirable feedback can occur rapidly after the initial valence processing.

We hypothesized that the frontocentral FRN would encode the desirability of feedback, with larger FRNs elicited by undesirable *vs* desirable feedback (Yeung and Sanfey, 2004; Heydari and Holroyd, 2016). However, the desirability of feedback was not observed to modulate FRN. The insensitivity of FRN to feedback valence in the belief updating task raises the possibility that EEs and reward prediction errors could reflect distinctive computational processes of error tracking (Sharot and Garrett, 2016). Specifically, FRNs are typically observed in reward-processing tasks during which feedback conveys monetary gains and losses (Hajcak et al., 2006; Proudfit, 2015; Heydari and Holroyd, 2016), whereas feedback in our task indicated numerical discrepancies between estimations of probabilities. Specifically, participants in the belief updating task needed to calculate the discrepancies between feedback probabilities and their initial estimations to guide belief updating. Such high-level inferential and calculation processes might make FRNs insensitive to the desirability of feedback in the present context. Moreover, FRNs have been suggested to be involved in both error tracking and behavioral adjustment. For example, in reward tasks, FRNs elicited by undesirable feedback (e.g. a loss) could guide behavioral adjustment to avoid losses (Holroyd and Coles, 2002; Cohen et al., 2007; Walsh and Anderson, 2012; Hu et al., 2015). However, in our belief updating task, participants preferentially used desirable rather than undesirable feedback to guide belief updating. The FRNs may reflect complex motivational-cognitive processes including both error tracking (in response to both desirable and undesirable feedback) and the signaling of behavioral adjustment (i.e. in response to desirable feedback). A mixture of these motivational-cognitive processes may thus lead to comparable FRNs in both desirable and undesirable conditions.

Regarding the delayed optimism biases, although our results suggest that time delay and offline processes contributed to the enhancement of optimism biases, it remains unknown whether sleep or wakefulness may differentially influence belief updating and memory biases. On one hand, enhancement of optimism biases may be time dependent rather than sleep dependent. Alternatively, given that sleep plays an important role in memory consolidation (Rasch and Born, 2013), sleep-based consolidation may be necessary for optimism biases to change. Future studies shall directly link sleep, offline consolidation processes and optimism biases to test this novel hypothesis.

Conclusion

To the best of our knowledge, the present preregistered studies are first to offer evidence regarding long-term optimism biases. Over time, optimism biases became larger, which were driven by significantly reduced belief updating following undesirable feedback. Both ERP and behavioral results highlighted roles of valence-dependent online and offline memory-related processes (e.g. encoding/consolidation) in the formation and change of optimism biases. Thus, the desirability of feedback not only biases initial online encoding but also subsequent offline consolidation processes, with both processes contributing to long-term optimism biases.

Acknowledgements

Funding

Career Schema (No. 27610617), and General Research Fund (No. 17601318) of the Hong Kong Research Grants Council to X.H.

Conflict of interest

The authors declare no conflicts of interest.

Supplementary data

Supplementary data are available at *SCAN* online.

References

Adan, A., Almirall, H. (1991). Horne & Östberg morningness-eveningness questionnaire: a reduced scale. *Personality and Individual Differences*, **12**(3), 241–53.

Azizian, A., Polich, J. (2007). Evidence for attentional gradient in the serial position memory curve from event-related potentials. *Journal of Cognitive Neuroscience*, **19**(12), 2071–81.

Baayen, R.H., Davidson, D.J., Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, **59**(4), 390–412.

Bastien, C.H., Vallières, A., Morin, C.M. (2001). Validation of the Insomnia Severity Index as an outcome measure for insomnia research. *Sleep Medicine*, **2**(4), 297–307.

Bates, D., Mächler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, **67**(1), 1–48.

Beck, A.T., Steer, R.A., Brown, G.K. (1996). Beck depression inventory–II. *San Antonio*, **78**, 490–98.

Buysse, D.J., Reynolds III, C.F., Monk, T.H., Berman, S.R., Kupfer, D.J. (1989). The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Research*, **28**(2), 193–213.

Chase, H.W., Swainson, R., Durham, L., Benham, L., Cools, R. (2011). Feedback-related negativity codes prediction error but not behavioral adjustment during probabilistic reversal learning. *Journal of Cognitive Neuroscience*, **23**(4), 936–46.

Chowdhury, R., Sharot, T., Wolfe, T., Düzel, E., Dolan, R.J. (2014). Optimistic update bias increases in older age. *Psychological Medicine*, **44**(9), 2003–12.

