Optimism Where There is None: Asymmetric Belief Updating Observed with Valence-Neutral Life Events

Jason W. Burton[1*], Adam J. L. Harris[2], Punit Shah[3], & Ulrike Hahn[1]

[1]Department of Psychological Sciences, Birkbeck, University of London, London, WC1E 7HX, UK
[2]Department of Experimental Psychology, University College London, London, WC1H 0AP, UK
[3]Department of Psychology, University of Bath, Bath, BA2 7AY, UK

* Corresponding author: Jason W. Burton (jburto03@mail.bbk.ac.uk; https://orcid.org/0000-0002-6797-2299)

Abstract

How people update their beliefs when faced with new information is integral to everyday life. A sizeable body of literature suggests that people's belief updating is optimistically biased, such that their beliefs are updated more in response to good news than bad news. However, recent research demonstrates that findings previously interpreted as evidence of optimistic belief updating may be the result of flaws in experimental design, rather than motivated reasoning. In light of this controversy, we conduct three pre-registered variations of the standard belief updating paradigm (combined N = 300) in which we test for asymmetric belief updating with *neutral*, non-valenced stimuli using analytic approaches found in previous research. We find evidence of seemingly biased belief updating with neutral stimuli — results that cannot be attributed to a motivational, valence-based, optimism account — and further show that there is uninterpretable variability across samples and analytic techniques. Jointly, these results serve to highlight the methodological flaws in current optimistic belief updating research.

Keywords: Belief updating, optimism, statistical artefact, rationality

Introduction

Over the past decade, research has argued that people are optimistically biased when updating their beliefs in light of new information, such that desirable information elicits greater updates than undesirable information. Given the practical implications of such motivational distortion, the phenomenon has attracted considerable cognitive and neuroscientific interest (Chowdhury et al., 2014; Garrett et al., 2014, 2018; Garrett & Sharot, 2017; Kappes et al., 2018; Kuzmanovic, Jefferson, et al., 2016; Kuzmanovic, Rigoux, et al., 2016; Kuzmanovic et al., 2018; Kuzmanovic & Rigoux, 2017; Marks & Baines, 2017; Moutsiana et al., 2013, 2015; Sharot et al., 2007; Sharot, 2011; Sharot et al., 2011; Sharot, Guitart-Masip, et al., 2012; Sharot, Kanai, et al., 2012), and seen application to high-stakes domains such as lay people's views on anthropogenic climate change (Sunstein et al., 2016). However, recent research has highlighted limitations of existing results and demonstrated that what appears to be optimistically asymmetric belief updating may in fact be attributable to a statistical artefact arising from the experimental design, rather than a self-enhancing motivational bias (Harris et al., 2013; Shah et al., 2016; but see Garrett & Sharot, 2017 for rebuttal). In this paper, we further investigate this "statistical artefact hypothesis" with three pre-registered experiments. All three demonstrate asymmetric belief updating in response to neutral information in our main analysis, and we further observe uninterpretable variability across samples and across various analytic techniques. Given there is no desirability in these trials, this asymmetry *cannot* be attributed to optimism.

The Update Method

Evidence for optimistic belief updating has primarily been obtained from "the update method" (Sharot et al., 2011). The simplest instantiation of the method proceeds as follows: Participants first estimate their chance of experiencing a negative life event (E1), then they are provided with the base rate statistic for experiencing that event (BR), and finally, they are asked to re-estimate their chance of experiencing said event (E2). For example, a participant, Sam, might be asked to estimate their chance of experiencing a negative life event, like getting divorced, to which they reply 5%. Sam is then presented with the BR of divorce (the actual proportion of the general population that gets divorced in their lifetime), which is 45%. Since the BR in this instance is greater than Sam's E1 for this negative event, this belief updating trial would be classed as one with undesirable information. Trials on which participants receive desirable information typically elicit greater updates than trials with undesirable information, which is interpreted as evidence of optimism in belief updating (e.g., Sharot et al., 2011).

The Statistical Artefact Hypothesis

While the update method has been accepted as the basis for a range of high-profile cognitive and neuroscientific work (e.g., Sharot et al., 2007, 2011; Sharot, Kanai, et al., 2012; Moutsiana et al., 2015; Garrett et al., 2018), it has also been subjected to critique. As the update method requires people to update a probabilistic belief, the appropriate normative standard against which to evaluate behavior so as to identify bias is Bayes' theorem (Hardman, 2009; Kahneman & Tversky, 1973; Phillips et al., 1966; Phillips & Edwards, 1966). Continuing our example, Sam is asked to report their risk of divorce. Normatively, this risk is comprised of two types of information: knowledge about the base rate of divorce and any individuating information Sam has to differentiate personal risk from the average person's. For instance, Sam might intend never to marry in the first place, which suggests that they should differentiate themself from the average person about marriage-related base rates (and this has long been understood to be the central challenge in assessing the accuracy of people's personal risk estimates for future life events, e.g., Weinstein, 1980). These two types of information — the base rate and individuating information — combine multiplicatively in Bayes' theorem, yet the update method neglects the influence of individuating information on normatively rational belief updating (c.f. Shah et al., 2016).

One way in which individuating information poses a difficulty for the update method is in the classification of trials with desirable versus undesirable information. For example, the primary component of Sam's belief may be the intention to never marry (whilst recognizing this may change later in life, and, as such, is greater than zero), but Sam estimates the base rate of divorce in the general population to be 50%, combining to a personal risk estimate of 5%. Upon learning the BR (45%) Sam should, in fact, slightly *decrease* the estimate of own, personal risk, not increase it as the classification of the desirability of BR (based on its comparison with E1) would suggest.

Even more worryingly, basic aspects of the probability scale — namely, that it is multiplicative and bounded — mean that over- and under-estimates relative to a given BR should not give rise to equal amounts of belief change once individuating knowledge comes into play. As Shah et al. (2016) demonstrate, it is mathematically impossible to equate the amount of BR error (the difference between one's estimate of the BR and the actual BR), the amount of individuating knowledge, and the normatively necessary degree of belief change across desirable and undesirable trials on the bounded, zero to 100 probability scale. Matching any two of these three will necessarily give rise to a divergence on the third (*SI, Fig. S6*).

As a result of both the issues above, entirely rational Bayesian agents will display seemingly optimistic belief updating (Shah et al., 2016). While Kuzmanovic and Rigoux (2017) have recently sought to remedy these issues (see *SI, Supplementary Analyses*, *1.3* for critique), the Shah et al. critique suggests that the vast majority of evidence for optimistic belief updating may be nothing more than a statistical artefact.

Why (Neutral) Valence Matters

Valence — whether information is good, bad, or neutral — is essential to the claims supporting the existence of optimistic belief updating. As Sharot and Garrett (2016) explain, optimistic belief updating is "a valence-dependent asymmetry in how people use favorable and unfavorable information" (p. 31). That is, optimistic belief updating is motivated by the perceived valence of encountered information. Therefore, establishing a causal relationship between asymmetries in belief updating (such as those recorded with the update method) and information valence is necessary to distinguish true, unrealistically optimistic belief updating from non-motivational, confounding effects.

Typically, belief updating studies use negative life events when asking participants for probability estimates with the update method (e.g., how likely is it that you will contract liver disease?). Negative life events are convenient because their perceived valence is considered to be universal, and one can readily gather accurate base rates from sources like the Office for National Statistics to use in experimental materials (e.g., Sharot et al., 2011). Given the statistical confounds associated with the update method, however, Shah et al. (2016) argued for the need to include both positive and negative life events in studies using the update method to evaluate directional, valence-driven effects (as in

Garrett & Sharot, 2017; Marks & Baines, 2017). The current study goes further by including neutral events as our focal stimuli. With negative events, evidence for optimistic updating has been claimed where participants update their beliefs more in trials where BR is less than E1 than where BR is greater than E1 (setting aside methodological concerns just outlined). If such an asymmetry can be observed with neutral events, this cannot be attributed to optimism, since there is no valence from which to ascribe (un)desirability. Thus, by parsimony, one should not appeal to a motivational explanation to explain such a pattern in situations where desirability is present.

In this paper, we present evidence for a non-motivational account of asymmetric belief updating by eliciting the statistical artefact with neutral life events, replicating it in three variations of the update method, and displaying its variability across samples and alternative analytic techniques (*SI, Supplementary Analyses 1.1-1.6*). As inherently non-valenced, an asymmetry in belief updating with neutral events suggests that what has been previously interpreted as evidence of optimistic belief updating is the result of statistical patterns rooted in methodological constraints, rather than a motivational bias. That this artefactual asymmetry varies so unpredictably further undercuts the interpretability of past and future results returned by the update method.

Method

Since studies utilizing the update method typically involve more than one stimulus (life event), there are variations in the way this experiment can be, and has been, run: Participants can provide E2s immediately after receiving the BR for each event (Kuzmanovic, Jefferson, et al., 2016; Marks & Baines, 2017), or they can provide all these at the end, after having provided E1s and received BRs for all events in the study (Garrett & Sharot, 2014, 2017). Such procedural changes have been made in the past without justification and it is unknown how they affect the results of the update task. We subsequently replicated our experiment in three ways for robustness (Fig. 1). While no predictions of different results between these studies were made, this project represents a direct replication of the same methodology using these three slightly different procedures. All data otherwise followed the same scoring plan.

Participants

One hundred participants were recruited for each study (N = 300). This sample size was deemed sufficient by a power analysis conducted with Judd et al.'s (2017) tool for calculating statistical power for mixed effects models. However, due to the quasi-experimental nature of this study, and in fact any study using the update method, the validity of such a power analysis is limited. Since we do not know how many life events participants will rate as neutral, we cannot know how many stimuli will be included in the main analysis, which focuses specifically on neutral life events. While we made a conscious effort to compile seemingly neutral life events, it is also plausible that some participant(s) will not rate any life events as neutral, and thereby be excluded from the main analysis. With these unknowns in mind, we arrived at the sample size of 100 because a hypothetical power analysis with 51 stimuli (the number of events used), an effect d of 0.5, and 100 participants returns a high power of 0.84. This is also roughly double the sample size used in Marks and Baines (2017), which is the only other optimistic belief updating study that uses the same main analysis (a linear mixed effects model).

The Prolific Academic online research platform was used for recruitment, and the participant pool location was restricted to the United Kingdom because the stimuli (the life events and base rate statistics) were compiled to suit this specific population, and certain events and base rates may not be relevant to participants living elsewhere. The sample for each study were independent of one another. In Study 1, participants' ages ranged from 18 to 81 (M = 32.82, SD = 11.52) with 73 females; in Study 2, ages ranged from 18 to 67 (M = 33.29, SD = 10.31) with 78 females; and in Study 3, ages ranged from 19 to 73 (M = 36.09, SD = 12.06) with 80 females.

Design

A 3 (event valence: negative, neutral, positive) x 2 (direction of error: upwards, downwards), quasi-experimental within-subjects design was implemented with the Qualtrics web-based survey software. Event valence (negative, neutral, or positive) signifies the self-reported desirability of experiencing a given life event. While the set of life events used (*SI, Table S1*) was intended to be comprised of neutrally-valenced events, event valence was coded trial-by-trial with participants' self-reports, meaning that an event had to have been rated as “neither positive nor negative” (three on a five-point Likert scale) to be classed as a neutral trial for a given participant. Direction of estimation error (henceforth referred to as “direction of error”) indicates whether a participant's initial self-estimate was less than or greater than the presented base rate on a given trial (whilst previous studies would refer to these trials as desirable vs. undesirable, such a categorization cannot be made for neutral events). Our dependent variable — the magnitude of belief updating — is the absolute difference between E1 and E2, signed as positive if the update is in the predicted direction (according to the direction of error), or negative if it is in the opposite direction. See the subsection on “Measures” below for further definition of these variables.

Stimuli

A total of 51 life events, each with an accompanying base rate statistic (M_{BR} = 38.39, SD_{BR} = 21.58), was presented to each participant (*SI, Table S1*). This set of life events is comprised of both previously used and novel material. Two life events were taken from Shah et al.'s (2016) materials that were deemed to be neutral at face value (“Be exactly the same weight in 10 years' time” and “Last the whole of winter without catching a minor cold”). Twenty-one life events were taken from the materials used by Garrett and Sharot (2017), which were found to be rated as neutral by pilot participants [3.00 ± 0.99 on a 1 (extremely negative) to 5 (extremely positive) scale]. Finally, 28 novel life events were compiled by gathering or calculating base rate estimates based on external sources (e.g., Ofcom, BBC) that were associated with seemingly regular, mundane life events [e.g., “How likely is it that the next store you visit is air conditioned?” (BR = 30%); “…that you use more than 3.7GB of mobile data over the next four weeks? (BR = 17%)].

Procedures

Study 1. Study 1 involved two parts (Fig. 1). In Part 1, each participant was presented with a life event and asked to estimate the likelihood of this event happening to them in the future (*E1*); participants were also asked to give a second estimate of the likelihood of the event happening to an average person (*eBR*; an estimate of the base rate). These two estimates were made on a 0% to 100% scale and there were no restrictions on participants' response time. The order of these two estimates was counterbalanced between subjects by randomly assigning each participant to one of two conditions: E1 followed by eBR, or eBR followed by E1. After these two initial estimates were recorded, participants were presented with the actual base rate statistic (*BR*). On the next page they were required to write down the BR correctly (*BR Recall*). If they incorrectly recalled the statistic, they were provided with the correct statistic on screen and required to enter that value. This ensured that the correct BR was attended to by all participants. Participants were then asked to re-estimate the likelihood of the event happening to them personally (*E2*). Once this sequence was completed for all 51 life events, in a randomized order, participants moved on to Part 2.

In Part 2, they were presented with each of the 51 life events again in a randomized order and asked to indicate the valence of each event, "How would you feel about experiencing this event?" on a five-point Likert scale (1 = extremely negative, 2 = somewhat negative, 3 = neither positive nor negative, 4 = somewhat positive, 5 = extremely positive) (*Valence Rating*). This rating would later be used to classify the trials as negatively, neutrally, or positively valenced according to each individual participant's subjective rating. Trials rated as 3 were considered neutral.

Study 2. Study 2 involved three parts (Fig. 1). The procedure followed that of Study 1, except the E2s for all events were recorded in Part 2, which was followed by the recording of all valence ratings in Part 3.

Study 3. Study 3 involved two parts (Fig. 1) and mirrored the procedure of Study 1, except the ordering of E2 and Valence Rating were swapped such that the updating occurred across the two parts (i.e., E1 in part one, E2 in Part 2). This procedure follows that used in Experiments 3A and 3B in Shah et al. (2016).

Exclusion Criteria

Following on from Shah et al. (2016), a pre-registered exclusion criterion was employed prior to analyzing the data from each study. Mean updates in each of the six conditions (i.e., negative/neutral/positive by upwards/downwards) were calculated and outliers were removed (±3 × the interquartile range)[1]. However, all of our results hold regardless of this criterion, except for a single supplementary result (noted below in the "Further Analyses" subsection). In addition, trials where E1 equalled the BR were necessarily excluded in our analyses, as they are in all analyses of the data from the update method, because the central, quasi-experimental classification of trials as desirable or undesirable — or in our case, upwards or downwards — cannot be applied. From the 5,100 trials recorded in each study, these criteria resulted in 339 (6.65%), 414 (8.12%), and 431 (8.45%) trials being excluded from studies 1, 2, and 3, respectively.

Measures

There are three measures that are central to interpreting results from the update method: direction of error, update value, and event valence. First, each trial's direction of error is determined by calculating the difference between BR and E1. If the difference is

[1] We note that on page 7 of the pre-registration this exclusion criterion is written as "±3 the interquartile range." This discrepancy is due to a clerical error whereby the multiplication sign was inadvertently removed when pasting the text into the OSF registration form and converting the Word file to PDF format. In addition, the mention of removing trials "in which a derived probability cannot be applied" is also erroneous as it is not applicable to the methodology presently used because we provided static, non-derived base rates.

positive (i.e., BR > E1) then the direction of error is classed as upwards, and vice versa for a downwards direction of error.

The update value in each trial, which indicates the magnitude of belief updating, is then determined by calculating the absolute difference between participants' re-estimates and initial estimates (|E2 - E1|). This is then labelled as positive if updating goes towards the

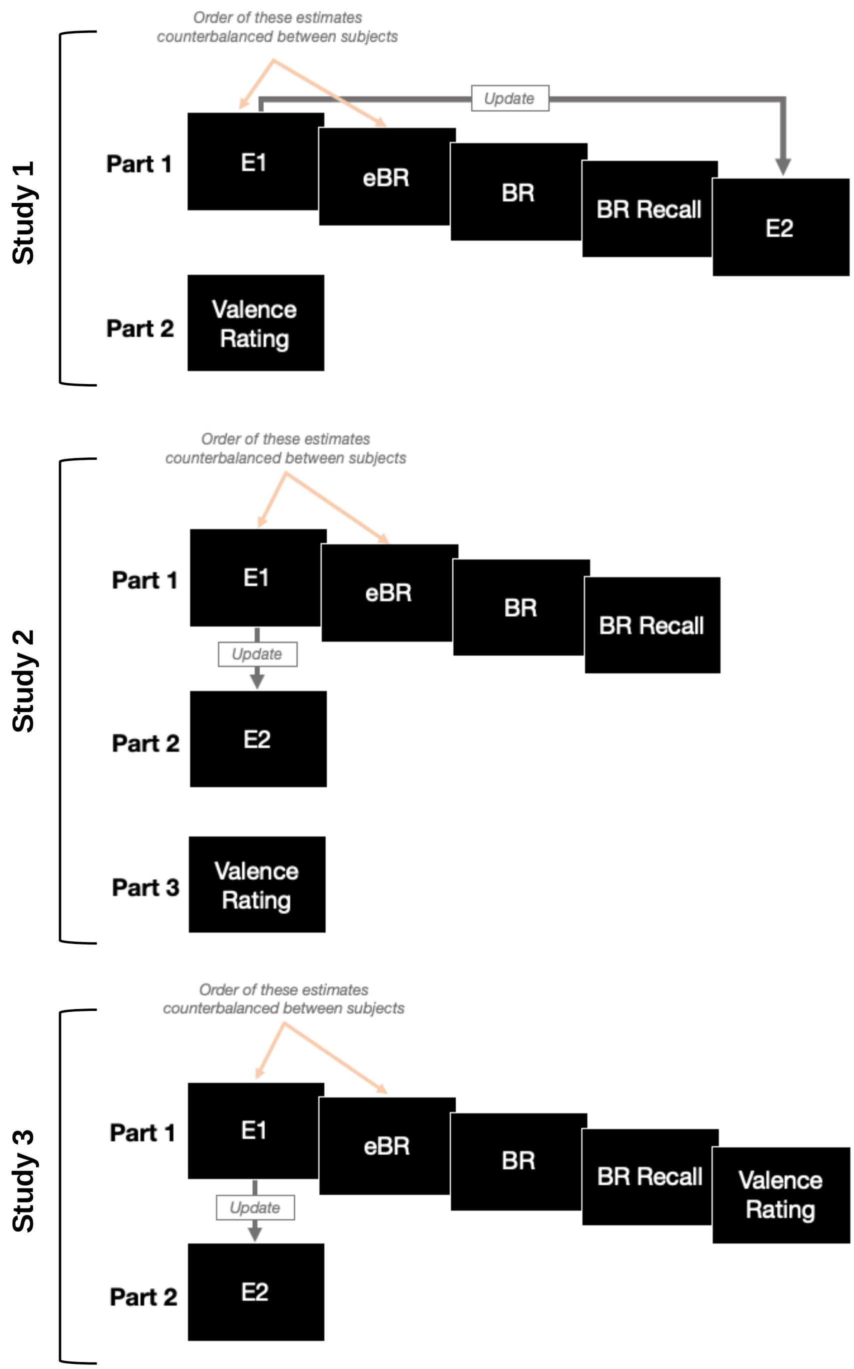


Fig. 1. Schematics of the procedure used in each study. Images depict the task sequence for a single trial, differing slightly across studies: participants make an initial self-estimate (E1) and an estimate of the base rate (eBR), view the base rate (BR), write down the base rate correctly (BR

Recall), rate the event's valence on a five-point Likert scale where a rating of 3 indicates the event is perceived as neutral (Valence Rating), and make a revised self-estimate (E2). The same stimuli were used in each study. We have followed the abbreviations of Kuzmanovic and Rigoux (2017). To aid cross-referencing, in Shah et al. (2016) E1 and E2 are SE1 and SE2, eBR is BR1, BR is actualBR.

BR (i.e., in accordance with the direction of error) or negative if updating goes away from the BR (i.e., *not* in accordance with the direction of error). For example, if E1 is 40, BR is 50, and E2 is 30, the update would be coded as -10; whereas if the BR was 10 then the update would be coded as +10. Additional pre-registered analyses in which direction of error is calculated in the normatively appropriate manner on the basis of participants' estimates of the BR (*eBR*), are presented in the supplementary materials (*SI, Supplementary Analyses, 1.1*).

Finally, event valence is determined trial-by-trial with the self-report question, "How would you feel about experiencing this event?", measured on a five-point Likert scale from "extremely negative" (1) to "neither positive nor negative" (3) to "extremely positive" (5). Following Garrett and Sharot (2017), we code ratings of 1 and 2 as negative, 3 as neutral, and 4 and 5 as positive. While our set of life events was compiled with neutral valence in mind, this procedure grants each participant the opportunity to provide updates for negative, positive, and neutral events, depending on their personal preferences.

Results

Main Analysis

To test our central hypothesis that participants will update asymmetrically in response to neutral information, we conducted three studies with independent samples in which 100 participants proceeded through the update method with 51 life events. Following our pre-registered analysis plan, we applied a linear mixed effects model (LMM) — as in Marks and Baines (2017) — to trials in which the stimulus (life event) was rated as neutral by the participant. Update value was entered as the dependent variable, direction of error (upwards/downwards) as a fixed factor, and participant as a random factor.

To select a model specification, we first fit the specification with the maximally complex random effects structure and then iteratively reduced model complexity until all degenerate random effects parameters were removed and the model was not singular (Bates et al., 2018). This procedure led us to a specification with only random intercepts by participant; however, the results hold across model specifications and we report the statistics of the maximally complex model specification in the Supplementary Information (*SI, Table S2*). Finally, we used Type III tests and Satterthwaite's approximation for degrees of freedom to calculate the statistical significance of the fixed effects. In all three studies, asymmetric belief updating was observed with neutral life events (Fig. 2).

In Study 1, there were 1,521 trials in which participants updated in response to a rated-as-neutral life event, with 801 trials with an upwards direction of error (M = 2.11, SD = 4.56) and 720 with a downwards direction of error (M = 11.33, SD = 15.91). An LMM determined that direction of error significantly affected the magnitude of participants' updating ($F(1,1515) = 244.47$, $p < 0.001$), such that an upwards direction of error (i.e., BR > E1) decreased update scores by approximately 9.13 percentage points (fixed effect estimate) ± 0.58 (standard error), as compared to downwards direction of error.

In Study 2, there were 1,699 trials in which participants updated in response to a rated-as-neutral life event, with 831 trials with an upwards direction of error (*M* = 4.16, *SD* = 10.01) and 868 with a downwards direction of error (*M* = 10.08, *SD* = 18.06). An LMM determined that direction of error significantly affected the magnitude of participants'

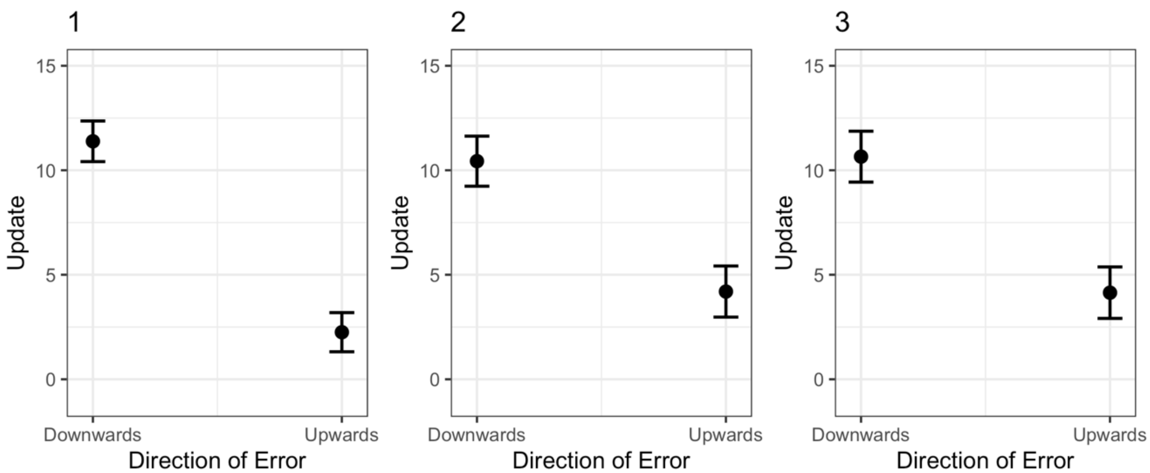


Fig. 2. Asymmetries in belief updating with neutral life events in each of the three studies. Points indicate the magnitude of belief updating predicted by the linear mixed effects model with bars representing 95% confidence intervals. Numbering of plots corresponds to the study.

updating ($F(1,1694) = 77.09$, $p < 0.001$), such that an upwards direction of error (i.e., BR > E1) decreased update scores by about 6.24 percentage points (fixed effect estimate) ± 0.71 (standard error), as compared to downwards direction of error.

In Study 3, there were 1,667 trials in which participants updated in response to a rated-as-neutral life event, with 828 trials with an upwards direction of error (*M* = 4.13, *SD* = 8.82) and 839 with a downwards direction of error (*M* = 10.56, *SD* = 20.80). An LMM determined that direction of error significantly affected the magnitude of participants' updating ($F(1,1662) = 68.48$, $p < 0.001$), such that an upwards direction of error (i.e., BR > E1) decreased update scores by about 6.51 percentage points (fixed effect estimate) ± 0.79 (standard error) as compared to downwards direction of error.

Secondary Analyses

Adding event valence as a second fixed factor. In addition to our central hypothesis testing, we also conducted a pre-registered analysis of each study's data with LMMs that included event valence as a second fixed factor in interaction with direction of error. This analysis parallels that of previous related work (Garrett & Sharot, 2017) and allows us to assess whether the supposed desirability of information influenced belief updating, as well as view that effect in comparison with the asymmetry in neutral trials described above (Fig. 3). To select a model specification, we followed the same procedure as in the main analysis, which led us to select a specification that includes only random slopes and intercepts by participant for direction of error and no correlation parameters. However, the results once again hold across model specifications, and we report the statistics of the maximally complex model in the supplementary information (*SI, Table S3*).[2]

In Study 1, there were significant main effects of both direction of error ($F(1,98) = 233.80$, $p < 0.001$) and event valence ($F(2,4713) = 47.98$, $p < 0.001$), plus a significant interaction term ($F(2,4706) = 61.80$, $p < 0.001$), whereby the updating asymmetry was smallest for the positive events (Fig. 2). Despite this, participants updated more in response to a downwards direction of error across both negative and positive events. In 1,662 trials with negative life events, participants updated more in response to a downwards direction of error ($n = 683$, $M = 10.73$, $SD = 14.23$) than upwards ($n = 979$, $M = 2.86$, $SD = 6.31$). Equally, in the 1,578 trials with positive life events, participants updated more in response to a downwards direction of error ($n = 882$, $M = 4.97$, $SD = 9.41$) than upwards ($n = 696$, $M = 3.10$, $SD = 6.65$).

In Study 2, there were again significant main effects of direction of error ($F(1,98) = 112.85$, $p < 0.001$) and event valence ($F(2,4656) = 35.49$, $p < 0.001$), as well as a significant interaction term ($F(2,4610) = 43.86$, $p < 0.001$). Among the 1,582 trials with negative events, participants updated more in response to a downwards direction of error ($n = 692$, $M = 13.08$, $SD = 21.33$) than upwards ($n = 890$, $M = 3.76$, $SD = 10.18$), in a similar fashion as Study 1. However, update values in the 1,405 trials with positive events displayed no asymmetry between downwards ($n = 835$, $M = 4.05$, $SD = 10.98$) and upwards direction of error ($n = 570$, $M = 4.48$, $SD = 10.62$).