Cohen, M.X., Elger, C.E., Ranganath, C. (2007). Reward expectation modulates feedback-related negativity and EEG spectra. *NeuroImage*, **35**(2), 968–78.

Conway, M., Csank, P.A., Holm, S.L., Blake, C.K. (2000). On assessing individual differences in rumination on sadness. *Journal of Personality Assessment*, **75**(3), 404–25.

Delorme, A., Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, **134**(1), 9–21.

Dricu, M., Kress, L., and Aue, T. (2020). The neurophysiological basis of optimism bias. In: Aue, T., and Okon-Singer, H., editors. *Cognitive biases in health and psychiatric disorders: Neurophysiological foundations*, San Diego, CA: Elsevier, 41–70.

Eil, D., Rao, J.M. (2011). The good news-bad news effect: asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics*, **3**(2), 114–38.

Garrett, N., Sharot, T., Faulkner, P., Korn, C.W., Roiser, J.P., Dolan, R.J. (2014). Losing the rose tinted glasses: neural substrates of unbiased belief updating in depression. *Frontiers in Human Neuroscience*, **8**, 639.

Hajcak, G., Moser, J.S., Holroyd, C.B., Simons, R.F. (2006). The feedback-related negativity reflects the binary evaluation of good versus bad outcomes. *Biological Psychology*, **71**(2), 148–54.

Hajcak, G., Foti, D. (2020). Significance? Significance! Empirical, methodological, and theoretical connections between the late positive potential and P300 as neural responses to stimulus significance: an integrative review. *Psychophysiology*, e13570. https://www.x-mol.com/paperRedirect/1249530166222151680.

Hauser, T.U., Iannaccone, R., Stämpfli, P., et al. (2014). The feedback-related negativity (FRN) revisited: new insights into the localization, meaning and network organization. *NeuroImage*, **84**, 159–68.

Heydari, S., Holroyd, C.B. (2016). Reward positivity: reward prediction error or salience prediction error? *Psychophysiology*, **53**(8), 1185–92.

Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., and Dement, W. C. (1973). Quantification of sleepiness: a new approach. *Psychophysiology*, **10**(4) 431–36.

Holroyd, C.B., Coles, M.G. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, **109**(4), 679.

Hu, X., Pornpattananangkul, N., Nusslock, R. (2015). Executive control-and reward-related neural processes associated with the opportunity to engage in voluntary dishonest moral decision making. *Cognitive, Affective and Behavioral Neuroscience*, **15**(2), 475–91.

Hu, X., Cheng, L.Y., Chiu, M.H., Paller, K.A. (2020). Promoting memory consolidation during sleep: a meta-analysis of targeted memory reactivation. *Psychological Bulletin*, **146**(3), 218.

Kamp, S.M., Potts, G.F., Donchin, E. (2015). On the roles of distinctiveness and semantic expectancies in episodic encoding of emotional words. *Psychophysiology*, **52**(12), 1599–609.

Korn, C., Sharot, T., Walter, H., Heekeren, H., Dolan, R.J. (2014). Depression is related to an absence of optimistically biased belief updating about future life events. *Psychological Medicine*, **44**(3), 579–92.

Kuzmanovic, B., Jefferson, A., Vogeley, K. (2016). The role of the neural reward circuitry in self-referential optimistic belief updates. *NeuroImage*, **133**, 151–62.

Kuzmanovic, B., Rigoux, L., Tittgemeyer, M. (2018). Influence of vmPFC on dmPFC predicts valence-guided belief formation. *Journal of Neuroscience*, **38**(37), 7996–8010.

Kuzmanovic, B., Rigoux, L., Vogeley, K. (2019). Brief report: reduced optimism bias in self-referential belief updating in high-functioning autism. *Journal of Autism and Developmental Disorders*, **49**(7), 2990–8.

Kuzmanovic, B., Rigoux, L. (2017). Valence-dependent belief updating: computational validation. *Frontiers in Psychology*, **8**, 1087.

Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software*, **82**, 13.

Lenth, R. (2018). Emmeans: Estimated marginal means, aka least-squares means. *R Package Version*1.4.7. https://CRAN.R-project.org/package=emmeans

Lopez-Calderon, J., Luck, S.J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, **8**, 213.

Marciano, D., Bentin, S., Deouell, L.Y. (2018). Alternative outcomes create biased expectations regarding the received outcome: evidence from event-related potentials. *Neuropsychologia*, **113**, 126–39.

Meteyard, L., Davies, R.A. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language*, **112**, 104092.