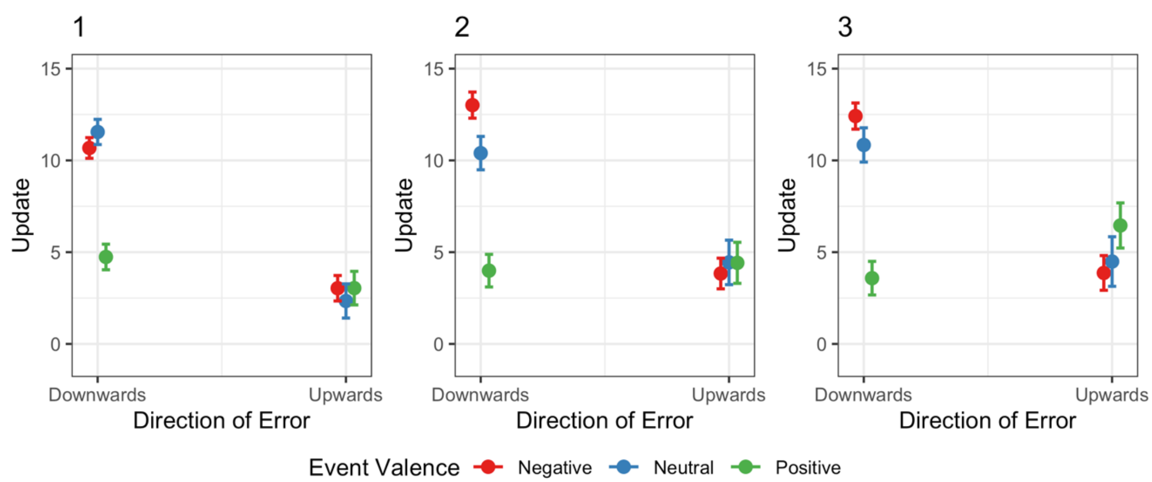


Fig. 3. Asymmetries in belief updating with event valence and direction of error as fixed factors. Points indicate the magnitude of belief updating predicted by the linear mixed effects model with bars representing 95% confidence intervals. Numbering of plots corresponds to the study.

[2] Because this secondary analysis is conditioned on a post-treatment variable (i.e., participants rated event valence after being provided with the BR), the results of this analysis may display a post-treatment bias (Montgomery et al., 2018). However, this potential issue also applies to past studies using the update method and life events of varying valence (e.g., Garrett & Sharot, 2014, 2017).

In Study 3, there were once again significant main effects of direction of error ($F(1,98) = 42.35$, $p < 0.001$) and event valence ($F(2,4636) = 18.51$, $p < 0.001$), as well as a significant interaction term ($F(2,4586) = 60.09$, $p < 0.001$). Interestingly, a flip in belief updating asymmetry was observed, which has previously been considered to be characteristic of optimistic belief updating. In 1,616 trials with negative life events, participants updated more in response to a downwards direction of error ($n = 639$, $M = 12.41$, $SD = 22.17$) than upwards ($n = 977$, $M = 3.62$, $SD = 9.51$); whereas participants updated *less* in response to a downwards direction of error ($n = 723$, $M = 3.45$, $SD = 10.42$) than upwards ($n = 663$, $M = 6.57$, $SD = 14.16$) in the 1,386 trials with positive life events.

Further analyses. Beyond the reported LMMs, there are a number of other analyses that have been applied to data produced by the update method. Relating to the critiques outlined here (i.e., the problems of misclassification and the bounded probability scale), Shah et al. (2016) analyzed the data with an alternative classification scheme and compare participants' updating with rational Bayesian predictions, and Kuzmanovic and Rigoux (2017) sought to incorporate in the influence of individuating information with a computational modelling technique based on reinforcement learning. We report the full results of each of these analytic approaches (all of which were pre-registered) in the *SI, Supplementary Analyses 1.1-1.3.* This shows that, despite their different attempts to remedy the limitations of the update method (attempts which have been critiqued – Harris et al., 2021; Shah et al, 2016), they each demonstrate the 'unmotivated' updating asymmetry with neutral events when the data from Studies 1-3 are aggregated. Whilst Analysis 1.1 yields such a result across all three studies, analyses 1.2 and 1.3 display unexplainable variability across the studies in terms of the significance of this result (Table 1).

Finally, we conducted an unregistered analysis in which we applied the procedure used in the original work of Sharot et al. (2011)[3]. This analysis uses regressions to assess correlations between "estimation error" (i.e., |E1 – BR|) and belief updating (*SI, Supplementary Analysis 1.4*). While it is meaningful to control for initial error (to ensure that observed asymmetries do not merely reflect an uneven distribution of errors), we did not preregister this particular analysis as it is conceptually flawed. As outlined in the Introduction above it is meaningless to subtract Sam's base rate estimate (BR) from his individual estimate (E1), because *normatively*, Sam's individual estimate should be a function of both of these quantities, and, *normatively*, Sam could even be required to revise the individual estimate in the opposite direction of that 'error', depending on what Sam took the base rate to be (see also, Shah et al., 2016). All of the (pre-registered) further analyses described in the previous paragraph (and reported in *Supplementary Analyses 1.1-1.3)* were proposed in the previous literature as conceptually meaningful alternative solutions for controlling for 'estimation error'.

The flawed analysis returns a statistically significant asymmetry with neutral events in Study 1, but there is neither a significant asymmetry in Studies 2 and 3, nor in the aggregated data (Table 1). Further, the statistically significant asymmetry in Study 1 does not hold when our exclusion criterion is not applied. To test the robustness of this result — and to explore why there is considerable, seemingly uninterpretable variability in results across studies — we ran simulations where we sampled subsets of events and participants from the aggregated data of Studies 1-3 and applied the regression analysis

[3] We included this analysis in response to reviewer comments.

to each sample. These simulations showed that materially different results can be obtained with the regression analysis depending on which events are used (*Supplementary Figure S7*). Building on this, we subsequently pre-registered and ran an additional Study 4 using 20 of the 51 events from Studies 1-3, which further demonstrates that consistent belief updating asymmetries, in line with optimism bias, are not observed (see *SI, Supplementary Study 4* for a write-up of these results).

Other analytic approaches may also be applicable, such as including stimuli (life events) as a random factor to heed the statistical literature pointing out that the failure to do so can inflate Type I error rates on fixed effect estimates (Judd et al., 2012; Yarkoni, 2020). However, this does not resolve the asymmetry observed with the LMMs (*SI, Supplementary Analyses, 1.6*). Further, one might argue that Bayesian modelling equivalents to the LMMs reported here should be used. But such an analysis introduces an additional “researcher degree of freedom” by requiring the specification of a prior distribution; and given that there is no precedent of this in the literature, we restricted our analyses to address our hypothesis. Nevertheless, all data has been made openly available on an OSF project page for further analysis.

Table 1: P values indicating the statistical significance of the asymmetries observed with neutral events according to four analytical techniques intended to resolve the flaws of the update method in each study and the aggregated data of Studies 1-3. Comparisons of upwards vs. downwards updating using the Bayesian difference measure and the regression analysis (with “estimation error” as the independent variable) were made with t-tests; comparisons using the Bayesian ratio measure and learning rates were made with non-parametric Wilcoxon signed rank tests. The analysis marked with `*` was not pre-registered. Full results for each analysis are reported in *SI, Supplementary Analyses 1.2-1.4*.

	Study 1	Study 2	Study 3	Aggregate
Bayesian difference measure (Shah et al., 2016)	0.049	0.044	0.013	< 0.001
Bayesian ratio measure (Shah et al., 2016)	< 0.001	0.784	0.449	0.010
Learning rates (Kuzmanovic & Rigoux, 2016)	0.001	0.704	0.324	0.016
Regression analysis* (Sharot et al., 2011)	< 0.001	0.662	0.540	0.081

Discussion

The present study provided the first targeted application of the update method to test for asymmetric belief updating with neutral life events. In the main analysis, the central result obtained was the thrice replicated asymmetry in belief updating with neutral stimuli (Fig. 2), which is coupled with uninterpretable variability among the results returned by various analytic techniques that have been proposed as fixes for the flaws of the update methods (Table 1). Such findings are unpredicted and unexplainable by motivational accounts. Consequently, our results contribute to the debate concerning the status of the optimistic belief updating phenomenon (Sharot et al., 2011) by demonstrating the empirical consequence of the statistical confounds previously highlighted (Shah et al., 2016) — empirical consequences that subsequent rebuttals have questioned (e.g., Garrett & Sharot, 2017; Kuzmanovic & Rigoux, 2017). While seemingly optimistic asymmetries can arise in the data, the current results suggest that such effects are not valence-driven. Given that an asymmetry can be observed in the absence of valence, and thus an absence of motivation, the presence of an asymmetry with valenced events is insufficient evidence for one to conclude that a motivational bias is present.

Previous research has countered the statistical artefact hypothesis by including positive events in the deployment of the update method (e.g., Garrett & Sharot, 2017; Marks & Baines, 2017). While such studies succeed in demonstrating asymmetric belief updating with positive events in a manner that is consistent with motivational optimism, they fail to appreciate “that the statistical artefacts will necessarily vary in expression as a function of events” (Shah et al., 2016, p. 106). Consequently, “the very nature of the artefacts that plague the update method mean that, given the right set of events [with the right statistical attributes], everything and anything could empirically be found, even in entirely unbiased agents” (Shah et al., 2016, p. 107). This is because the statistical artefact hypothesis is multi-faceted, being driven by both the base rate and the amount of individuating information the individual believes they possess about their chance of experiencing a particular event[4]. In fact, the variability of these subjective confounds can be seen even when an identical set of life events is used. Across our studies, there are differences in the distributions of implied likelihood ratios (a measure of participants’ individuating information derived from Bayes Theorem), estimation error (|E1 – BR|), and base rate error (|eBR – BR|) (see *SI, Figs. S3-S5* for the empirical distributions of these variables in each study; also see *SI, Fig. S6* for a numerical simulation of the relationship between these variables). It may be these discrepant statistical characteristics that give rise to varying results across studies.

One tempting approach to side-step the variability in results across studies and analyses is to focus solely on the aggregated data from Studies 1-3 as the single “answer.” This approach is attractive as it achieves the highest possible statistical power. Equally, however, this strategy masks the strikingly variable results across the individual studies, which were already adequately powered with 100 participants each. This point is especially striking in view of the fact that the update method has most widely been used in cognitive neuroscience studies that contained far fewer participants. The original work of Sharot et al. (2011) relied on a sample of just 19 participants, Garrett et al. (2014) had 30 participants, Garrett and Sharot (2014) had 32 participants, Kuzmanovic and Rigoux (2017) had 27 participants, and even the large (by neuroscience standards) sample of Moutsiana et al. (2013) was just 52 participants.

Our findings suggest that there is not a consistent, behaviourally-large, motivational asymmetry in belief updating, as would be required to inspire confidence in the neural correlates – and therefore the existence – of the optimistic bias phenomenon. While variability with “sensible” stimuli (i.e., negative or positive events) is often attributed to natural noise in the world, we observed similar variability with entirely nonsensical stimuli

[4] Note that the same sized bias for positive events as for negative events (greater updating in a downwards direction) would be predicted in unbiased agents in the very unlikely situation where positive stimuli are the exact analogues of the negative stimuli, with the same base rates, but also being “exactly matched for diagnostic knowledge” (Shah et al., 2016, p. 106).

(i.e., neutral events). The dominant theory for a general optimism bias is that it is an adaptive, self-serving mechanism that enhances exploratory behavior and reduces stress and anxiety as a regular feature of healthy human cognition (Sharot, 2011). Further, it has been argued that the ability to integrate desirable and undesirable information reflects two dissociated processes with different developmental trajectories in the human brain (Chowdhury et al., 2014; Moutsiana et al., 2013). On neither perspective would one expect the findings presented here. Were the update method a suitable tool for probing optimistic bias, it simply should not show “bias” with valence-neutral events. The fact that it can, and that results vary across samples and analyses, renders both the behavioral results of earlier studies and the underlying neurological correlates, based on small samples, uninterpretable.

Whilst our investigation included small changes in methodology across studies and covered several supplementary analyses, readers familiar with the optimistic belief updating literature might point out that we did not control for certain covariates that have been considered in past studies, such as participants’ familiarity, vividness, emotional arousal, and perceived controllability of the events presented. While this could indeed be viewed as a limitation of the present work, we do not see it as such because past studies consistently report no change in results with or without these controls (e.g., Sharot et al., 2011; Moutsiana et al., 2013; Shah et al. 2016), and moreover, certain covariates like emotional arousal are, by definition, irrelevant for neutral life events. Likewise, it could be argued that our results are limited due to the fact that we only considered one framing; we only asked participants how likely they were to experience the events presented and not how likely they were to *not* experience the events. However, the potential of a confounding framing effect also seems unlikely given that the past studies that do control for it report there to be no consequence (e.g., Garrett et al., 2014; Garrett & Sharot, 2014; Korn et al., 2014; Sharot et al., 2011).

Other methods for optimism research need to be used, but there is no quick fix. One plausible direction is to test for bias as asymmetric deviations from Bayes’ theorem, which acknowledges that updates in different parts of the probability scale cannot be mathematically equated to one another. Yet, Shah et al. (2016) demonstrate that the fundamental problems with the scale are inherited in such analyses given that participants’ responses will be noisy (see *SI, Supplementary Analyses, 1.2*). Alternatively, some investigators have traded off external validity and utilised “lottery” methods incorporating objective probabilities (Eil & Rao, 2011; Mobius et al., 2014; Barron, 2019; Buser et al., 2018; Coutts, 2019; Ertac, 2011; Gotthard-Real, 2017). These studies have produced markedly inconsistent results, with some observing an optimistic asymmetry (Eil & Rao, 2011; Mobius et al., 2014), and others finding no evidence thereof (Barron, 2019; Buser et al., 2018; Coutts, 2019; Ertac, 2011; Gotthard-Real, 2017).

Conclusion

Optimism research is of importance to both researchers and practitioners across disciplines. Yet, for there to indeed be a true optimism bias in belief updating, the evidence should not be able to be produced by rational, unbiased agents (e.g., Shah et al., 2016) or in cases where the variable upon which it depends — valence — is removed. The update method thus remains unfit for purpose, and assuming evidence produced by it to be solid is ill-advised. Optimistic belief updating, along with optimism’s other forms, may very well exist in the real-world. At present, however, the foundational method that research is building upon continues to fail critical tests. Here, it has failed to display a consistent valence-dependence, an inherent attribute of optimistic belief updating's very definition.