Otten, L.J., Donchin, E. (2000). Relationship between P300 amplitude and subsequent recall for distinctive events: dependence on type of distinctiveness attribute. *Psychophysiology*, **37**(5), 644–61.

Otten, L.J., Rugg, M.D. (2001). Electrophysiological correlates of memory encoding are task-dependent. *Cognitive Brain Research*, **12**(1), 11–8.

Paller, K.A., Kutas, M., Mayes, A.R. (1987). Neural correlates of encoding in an incidental learning paradigm. *Electroencephalography and Clinical Neurophysiology*, **67**(4), 360–71.

Paller, K.A., Wagner, A.D. (2002). Observing the transformation of experience into memory. *Trends in Cognitive Sciences*, **6**(2), 93–102.

Payne, J.D., Kensinger, E.A. (2018). Stress, sleep, and the selective consolidation of emotional memories. *Current Opinion in Behavioral Sciences*, **19**, 36–43.

Pion-Tonachini, L., Kreutz-Delgado, K., Makeig, S. (2019). ICLabel: an automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, **198**, 181–97.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology*, **118**(10), 2128–48.

Polich, J. (2012). Neuropsychology of P300. *Oxford Handbook of Event-related Potential Components*, **159**, 159–88.

Proudfit, G.H. (2015). The reward positivity: from basic research on reward to a biomarker for depression. *Psychophysiology*, **52**(4), 449–59.

Rasch, B., and Born, J. (2013). About sleep's role in memory. *Physiological reviews*, **93**(2), 681–66.

Rigney, A.E., Schnyer, D.M., Hu, X., Beer, J.S. (2020). Mechanisms of a spotless self-image: navigating negative, self-relevant feedback. *Self and Identity*, 1–20.

Sambrook, T.D., Goslin, J. (2015). A neural reward prediction error revealed by a meta-analysis of ERPs using great grand averages. *Psychological Bulletin*, **141**(1), 213.

San Martín, R. (2012). Event-related potential studies of outcome processing and feedback-guided learning. *Frontiers in Human Neuroscience*, **6**, 304.

Schad, D.J., Vasishth, S., Hohenstein, S., Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: a tutorial. *Journal of Memory and Language*, **110**, 104038.

Scheier, M.F., Carver, C.S., Bridges, M.W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the Life Orientation Test. *Journal of Personality and Social Psychology*, **67**(6), 1063.

Schupp, H.T., Flaisch, T., Stockburger, J., Junghöfer, M. (2006). Emotion and attention: event-related brain potential studies. *Progress in Brain Research*, **156**, 31–51.

Sharot, T., Korn, C.W., Dolan, R.J. (2011). How unrealistic optimism is maintained in the face of reality. *Nature Neuroscience*, **14**(11), 1475.

Sharot, T., Garrett, N. (2016). Forming beliefs: why valence matters. *Trends in Cognitive Sciences*, **20**(1), 25–33.

Spielberger, C.D. (1983). Manual for the State-Trait Anxiety Inventory STAI (form Y) ("Self-Evaluation Questionnaire"). Palo Alto, CA: Consulting Psychologists.

Stickgold, R., Walker, M.P. (2013). Sleep-dependent memory triage: evolving generalization through selective processing. *Nature Neuroscience*, **16**(2), 139.

Talmi, D., Atkinson, R., and El-Deredy, W., (2013). The feedback-related negativity signals salience prediction errors, not reward prediction errors. *Journal of Neuroscience*, **33**(19), 8264–269.

Tambini, A., Davachi, L. (2019). Awake reactivation of prior experiences consolidates memories and biases cognition. *Trends in Cognitive Sciences*, **23**(10), 876–90.

RC Team. (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Available: https://www.r-project.org/.

Voeten, C.C. (2020). buildmer: Stepwise elimination and term reordering for mixed-effects regression. *R package version* 1.6. https://CRAN.R-project.org/package=buildmer

Walsh, M.M., Anderson, J.R. (2012). Learning from experience: event-related potential correlates of reward processing, neural adaptation, and behavioral choice. *Neuroscience and Biobehavioral Reviews*, **36**(8), 1870–84.

Weinstein, N.D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology*, **39**(5), 806–20.

Wu, Y., Zhou, X. (2009). The P300 and reward valence, magnitude, and expectancy in outcome evaluation. *Brain Research*, **1286**, 114–22.

Yeung, N., Sanfey, A.G. (2004). Independent coding of reward magnitude and valence in the human brain. *Journal of Neuroscience*, **24**(28), 6258–64.