Data and Code Availability

The datasets and code used in this paper are available on the OSF project page.

Ethics Statement

This research was approved by the Department of Psychological Sciences Research Ethics Committee of Birkbeck, University of London (reference 161755). All participants provided their informed consent.

Declaration of Interest

The authors declare no competing interest.

References

Barron, K. (2019). *Belief updating: Does the "good-news, bad-news" asymmetry extend to purely financial domains?* (SP II 2016-309r; WZB Discussion Paper,).

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2018). Parsimonious Mixed Models. *ArXiv:1506.04967 [Stat]*. http://arxiv.org/abs/1506.04967

Buser, T., Gerhards, L., & van der Weele, J. (2018). Responsiveness to feedback as a personal trait. *Journal of Risk and Uncertainty*, *56*(2), 165–192. https://doi.org/10.1007/s11166-018-9277-3

Chowdhury, R., Sharot, T., Wolfe, T., Düzel, E., & Dolan, R. J. (2014). Optimistic update bias increases in older age. *Psychological Medicine*, *44*(9), 2003–2012. https://doi.org/10.1017/S0033291713002602

Coutts, A. (2019). Good news and bad news are still news: Experimental evidence on belief updating. *Experimental Economics*, *22*(2), 369–395. https://doi.org/10.1007/s10683-018-9572-5

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), 114–138. https://doi.org/10.1257/mic.3.2.114

Ertac, S. (2011). Does self-relevance affect information processing? Experimental evidence on the response to performance and non-performance feedback. *Journal of Economic Behavior and Organization*, *80*(3), 532–545. https://doi.org/10.1016/j.jebo.2011.05.012

Garrett, N., González-Garzón, A. M., Foulkes, L., Levita, L., & Sharot, T. (2018). Updating beliefs under perceived threat. *Journal of Neuroscience*, *38*(36), 7901–7911. https://doi.org/10.1523/JNEUROSCI.0716-18.2018

Garrett, N., & Sharot, T. (2014). How robust is the optimistic update bias for estimating self-risk and population base rates? *PLoS ONE*, *9*(6), 1–8. https://doi.org/10.1371/journal.pone.0098848

Garrett, N., & Sharot, T. (2017a). Optimistic update bias holds firm: Three tests of robustness following Shah et al. *Consciousness and Cognition*, *50*, 12–22. https://doi.org/10.1016/j.concog.2016.10.013

Garrett, N., & Sharot, T. (2017b). Optimistic update bias holds firm: Three tests of robustness following Shah et al. *Consciousness and Cognition*, *50*, 12–22. https://doi.org/10.1016/j.concog.2016.10.013

Garrett, N., Sharot, T., Faulkner, P., Korn, C. W., Rosier, J. P., & Dolan, R. J. (2014). Losing the rose tinted glasses: Neural substrates of unbiased belief updating in depression. *Frontiers in Human Neuroscience*, *8*(August), 1–9. https://doi.org/10.3389/fnhum.2014.00639

Gotthard-Real, A. (2017). Desirability and information processing: An experimental study. *Economics Letters*, *152*, 96–99. https://doi.org/10.1016/j.econlet.2017.01.012

Hardman, D. (2009). *Judgment and decision making: Psychological perspectives (Vol. 11)*. John Wiley & Sons.

Harris, A. J. L., Shah, P., Catmur, C., Bird, G., & Hahn, U. (2013). Autism, optimism and positive events: Evidence against a general optimistic bias. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *35*(555–560).

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54–69. https://doi.org/10.1037/a0028347

Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with More Than One Random Factor: Designs, Analytic Models, and Statistical Power. *Annual Review of Psychology*, *68*(1), 601–625. https://doi.org/10.1146/annurev-psych-122414-033702

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

Kappes, A., Faber, N. S., Kahane, G., Savulescu, J., & Crockett, M. J. (2018). Concern for Others Leads to Vicarious Optimism. *Psychological Science*, *29*(3), 379–389. https://doi.org/10.1177/0956797617737129

Korn, C. W., Sharot, T., Walter, H., Heekeren, H. R., & Dolan, R. J. (2014). Depression is related to an absence of optimistically biased belief updating about future life events. *Psychological Medicine*, *44*(3), 579–592. https://doi.org/10.1017/S0033291713001074

Kuzmanovic, B., Jefferson, A., & Vogeley, K. (2016). NeuroImage The role of the neural reward circuitry in self-referential optimistic belief updates. *NeuroImage*, *133*, 151–162. https://doi.org/10.1016/j.neuroimage.2016.02.014

Kuzmanovic, B., & Rigoux, L. (2017). Valence-dependent belief updating: Computational validation. *Frontiers in Psychology*, *8*(JUN), 1–11. https://doi.org/10.3389/fpsyg.2017.01087

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

Kuzmanovic, B., Rigoux, L., & Vogeley, K. (2016). Reduced optimism bias in self-referential belief updating in high-functioning autism. *Journal of Autism and Developmental Disorders*, 1–9. https://doi.org/doi:10.1007/s10803-016-2940-0

Marks, J., & Baines, S. (2017). Optimistic belief updating despite inclusion of positive events. *Learning and Motivation*, *58*(May), 88–101. https://doi.org/10.1016/j.lmot.2017.05.001

Mobius, M. M., Niederle, M., Niehaus, P., & Rosenblat, T. S. (2014). *Managing Self-Confidence*.

Montgomery, J. M., Nyhan, B., & Torres, M. (2018). How Conditioning on Post-treatment Variables Can Ruin Your Experiment and What to Do about It [Data set]. In *American Journal of Political Science* (Vol. 62, Issue 3, pp. 760–775). https://doi.org/10.7910/DVN/EZSJ1S

Moutsiana, C., Charpentier, C. J., Garrett, N., Cohen, M. X., & Sharot, T. (2015). Human frontal–subcortical circuit and asymmetric belief updating. *Journal of Neuroscience*, *35*(42), 14077–14085. https://doi.org/10.1523/JNEUROSCI.1120-15.2015

Moutsiana, C., Garrett, N., Clarke, R. C., Lotto, R. B., Blakemore, S. J., & Sharot, T. (2013). Human development of the ability to learn from bad news. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(41), 16396–16401. https://doi.org/10.1073/pnas.1305631110

Phillips, L. D., & Edwards, W. (1966). Conservatism in a simple probability inference task. *Journal of Experimental Psychology*, *72*(3), 346–354.

Phillips, L. D., Hays, W. L., & Edwards, E. (1966). Conservatism in complex probabilistic inference. *IEEE Transactions on Human Factors in Electronics*, *1*, 7–18.

Shah, P., Harris, A. J. L., Bird, G., Catmur, C., & Hahn, U. (2016). A pessimistic view of optimistic belief updating. *Cognitive Psychology*, *90*, 71–127. https://doi.org/10.1016/j.cogpsych.2016.05.004

Sharot, T. (2011). The optimism bias. *Current Biology*, *21*(23), R941–R945. https://doi.org/10.1016/j.cub.2011.10.030

Sharot, T., & Garrett, N. (2016). Forming Beliefs: Why Valence Matters. *Trends in Cognitive Sciences*, *20*(1), 25–33. https://doi.org/10.1016/j.tics.2015.11.002

Sharot, T., Guitart-Masip, M., Korn, C. W., Chowdhury, R., & Dolan, R. J. (2012). How dopamine enhances an optimism bias in humans. *Current Biology*, *22*(16), 1477–1481. https://doi.org/10.1016/j.cub.2012.05.053

Sharot, T., Kanai, R., Marston, D., Korn, C. W., Rees, G., & Dolan, R. J. (2012). Selectively altering belief formation in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(42), 17058–17062. https://doi.org/10.1073/pnas.1205828109

Sharot, T., Korn, C. W., & Dolan, R. J. (2011). How unrealistic optimism is maintained in the face of reality. *Nature Neuroscience*, *14*(11), 1475–1479. https://doi.org/10.1038/nn.2949

Sharot, T., Riccardi, A. M., Raio, C. M., & Phelps, E. A. (2007). Neural mechanisms mediating optimism bias. *Nature*, *450*(7166), 102–105. https://doi.org/10.1038/nature06280

Sunstein, C. R., Bobadilla-Suarez, S., Lazzaro, S. C., & Sharot, T. (2016). How people update beliefs about climate change: Good news and bad news. *Cornell Law Review*, *102*, 1431.

Weinstein, D. (1980). Unrealistic Optimism About Future Life Events. *Journal of Personality and Social Psychology*, *39*(5), 806. https://doi.org/10.1037/0022-3514.39.5.806

Yarkoni, T. (2020). The generalizability crisis. *Behavioral and Brain Sciences*, 1–37. https://doi.org/10.1017/S0140525X20001685

Supplementary Information for

Optimism Where There is None: Asymmetric Belief Updating Observed with Valence-Neutral Life Events

Jason W. Burton[1*], Adam J. L. Harris[2], Punit Shah[3], & Ulrike Hahn[1]

[1]Department of Psychological Sciences, Birkbeck, University of London, London, WC1E 7HX, UK
[2]Department of Experimental Psychology, University College London, London, WC1H 0AP, UK
[3]Department of Psychology, University of Bath, Bath, BA2 7AY, UK

* Corresponding author: Jason W. Burton (jburto03@mail.bbk.ac.uk; https://orcid.org/0000-0002-6797-2299)

This PDF file includes:

✓ Indicates a pre-registered analysis
X Indicates an unregistered analysis

1. Supplementary Analyses

1.1 Accounting for direction of error misclassification. As described in the main text, a central limitation of the standard update method is its neglect of individuating information. Participants may hold one estimate of their personal likelihood of experiencing each event (E1), which is influenced by individuating information, and another estimate of the likelihood of an average person experiencing each event (an estimate of the base rate — eBR), which is not influenced by individuating information[1]. This means that by classifying direction of error (or, in the case of previous studies of optimistic belief updating, the desirability) in each trial on the basis of E1 instead of eBR, trials can be misclassified and subsequently muddle the results. To assess the empirical consequence of misclassification, we re-analyzed the data with an alternative direction of error assigned by comparing eBR to BR in each trial (i.e., eBR > BR is downwards, and vice versa). Across all of the data collected, 25.30% (n = 3,571) of trials were misclassified. In the following paragraphs we report the reduction in the fixed effect estimates produced by LMMs that exclusively analyzed neutral trials (as in the main analysis), as well as the altered effects produced by LMMs that included event valence as a second fixed factor[5]. In each instance we followed the procedure for the LMMs reported in the main text, whereby we first fit the model specification with the maximally complex random effects structure and then iteratively reduce the random effects structure until all degenerate random effects parameters are removed and the model is not singular. For the LMMs with only neutral trials, this left us with only random intercepts by participant; and for the LMMs with event valence as a second fixed factor this left us with random slopes and intercepts for direction of error by participant, and no correlation parameters. While the maximally complex random effects specifications were singular, we also report the results of these specification in Tables S4-S5 for comparison. It should be noted that while this analysis alleviates the issue of misclassification, it should be noted that it does not address the issue of the bounded probability scale.

In Study 1, accounting for the misclassification of direction of error reduced the fixed effect estimate by 78%, from 9.13 (SE = 0.58, p < 0.001) to 1.98 (SE = 0.44, p < 0.001), but still, an LMM exclusively testing trials with neutral events displayed a significant asymmetry (F(1,1436) = 20.73, p < 0.001). An LMM including event valence as a second fixed factor produced significant but reduced main effects of direction of error (F(1,96) = 71.52, p < 0.001) and event valence (F(2,4541) = 26.95, p < 0.001), and a weakened interaction term (F(2,4509) = 3.29, p = 0.037). While the same asymmetries are displayed (i.e., downwards direction of error elicited significantly greater updates than upwards for neutral, negative, and positive trials), there is a prominent reduction in the magnitude of these asymmetries (Figure S1).

---

[5]This analysis differs slightly from the pre-registered analysis plan in which we stated that we would test for an interaction between these classification schemes. Since the reclassification of direction of error also means that update values can change (i.e., an update of -10 would change to +10 if the direction of error is reclassified), different observations were identified as outliers by our exclusion criteria (i.e., ±3 × the interquartile range for a given condition). This in turn results in two distinct datasets: one where the standard classification scheme is applied and one where misclassification is accounted for. For this reason, it was not possible to test for an interaction by adding the classification scheme as a fixed factor in our LMMs that tests for effects within a dataset, and we instead provide qualitative comparisons between analyses.

Accounting for misclassification in Study 2 reduced the fixed effect estimate by 63%, from 6.24 (*SE* = 0.71, $p < 0.001$) to 2.32 (*SE* = 0.71, $p = 0.001$), but an LMM exclusively testing trials with neutral events again displayed an asymmetry ($F(1,1639) = 10.63$, $p = 0.001$). Reduced effects were also observed once event valence was included as a second fixed factor in the LMM, but nevertheless, there were still significant main effects of direction of error ($F(1,93) = 31.81$, $p < 0.001$) and event valence ($F(2,4534) = 9.55$, $p < 0.001$), and an interaction ($F(2,4410) = 8.33$, $p < 0.001$). Once again, the same significant asymmetries persisted but their magnitudes were heavily undercut (Figure S1).

In Study 3, misclassification accounted for 67% of the fixed effect estimate produced by an LMM exclusively testing trials with neutral events, reducing 6.51 (*SE* = 0.79, $p < 0.001$) to 2.15 (*SE* = 0.71 , $p = 0.002$). But, again, the asymmetry in trials with neutral events remained ($F(1,1565) = 9.23$, $p = 0.002$). However, once event valence is included in the LMM as a second fixed factor, the "flip" in asymmetries commonly interpreted as a result of valence-dependent updating disappears. While the effect of direction of error ($F(1,101) = 5.28$, $p = 0.024$), event valence ($F(2,4400) = 21.10$, $p < 0.001$), and the interaction remained significant ($F(1,4149) = 5.91$, $p = 0.003$), there is another notable reduction in the magnitude (Figure S1).

1.2 Comparisons with rational Bayesian predictions. Given that updates in different parts of the scale cannot be mathematically equated to one another, a seemingly sensible analysis of data produced by the update method is to compare participants' actual updating behavior to rational Bayesian predictions. As done in Shah et al. (2016), the collection of eBRs from participants in our studies allowed for the calculation of implied likelihood ratios (LHR) for each trial following the logic of Eq. 1-2:

$$Posterior\ Odds = Prior\ Odds\ \times\ LHR \qquad [1]$$

$$\frac{P(h|e)}{1-P(h|e)} = \frac{P(h)}{1-P(h)} \times\ LHR \qquad [2]$$

If eBR and E1 are then divided by 100, the equation can be rewritten with the terminology of the present experiments as follows in Eq. 3:

$$LHR = \frac{E1}{1-E1} \div \frac{eBR}{1-eBR} \qquad [3]$$

With the implied LHRs serving as a measure of individuating information participants believe they possess, we subsequently calculated the predicted posterior odds for each trial (Eq. 4), which could then be used to indicate how much a rational Bayesian agent "should" update in each trial (Eq. 5):

$$Posterior\ Odds = \frac{BR}{1-BR} \times\ LHR \qquad [4]$$

$$Bayesian\ Update = \left|\frac{E1 - Posterior\ Odds}{1 + Posterior\ Odds}\right| \qquad [5]$$

From here, we tested for asymmetries in belief updating with two measures across conditions, within participants: a Bayesian difference measure (i.e., predicted belief change – observed belief change) and a Bayesian ratio measure (i.e., observed belief change ÷ predicted belief change).

Our results show that there is variability between studies/samples. For instance, comparisons of upwards versus downwards updating with the Bayesian ratio measure indicates no asymmetry in trials with neutral events in Studies 2 and 3, but there is a statistically significant asymmetry observed in Study 1 and in the aggregated data of Studies 1-3 (Tables S6-S7). This observation highlights the inherited flaws of this analysis — although the comparisons are normatively appropriate, both the difference and ratio measures are susceptible to artefacts produced by the bounded probability scale and uneven effects of response noise (Shah et al., 2016). When a participant is required to translate a perceived personal risk estimate onto the probability scale, response noise will arise where a participant's non-integer estimates are forcibly rounded, where a participant misinterprets his or her internal state, or where a participant simply mis-types (e.g., entering "15" instead of "14"). The influence of such response noise will depend on where updating is taking place on the probability scale. For instance, as one approaches either end of the scale, response noise will constitute different proportions of the probability estimate. This issue is in turn reflected in the Bayesian comparison measures, deeming them insufficient to address the statistical artefact.

1.3 Analysis of learning rates derived from Kuzmanovic and Rigoux's (2017) computational model. A recent paper by Kuzmanovic and Rigoux (2017) proposed two new modelling techniques for investigating optimistic belief updating. First, they present a Bayesian model, in which they fit a scaling (S) and asymmetry (A) parameter to model participants subjective updates:

$$subjectiveUpdate_{good} = bayesianUpdate \times (S + A) \quad [6]$$

$$subjectiveUpdate_{bad} = bayesianUpdate \times (S - A) \quad [7]$$

The logic of these equations is that if participants update equally on desirable and undesirable information, the asymmetry parameter, *A*, will equal zero, and hence the right-hand bracketed expression will be constant across both equations. Thus, the modelling represented in these equations can be considered computationally equivalent to determining whether $\frac{subjectiveUpdate_{good}}{bayesianUpdate} = \frac{subjectiveUpdate_{bad}}{bayesianUpdate}$, which Shah et al. (2016), and the present work introduce, as the Bayesian ratio measure (Section 2.2).

Kuzmanovic and Rigoux (2017) also propose a reinforcement learning model. The full model presented is:

$$beliefUpdate = learningRate \times predictionError \times (1 - rP \times W) \quad [8]$$

where *beliefUpdate* represents the update value, *predictionError* represents the difference between eBR and BR, and *rP* represents "relative personal knowledge," and *W* is a free parameter to account for participants' individual variability in their sensitivity to *rP* (*W* is thus irrelevant when considering rational Bayesian agents) (Kuzmanovic & Rigoux, 2017). We additionally assessed the implications of the reinforcement learning model by addressing the crux of the argument: do learning rates differ across conditions? To do so, we simply rearranged Eq. 8 to permit a trial-by-trial calculation of learning rates:

$$learningRate = \frac{beliefUpdate}{predictionError \times (1 - rP)} \qquad [9]$$

Using Wilcoxon signed rank tests, we then compared learning rates across conditions, within participants. Once again, we observed unexplained variability in the results of each study — statistically significant asymmetries were observed in Study 1 and in the aggregated data, but not in Studies 2 and 3 — suggesting that the statistical artefact pervades this approach too with asymmetrical learning rates capable of being seen in trials with valence-neutral events (Table S8). Why this approach fails to address the artefact can be traced back to its failure to appropriately capture the influence of individuating information (Harris, Hahn, Burton, 2021). In the model, the *rP* parameter is used to account for individuating information, but it is not the LHR, which is the normatively appropriate way to capture this.

1.4 Regression analysis of updating behavior. Previous work analyzed participants' updating behavior by fitting linear regressions in which updates are entered as the dependent measure and "estimation error" (|E1 – BR|) is entered as the independent measure (Moutsiana et al., 2013; Sharot et al., 2011). Since "estimation error" is presumably correlated with the magnitude of updates, comparisons of regression coefficients across conditions, within participants is expected to display potential asymmetries while naturally controlling for the magnitude of "estimation error". In other words, if desirable trials have a larger regression coefficient than undesirable trials within participants, it would seem that participants are more conservative in belief updating when faced with bad news as compared to good news.

While not included in our pre-registered analysis plan, we followed this regression analysis procedure to further examine our data. Similar to the other supplementary analyses, this analysis displays considerable variability across studies/samples: there is a statistically significant asymmetry in trials with neutral events in Study 1, but not Studies 2 and 3, or in the aggregated data (Table S9). In addition, we repeated the regression analysis with base rate error (|eBR – BR|) entered as the independent variable, instead of "estimation error". Here we once again see variability in the results with statistically significant asymmetries observed in Study 1 and in the aggregated data, but not in Studies 2 and 3 (Table S10).

It is difficult to interpret these results because this regression analysis falsely equivocates upwards and downwards updating on the compressed probability scale (Figure S5). Normatively, the degree to which one should update his or her beliefs is the product of individuating information and the base rate. This means that even if two individuals are faced with the same BR, have identical likelihood ratios, and provide E1s that are equal absolute distances from the BR — but one agent's E1 is above the BR and the other's is below the BR — their prescribed Bayesian updates will differ (for further details see Shah et al. 2016). Yet, the regression analysis cannot account for this because it only considers the raw belief change and either estimation or BR error, while neglecting the influence of individuating information.

1.5 Accounting for post-treatment bias. As is the case for existing studies that use the update method and life events of varying valence (e.g., Garrett & Sharot, 2014, 2017), there is a possibility that a post-treatment bias may influence our models' estimates (see Montgomery et al., 2018 for a detailed exposition of post-treatment bias). Since participants provide their ratings of valence for each life event after having received the BR, the provision of the BR might influence the subsequent valence rating and the subsequent belief update. To remedy this potential problem in our main analysis, we re-ran the analysis as if every event were rated as neutral by the participants. Given that we aimed to compile a set of life events that could plausibly be rated as neutral by participants, this analysis is consistent with our research objective of detecting an asymmetry with valence-neutral events, despite its neglect of the variability in participants' perceptions of event valence. The results of this analysis mirror those of the main analysis in the main text, albeit slightly attenuated, meaning that an asymmetry was observed in upwards versus downwards updating across all life events.

In Study 1, there were 2,482 trials with an upwards direction of error (M = 2.72, SD = 5.90) and 2,336 with a downwards direction of error (M = 9.36, SD = 14.31). An LMM determined that direction of error significantly affected the magnitude of participants' updating ($F(1,4798)$ = 434.00, $p < 0.001$), such that an upwards direction of error (i.e., BR > E1) decreased update scores by approximately 6.43 percentage points (fixed effect estimate) ± 0.31 (standard error), as compared to downwards direction of error.

In Study 2, there were 2,288 trials with an upwards direction of error (M = 4.06, SD = 10.17) and 2,459 with a downwards direction of error (M = 9.16, SD = 18.22). An LMM determined that direction of error significantly affected the magnitude of participants' updating ($F(1,4735)$ = 136.70, $p < 0.001$), such that an upwards direction of error (i.e., BR > E1) decreased update scores by about 5.02 percentage points (fixed effect estimate) ± 0.43 (standard error), as compared to downwards direction of error.

In Study 3, there were 2,429 trials with an upwards direction of error (M = 4.04, SD = 9.60) and 2,278 with a downwards direction of error (M = 8.95, SD = 20.80). An LMM determined that direction of error significantly affected the magnitude of participants' updating ($F(1,4701)$ = 118.21, $p < 0.001$), such that an upwards direction of error (i.e., BR > E1) decreased update scores by about 4.78 percentage points (fixed effect estimate) ± 0.44 (standard error) as compared to downwards direction of error.

1.6 Adding stimuli as a random factor. In the LMM in our main analysis we included participants as a random factor to follow Marks and Baines (2017) and account for the nested structure of the data. Given that the main objective of the present work is to demonstrate that the update method — as it has been employed in the literature — can elicit asymmetric belief updating with neutral events, it was deemed crucial to follow analysis plans with precedent in the literature. However, it can be argued that the design of the update method warrants the inclusion of stimuli (life events) as a random factor, and that not doing so could inflate Type I error rates on the fixed effect estimates (Judd et al., 2012; Yarkoni, 2020). As a check of robustness, we therefore conducted an additional analysis where we re-fit the LMMs in our main analysis with stimuli as a random factor[6]. In each study, the asymmetry in belief updating with neutral life events remained with slightly attenuated fixed effect estimates.

In Study 1, an LMM determined that direction of error significantly affected the magnitude of participants' updating ($F(1,1507)$ = 222.13, $p < 0.001$), such that an upwards direction of error (i.e., BR > E1) decreased update scores by approximately 8.95 percentage points (fixed effect estimate) ± 0.60 (standard error), as compared to downwards direction of error.

[6] We used the same procedure to select a model specification as described in the main analysis in the main text, which led us to reduce the complexity of the random effects structure to include only random intercepts by participant and random intercepts by stimuli. However, results also hold in the maximally complex model specifications.

In Study 2, an LMM determined that direction of error significantly affected the magnitude of participants' updating ($F(1,1505) = 64.77$, $p < 0.001$), such that an upwards direction of error (i.e., BR > E1) decreased update scores by about 5.96 percentage points (fixed effect estimate) ± 0.74 (standard error), as compared to downwards direction of error.

In Study 3, an LMM determined that direction of error significantly affected the magnitude of participants' updating ($F(1,1442) = 61.05$, $p < 0.001$), such that an upwards direction of error (i.e., BR > E1) decreased update scores by about 6.42 percentage points (fixed effect estimate) ± 0.82 (standard error) as compared to downwards direction of error.

2. Supplementary Study 4

During the peer-review process special attention was given to Supplementary Analysis 1.4, which is the regression analysis used in the original work of Sharot et al. (2011), and which displays results seemingly in line with motivational optimism in the aggregated data from Studies 1-3 (Table S9). However, the results of this analysis are also seen to vary across Studies 1-3 and return a statistically significant asymmetry with neutral events in Study 1. As pointed out in the discussion section of the main text, this variability across studies with 100 participants each is particularly troubling because the update method is frequently used in neuroscientific studies with few participants, and we would thus expect such studies to be even noisier than those reported here. Nevertheless, we sought to better understand this variability and assess the robustness of these results by running simulations and subsequently running an additional pre-registered experiment, Study 4. All data and code for Study 4 is available on the OSF project page.

Our simulations, presented in the Study 4 pre-registration, stem from the statistical artefact hypothesis' explanation that "the very nature of the artefacts that plague the update method mean that, given the right set of events, everything and anything could empirically be found, even in entirely unbiased agents" (Shah et al., 2016, p. 107)." That is, the statistical artefact hypothesis predicts that the results of the regression analysis could change if we were to have sampled stimuli (life events) with different statistical properties (e.g., the events' average base rate error), which, crucially, are not valence-dependent. By simulating 500 "experiments" where we sample participants and events from the aggregated data of Studies 1-3 we found that the results of the regression analysis do indeed seem to be driven by statistical properties of the events used: in response to events with low average base rate error (i.e., |eBR-BR|) participants display asymmetric updating when they rate those events to be neutrally-valenced, but not when they rate them to be positively- or negatively-valenced (Figure S7). This is of course a nonsensical result that cannot be attributed to motivational optimism.

In order to empirically test the conclusions of our simulations we conducted Study 4, which involved recruiting 200 participants via the Prolific Academic platform ($M_{age} = 30.66$, $SD_{age} = 11.35$; 133 female, 63 male, 4 other) and presenting them with the 20 life events that elicited the lowest average base rate error in Studies 1-3 (see events list in Study 4 pre-registration). We applied the same analyses that were used to analyze Studies 1-3: LMMs as in the main text, and Supplementary Analyses 1.1-1.6. The LMMs and Supplementary Analyses 1.1, 1.5, and 1.6 replicated the results of Studies 1-3 and showed statistically significant asymmetries with neutral events; however, the results of Supplementary Analyses were inconclusive. Supplementary Analyses 1.2 (Bayesian comparisons) and 1.3 (learning rates) returned non-significant results for each event type, neutral, negative, and positive. Supplementary Analysis 1.4 (regression analysis) returned non-significant results for neutral [$t(130) = 1.08$, $p = 0.282$] and negative events

[$t(62) = 0.48$, $p = 0.635$], and a statistically significant asymmetry with positive events [$t(163) = -2.06$, $p = 0.041$].

We also note that there were noticeable departures in the distributions of how the participants in Study 4 estimated the relevant statistical properties of the new event sub-set vis a vis the data on which our selection had been based (Figure S8). This underscores further the limits of the current update methodology.

3. Tables

Table S1: Set of life events and accompanying base rate statistics to be used as stimuli. Participants will be asked to "Please estimate how likely this event is to happen to you," and "Please estimate how likely this event is to happen to the average person." Source indicates where the life event and base rate was obtained, with "new" indicating events that have not be previously used in research.

Source	ID	Life event	Base rate (%)
Shah et al. (2016)	1	Be exactly the same weight in 10 years' time	26
	2	Last the whole of next winter without catching a minor cold	20
Garrett & Sharot (2017)	3	Participate in a game of sport in the next four weeks	29
	4	Clean the bathroom in the next four weeks	78
	5	50 or more hours of sleep in a single week in the next four weeks	56
	6	Fix a broken possession in the next four weeks	39
	7	Get a haircut in the next four weeks	45
	8	Have your photo taken in the next four weeks	75
	9	Play a board game in the next four weeks	29
	10	Shop for clothes in the next four weeks	56
	11	Try a new hobby, craft, or sport in the next four weeks	31
	12	Receive a utility bill in the next four weeks	78
	13	Win a competitive game of sport in the next four weeks	22
	14	Burn something that you are cooking in the next four weeks	41
	15	Embarrass yourself in the next four weeks	60
	16	Get lost in the next four weeks	26
	17	Have a disagreement with a friend in the next four weeks	43
	18	Have a headache in the next four weeks	82
	19	Be ill one day because of over-drinking in the next four weeks	21

	20	Stay up past 2 AM for school or work in the next four weeks	40
	21	Get teased at/made fun of in the next four weeks	35
	22	Get lied to in the next four weeks	60
	23	Get stuck in traffic in the next four weeks	71
New	24	The next car that passes is a BMW	14
	25	Have a vegan meal in the next four weeks	14
	26	Make a purchase by contactless card in the next four weeks	29
	27	Check your phone more that 100 times in one day in the next four weeks	45
	28	The next car that passes is the colour black	20
	29	Receive a phone call from an unknown number in the next four weeks	66
	30	Buy a non-dairy milk alternative in the next four weeks	48
	31	Spend more than £121 on dinners out over the next four weeks	19
	32	Spend less than £89 on commuting over the next four weeks	33
	33	Send fewer than 106 text messages over the next four weeks	15
	34	Feel a phantom phone vibration in the next four weeks	80
	35	Walk less than seven miles over the next four weeks	17
	36	That your next flight will have a minor delay (i.e., 15 minutes or less)	26
	37	That the next store you visit is air conditioned	30
	38	Receive junk mail in the next four weeks	71
	39	Drink between 56 and 84 cups of coffee over the next four weeks	43
	40	Make your bed every day for the next four weeks	21
	41	Use more than 3.7GB of mobile data over the next four weeks	17
	42	Check your mobile data usage in your phone's settings in the next four weeks	13
	43	Spend more than 40 hours online in the next week	81

44	The next car you ride in, other than your own, is the colour white	19
45	Take the Eurostar train service in the future	16
46	Own a pet	45
47	Live in a home that was originally built before 1900	20
48	Move homes more than 10 times in your lifetime	18
49	Enrol in private health insurance	11
50	Meet your future spouse through an online dating service	38
51	Marry someone with a different political affiliation to you	26

Table S2: Results of linear mixed effects model with only neutral trials and the maximally complex random effects structure. This specification includes random slopes and intercepts by participant for direction of error, plus correlations between random effects. This model specification is singular, hence the reporting of a simpler specification in the main text. Statistics pertain to Type III tests of the fixed effect of the direction of error on belief updating. Degrees of freedom are approximated with Satterthwaite's method (dfn refers to the numerator degrees of freedom and dfd refers to the denominator degrees of freedom).

Study	*dfn*	*dfd*	*F*	*p-value*
1	1	112.57	131.46	< 0.001
2	1	139.75	61.57	< 0.001
3	1	107.53	45.64	< 0.001

Table S3: Results of linear mixed effects model with direction of error, event valence, and an interaction term and the maximally complex random effects structure. This specification includes random slopes and intercepts by participant for direction of error, event valence, and the interaction term, plus correlations between random effects. Fitting this model led to singularities and negative eigenvalues, hence the reporting of a simpler specification in the main text. Statistics pertain to Type III tests of the models' fixed effects. Degrees of freedom are approximated with Satterthwaite's method (dfn refers to the numerator degrees of freedom and dfd refers to the denominator degrees of freedom).

Study	*Fixed Factor*	*dfn*	*dfd*	*F*	*p-value*
	Direction of Error	1	114.31	198.47	< 0.001
1	Event Valence	2	127.62	33.66	< 0.001
	Interaction	2	160.89	49.19	< 0.001
	Direction of Error	1	123.41	104.90	< 0.001
2	Event Valence	2	136.27	29.82	< 0.001
	Interaction	2	154.78	38.96	< 0.001
	Direction of Error	1	95.12	48.67	< 0.001
3	Event Valence	2	133.32	13.72	< 0.001
	Interaction	2	143.60	47.26	< 0.001

Table S4: Results of linear mixed effects model with only neutral trials and the maximally complex random effects after accounting for misclassification. This specification includes random slopes and intercepts by participant for direction of error, plus correlations between random effects. This model specification is singular, hence the reporting of a simpler specification in the supplementary text. Statistics pertain to Type III tests of the fixed effect of the direction of error on belief updating. Degrees of freedom are approximated with Satterthwaite's method (dfn refers to the numerator degrees of freedom and dfd refers to the denominator degrees of freedom).

Study	*dfn*	*dfd*	*F*	*p-value*
1	1	246.67	19.47	< 0.001
2	1	473.51	10.81	0.001
3	1	162.2	7.84	0.006

Table S5: Results of linear mixed effects model with direction of error, event valence, and an interaction term and the maximally complex random effects structure after accounting for misclassification. This specification includes random slopes and intercepts by participant for direction of error, event valence, and the interaction term, plus correlations between random effects. Fitting this model led to singularities and negative eigenvalues, hence the reporting of a simpler specification in the supplementary text. Statistics pertain to Type III tests of the models' fixed effects. Degrees of freedom are approximated with Satterthwaite's method (dfn refers to the numerator degrees of freedom and dfd refers to the denominator degrees of freedom).

Study	*Fixed Factor*	*dfn*	*dfd*	*F*	*p-value*
	Direction of Error	1	178.04	65.14	< 0.001
1	Event Valence	2	195.49	19.63	< 0.001
	Interaction	2	168.84	2.33	0.101
	Direction of Error	1	601.39	30.06	< 0.001
2	Event Valence	2	318.13	8.47	< 0.001
	Interaction	2	815.00	8.13	< 0.001
	Direction of Error	1	217.40	5.77	0.017
3	Event Valence	2	277.15	19.77	< 0.001
	Interaction	2	211.56	5.95	0.003

Table S6: Results of paired t-tests comparing Bayesian difference measures that compare participants' updating to rational Bayesian predictions.

Study	*Event Valence*	*Mean of difference measure for downwards trials*	*Mean of difference measure for upwards trials*	*t*	*p-value*
1	Positive	0.13	0.09	-4.31	< 0.001
	Neutral	0.10	0.08	-2.00	0.049
	Negative	0.09	0.10	1.54	0.128
2	Positive	0.14	0.07	-4.57	< 0.001
	Neutral	0.11	0.08	-2.04	0.044
	Negative	0.06	0.10	2.63	0.009
3	Positive	0.12	0.07	-2.79	0.006
	Neutral	0.10	0.07	-2.52	0.013
	Negative	0.06	0.11	2.84	0.006
Aggregate	Positive	0.13	0.08	-6.47	< 0.001
	Neutral	0.10	0.07	-3.81	< 0.001
	Negative	0.07	0.10	4.14	< 0.001

Table S7: Results of Wilcoxon signed rank tests comparing Bayesian ratio measures that compare participants' updating to rational Bayesian predictions.

Study	*Event Valence*	*Median ratio measure for downwards trials*	*Median ratio measure for upwards trials*	*Z*	*p-value*
1	Positive	0.00	0.00	1.41	0.921
	Neutral	0.57	0.04	-3.73	< 0.001
	Negative	0.49	0.00	-4.92	< 0.001
2	Positive	0.00	0.17	-1.19	0.116
	Neutral	0.53	0.28	0.79	0.784
	Negative	0.51	0.07	-2.26	0.012
3	Positive	0.00	0.46	-2.14	0.016
	Neutral	0.53	0.28	-0.13	0.449
	Negative	0.51	0.09	-3.71	< 0.001
Aggregate	Positive	0.00	0.10	-2.36	0.009
	Neutral	0.53	0.25	-2.34	0.010
	Negative	0.51	0.00	-6.25	< 0.001

Table S8: Results of Wilcoxon signed rank tests comparing the learning rate measure derived from the reinforcement learning model presented by Kuzmanovic and Rigoux (2017).

Study	*Event Valence*	*Median learning rate for downwards trials*	*Median learning rate for upwards trials*	*Z*	*p-value*
1	Positive	0.00	0.00	0.28	0.610
	Neutral	0.61	0.04	-3.04	0.001
	Negative	0.60	0.00	-4.83	< 0.001
2	Positive	0.00	0.19	-0.89	0.187
	Neutral	0.68	0.45	0.53	0.704
	Negative	0.56	0.14	-1.86	0.032
3	Positive	0.00	0.55	-2.45	0.007
	Neutral	0.64	0.30	-0.46	0.324
	Negative	0.57	0.12	-4.02	< 0.001
Aggregate	Positive	0.00	0.15	-2.15	0.016
	Neutral	0.66	0.28	-2.14	0.016
	Negative	0.58	0.00	-6.12	< 0.001

Table S9: Results of paired t-tests comparing regression coefficients whereby “estimation error” is used to predict update values.

Study	*Event Valence*	*Mean coefficient for downwards trials*	*Mean coefficient for upwards trials*	*t*	*p-value*
1	Positive	-0.02	0.10	-1.90	0.060
	Neutral	0.20	0.00	3.57	< 0.001
	Negative	0.20	0.20	2.23	0.028
2	Positive	0.04	0.30	-3.22	0.002
	Neutral	0.11	0.06	0.44	0.662
	Negative	0.23	0.11	1.40	0.165
3	Positive	-0.14	0.11	-2.06	0.042
	Neutral	0.13	0.08	0.62	0.540
	Negative	0.40	0.13	1.52	0.132
Aggregate	Positive	-0.04	0.17	-4.03	< 0.001
	Neutral	-0.15	0.05	1.75	0.081
	Negative	0.27	0.15	2.76	0.006

Table S10: Results of paired t-tests comparing regression coefficients whereby base rate error is used to predict update values.

Study	*Event Valence*	*Mean coefficient for downwards trials*	*Mean coefficient for upwards trials*	*t*	*p-value*
1	Positive	0.04	0.11	-0.85	0.398
	Neutral	0.21	0.04	3.04	0.003
	Negative	0.30	0.21	3.78	< 0.001
2	Positive	0.08	0.06	0.28	0.777
	Neutral	0.21	0.06	1.83	0.071
	Negative	0.33	0.21	1.07	0.285
3	Positive	-0.04	-0.06	0.08	0.934
	Neutral	0.17	0.01	1.89	0.061
	Negative	0.31	0.17	1.80	0.075
Aggregate	Positive	0.03	0.04	-0.13	0.895
	Neutral	0.19	0.04	3.62	< 0.001
	Negative	0.31	0.19	3.20	0.002

4. Figures

Figure S1: Plots of the observed asymmetries in belief updating once the misclassification of direction of error is accounted for in each study. Points indicate the estimated marginal means of belief updating as predicted by the linear mixed effects model with bars representing 95% confidence intervals. Numbering of plots corresponds to the study.

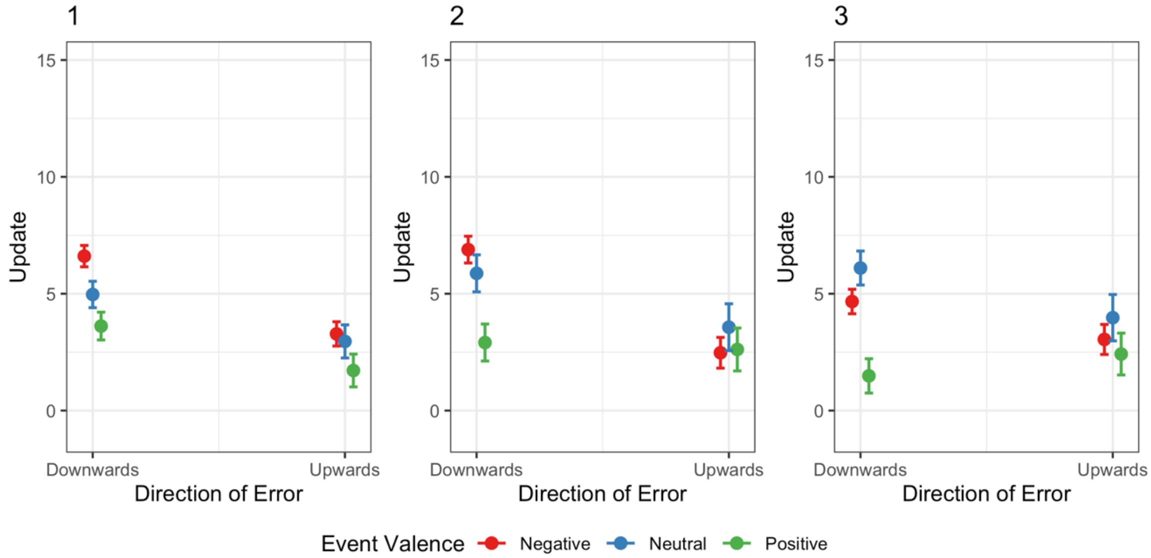


Figure S2: Density plots displaying the distributions of event base rates across studies (top labels) and event valence (right labels). It should be noted, however, that because each participant self-rates the valence of each event, each participant is likely to encounter different distributions of base rate statistics.

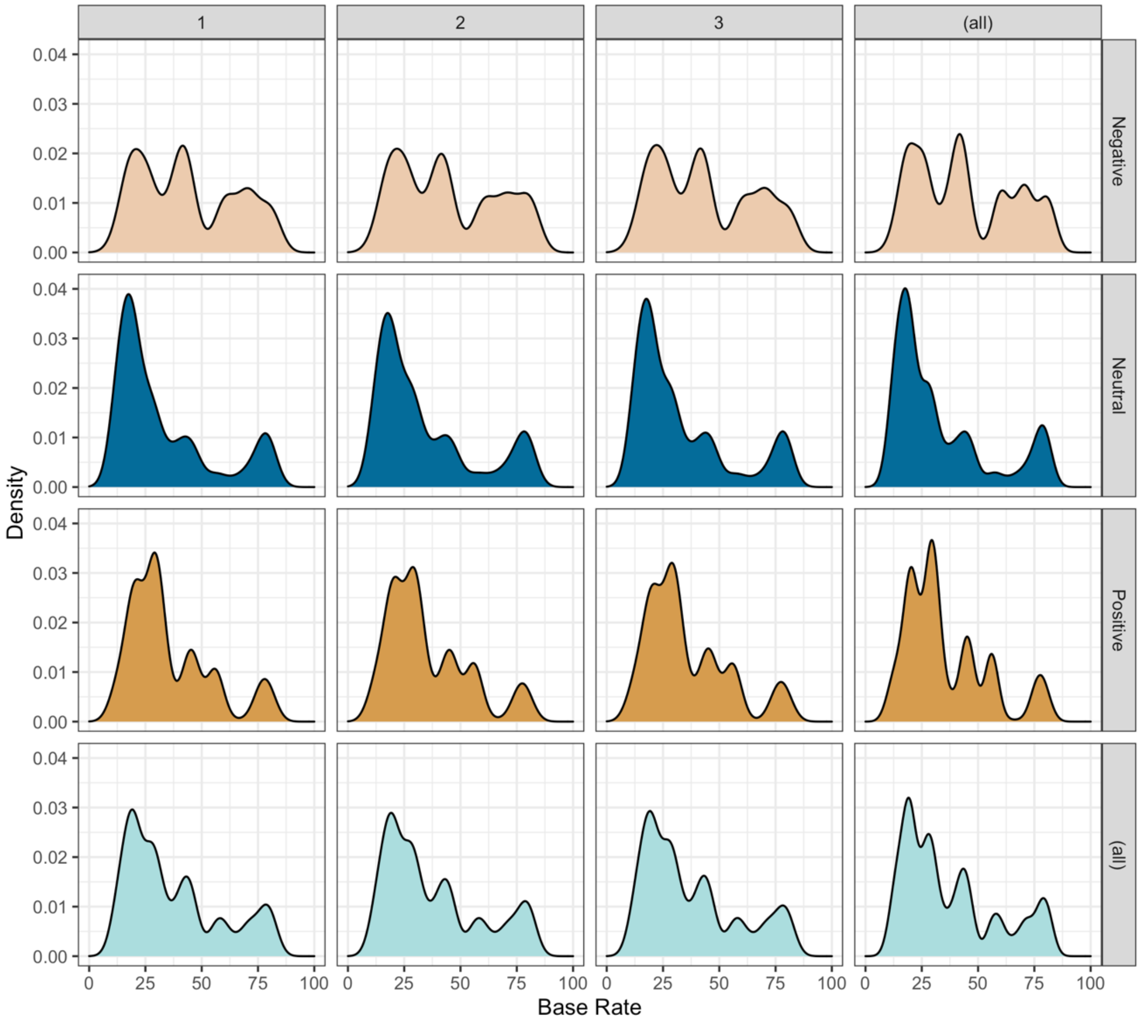


Figure S3: Density plots displaying the distributions of log transformed likelihood ratios across studies (top labels) and event valence (right labels). Implied likelihood ratios were calculated for each trial as: $\frac{E1}{1-E1} \div \frac{eBR}{1-eBR}$.

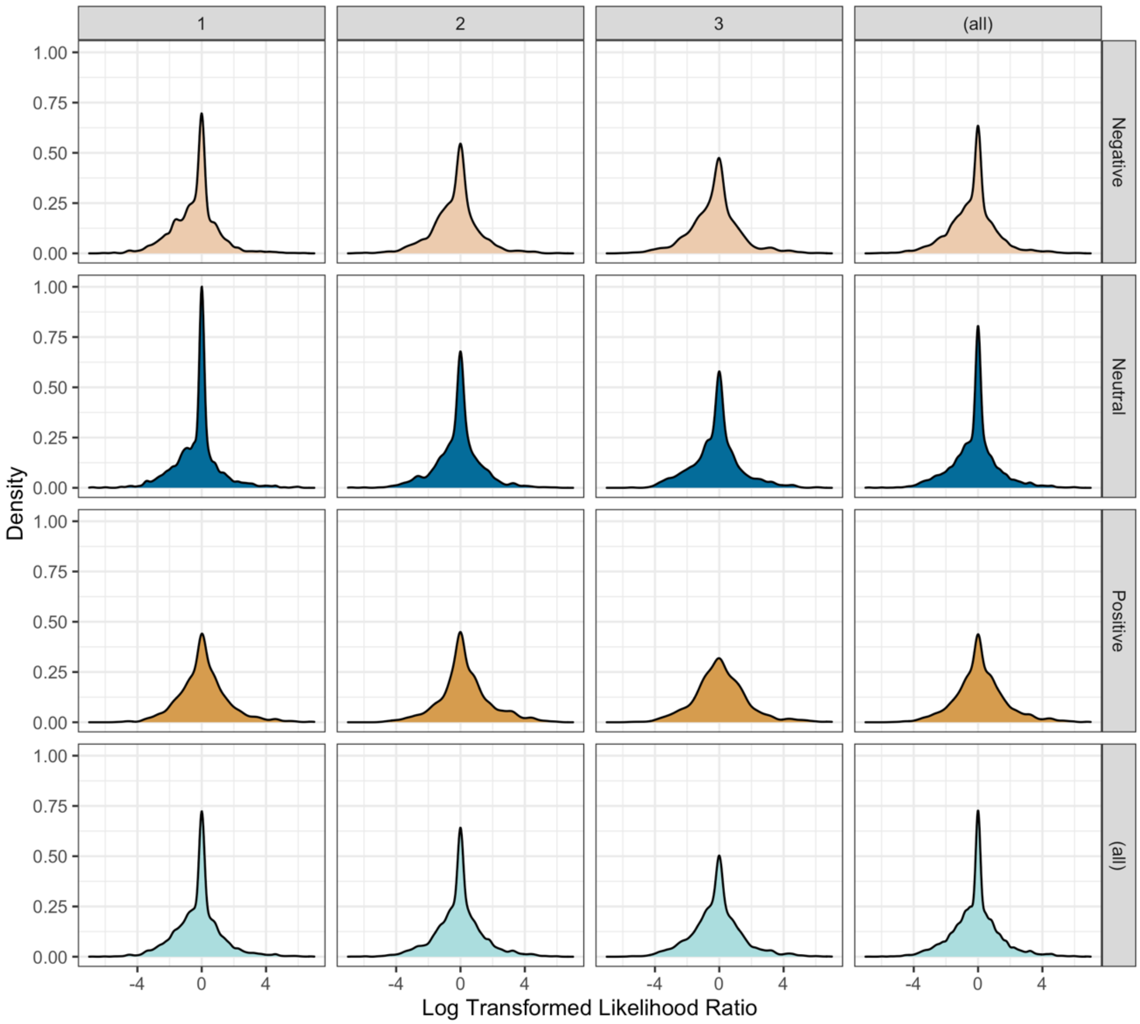


Figure S4: Density plots displaying the distributions of base rate error across studies (top labels) and event valence (right labels). Base rate error was calculated for each trial as: $|\boldsymbol{eBR} - \boldsymbol{BR}|$.

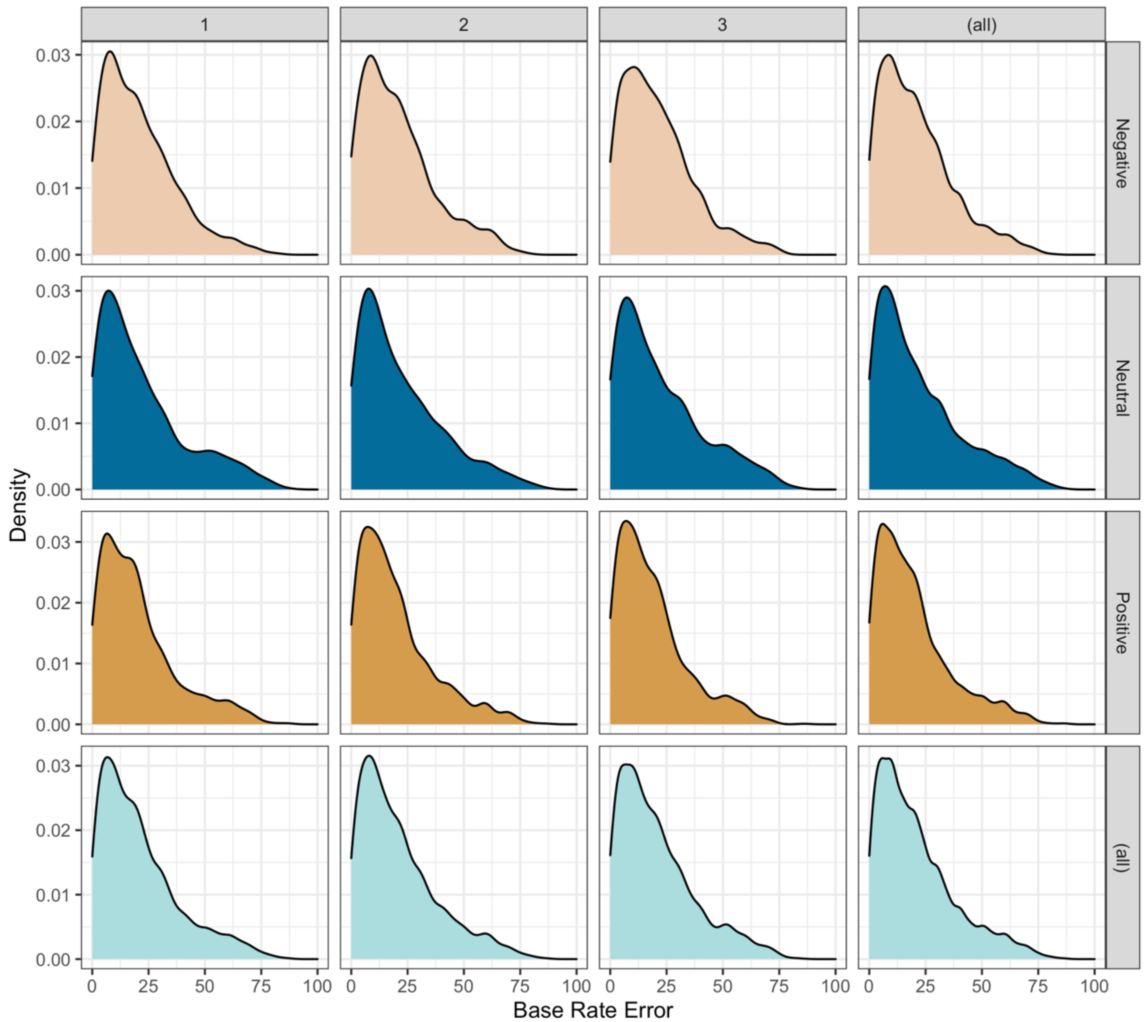


*Figure S5: Density plots displaying the distributions of "**estimation error**" across studies (top labels) and event valence (right labels). Estimation error was calculated for each trial as:* $|\boldsymbol{E1} - \boldsymbol{BR}|$.

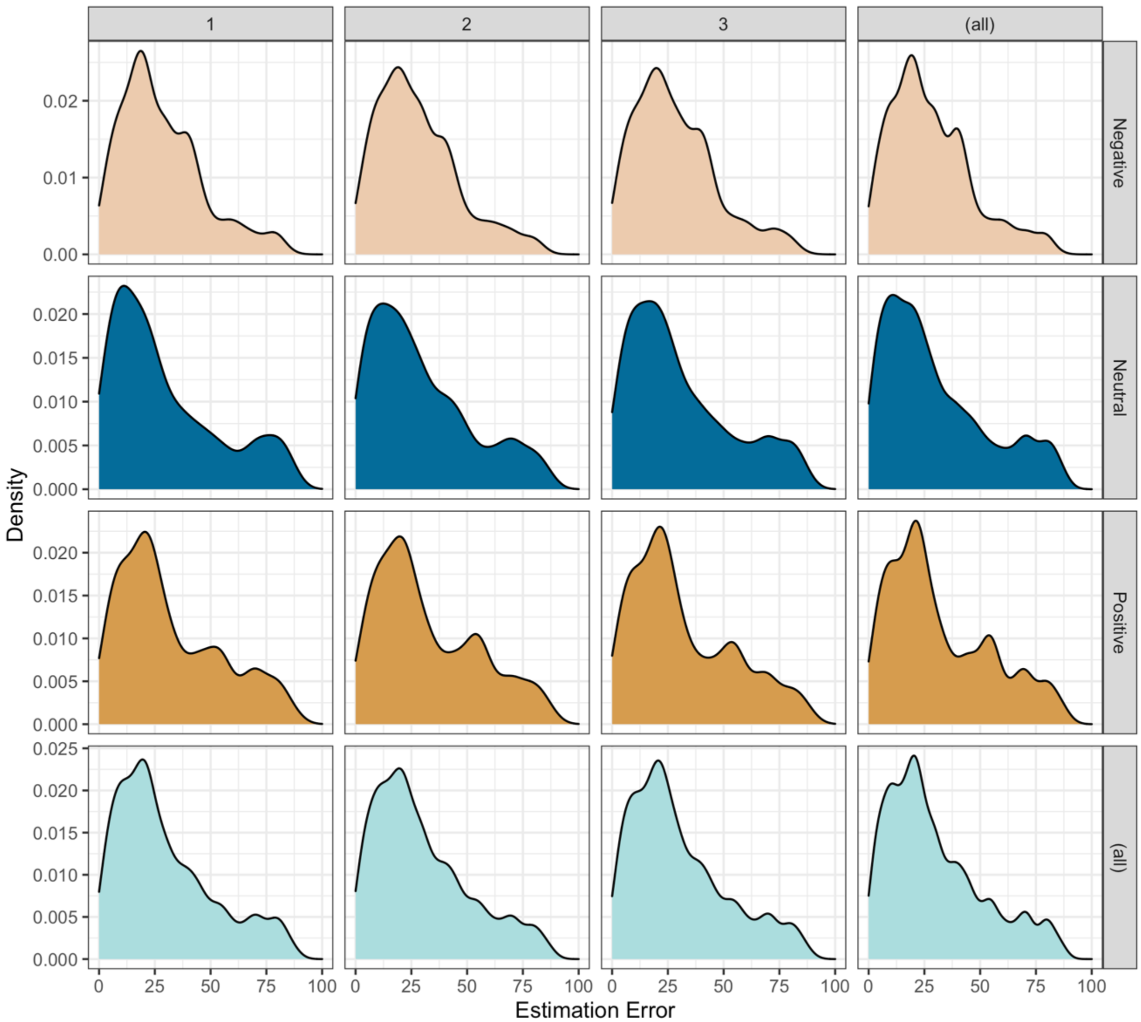


Figure S6: Probability scale compression and the relationship between base rate (BR) error, belief change, and implied likelihood ratios (LHR). **[A]** Ten simulated observations (x-axis) of paired base rates and posterior probabilities (y-axis). Across all ten pairs the LHR, ($\frac{E1}{1-E1} \div \frac{eBR}{1-eBR}$), that is, the degree of individuating knowledge is the same (in this plot, 0.4), and the posterior probability is derived via Bayes' theorem, by combining that individuating knowledge with the respective base rate. **[B]** Observations 2, 5, and 8 from Plot A. While the LHR (0.4) and absolute BR error (0.15) is held constant, the absolute belief change cannot be equal when updating in opposing directions (0.16 moving upwards on the scale from observation 5 to 2; 0.10 moving downwards on the scale from observation 5 to 8). **[C]** Observations 2, 5, and 8 when absolute belief change (0.15) and LHRs (0.4) are held constant, which results in unequal absolute BR errors (0.12 to move upwards on the scale from observation 5 to 2; 0.06 to move downwards on the scale from observation 5 to 8).

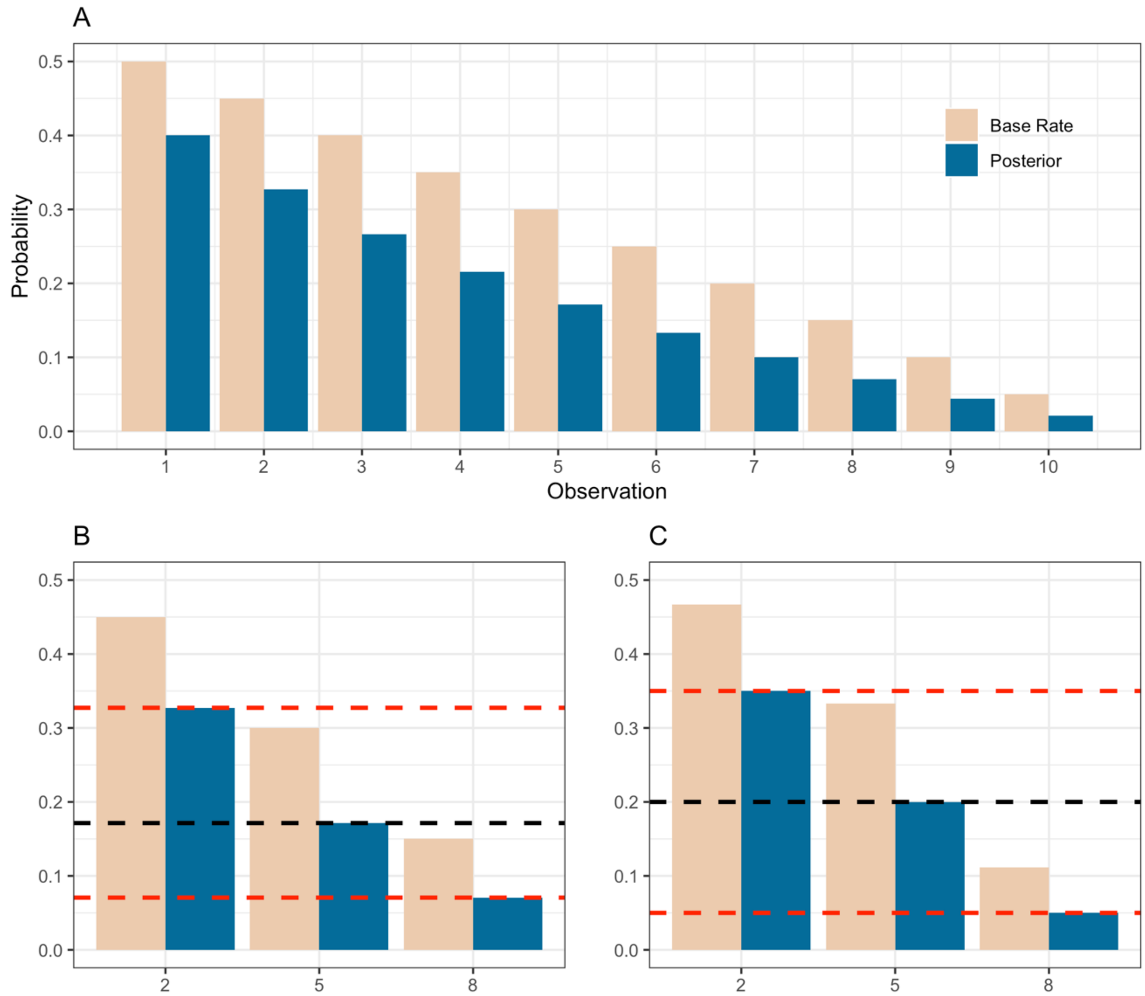


Figure S7: Results of the regression analysis using estimation error as the independent variable in 500 simulated "experiments." Each iteration, or experiment, involved sampling 200 participants and their responses to the 20 events with the lowest average base rate error ($|eBR - BR|$) from the aggregated data of Studies 1-3. Results are split out by event valence (y-axis). Blue points represent statistically significant ($p < 0.05$) asymmetries returned by paired samples t-tests comparing the regression coefficients in upwards vs. downwards updating. Red points represent non-significant asymmetries. The direction and magnitude of asymmetries is indicated by the t statistic (x-axis). Results suggest that, in response to events with low average base rate error, participants more frequently display asymmetric updating when they rate those events to be neutrally-valenced as compared to positively- or negatively-valenced.

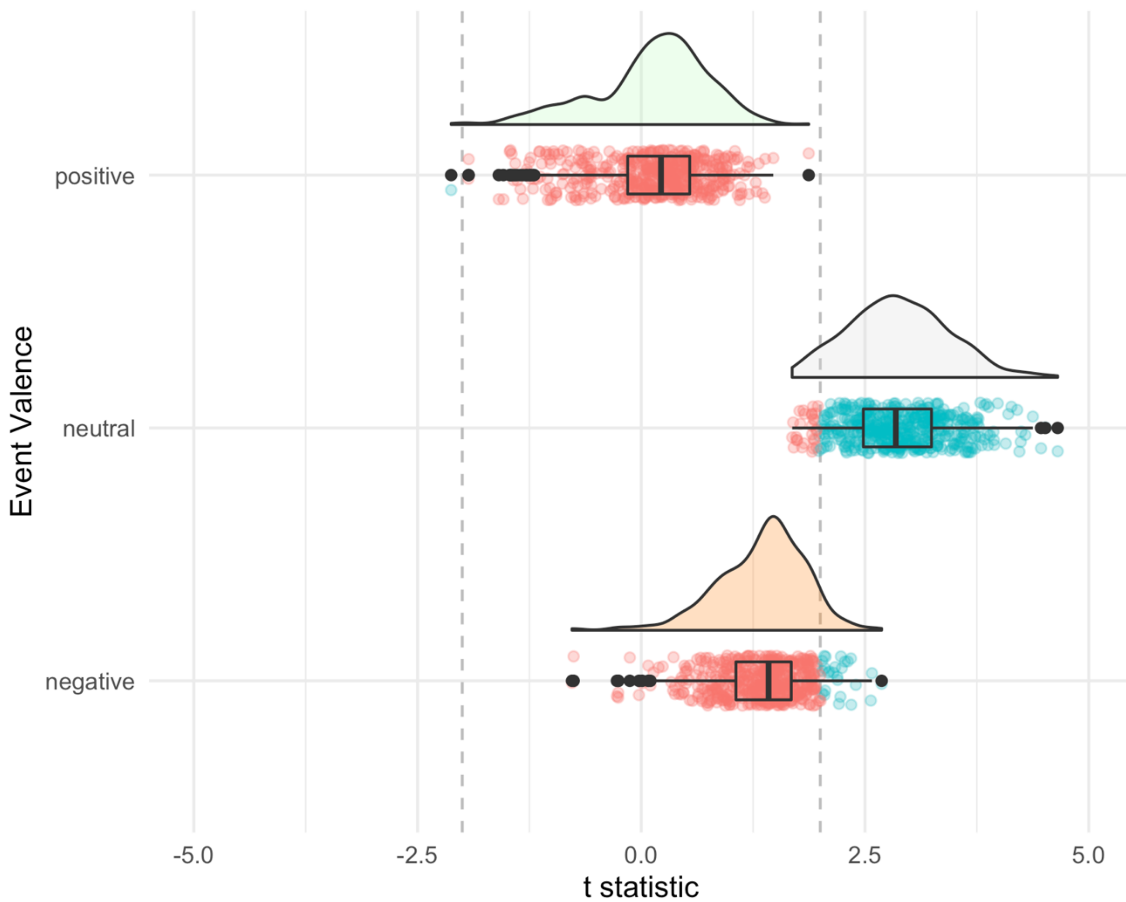
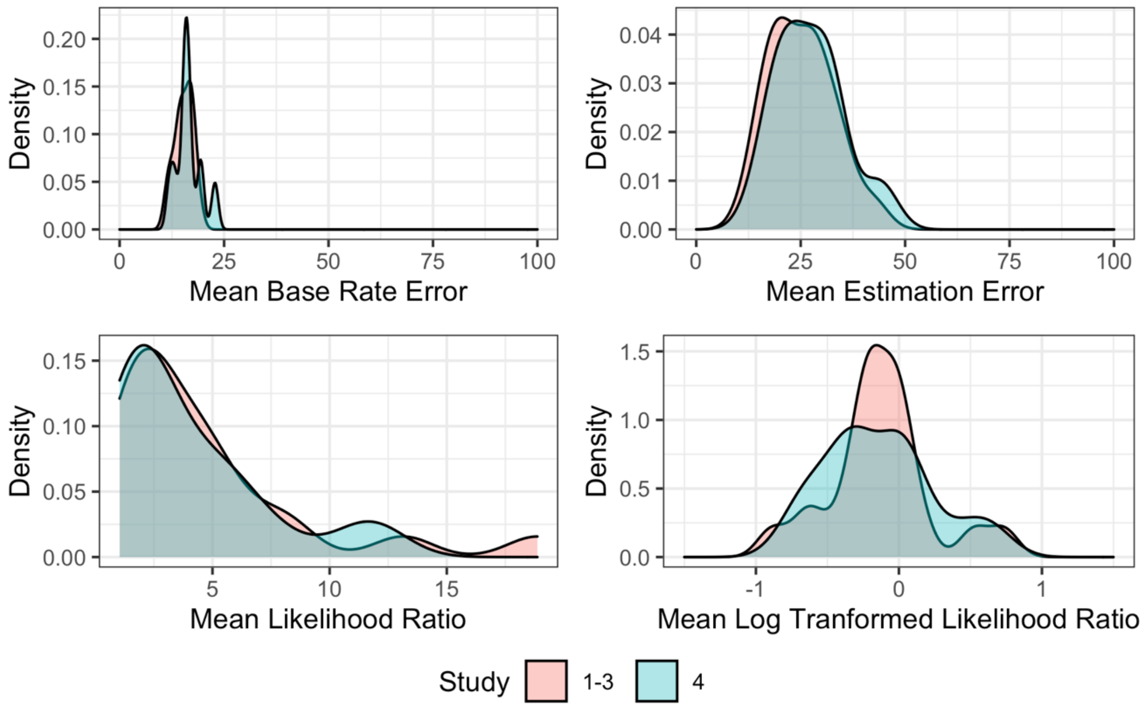


Figure S8: Density plots displaying the distributions of how participants estimated the statistical attributes of the 20 events used in Study 4 as compared to what was observed in Studies 1-3.



5. SI References

Garrett, N., & Sharot, T. (2014). How Robust Is the Optimistic Update Bias for Estimating Self-Risk and Population Base Rates? *PLoS ONE*, *9*(6), e98848. https://doi.org/10.1371/journal.pone.0098848

Garrett, N., & Sharot, T. (2017). Optimistic update bias holds firm: Three tests of robustness following Shah et al. *Consciousness and Cognition*, *50*, 12–22. https://doi.org/10.1016/j.concog.2016.10.013

Harris, A. J. L., Hahn, U., & Burton, J. W. (2021). Problems with optimistic belief updating tasks compromise behavioral and neuroscientific conclusions. *Manuscript in preparation (invited resubmission)*.

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54–69. https://doi.org/10.1037/a0028347

Kuzmanovic, B., & Rigoux, L. (2017). Valence-Dependent Belief Updating: Computational Validation. *Frontiers in Psychology*, *8*, 1087. https://doi.org/10.3389/fpsyg.2017.01087

Montgomery, J. M., Nyhan, B., & Torres, M. (2018). How Conditioning on Post-treatment Variables Can Ruin Your Experiment and What to Do about It [Data set]. In *American Journal of Political Science* (Vol. 62, Issue 3, pp. 760–775). https://doi.org/10.7910/DVN/EZSJ1S

Moutsiana, C., Garrett, N., Clarke, R. C., Lotto, R. B., Blakemore, S.-J., & Sharot, T. (2013). Human development of the ability to learn from bad news. *Proceedings of the National Academy of Sciences*, *110*(41), 16396–16401. https://doi.org/10.1073/pnas.1305631110

Shah, P., Harris, A. J. L., Bird, G., Catmur, C., & Hahn, U. (2016). A pessimistic view of optimistic belief updating. *Cognitive Psychology*, *90*, 71–127. https://doi.org/10.1016/j.cogpsych.2016.05.004

Sharot, T., Korn, C. W., & Dolan, R. J. (2011). How unrealistic optimism is maintained in the face of reality. *Nature Neuroscience*, *14*(11), 1475–1479. https://doi.org/10.1038/nn.2949

Yarkoni, T. (2020). The generalizability crisis. *Behavioral and Brain Sciences*, 1–37. https://doi.org/10.1017/S0140525X20001685