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[1]. 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).

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

[1]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.

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:

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

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[2]. 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.

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.,

[2] 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.

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*
1	Direction of Error	1	114.31	198.47	< 0.001
	Event Valence	2	127.62	33.66	< 0.001
	Interaction	2	160.89	49.19	< 0.001
2	Direction of Error	1	123.41	104.90	< 0.001
	Event Valence	2	136.27	29.82	< 0.001
	Interaction	2	154.78	38.96	< 0.001
3	Direction of Error	1	95.12	48.67	< 0.001
	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*
1	Direction of Error	1	178.04	65.14	< 0.001
	Event Valence	2	195.49	19.63	< 0.001
	Interaction	2	168.84	2.33	0.101
2	Direction of Error	1	601.39	30.06	< 0.001
	Event Valence	2	318.13	8.47	< 0.001
	Interaction	2	815.00	8.13	< 0.001
3	Direction of Error	1	217.40	5.77	0.017
	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.

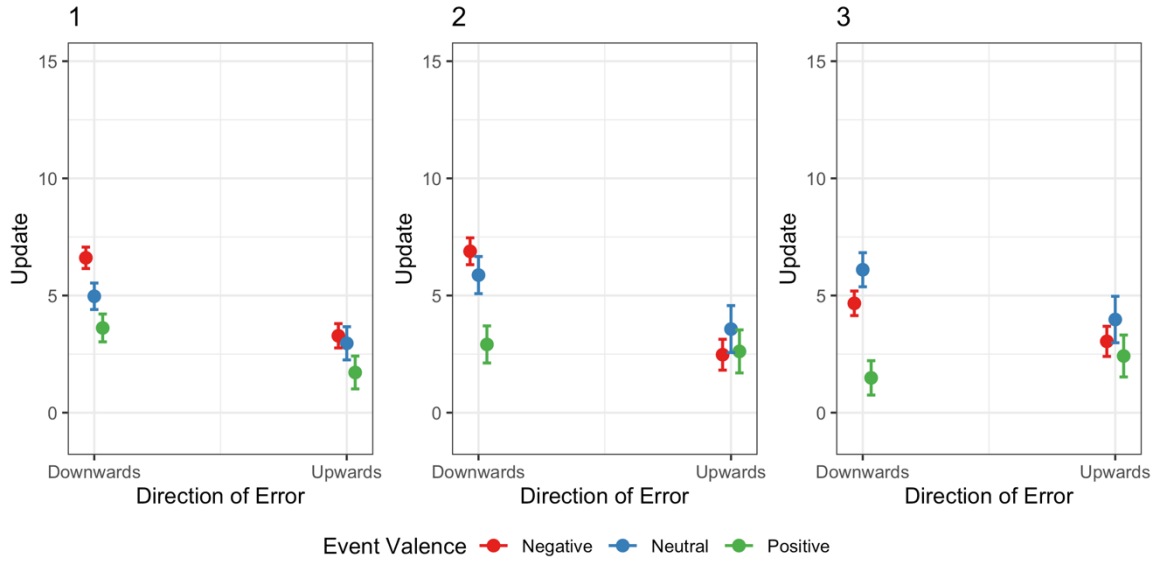


Figure S2: Density plots displaying the distributions of event base rates across studies (top labels) and event valence (right labels). It should be noted, however, that because each participant self-rates the valence of each event, each participant is likely to encounter different distributions of base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

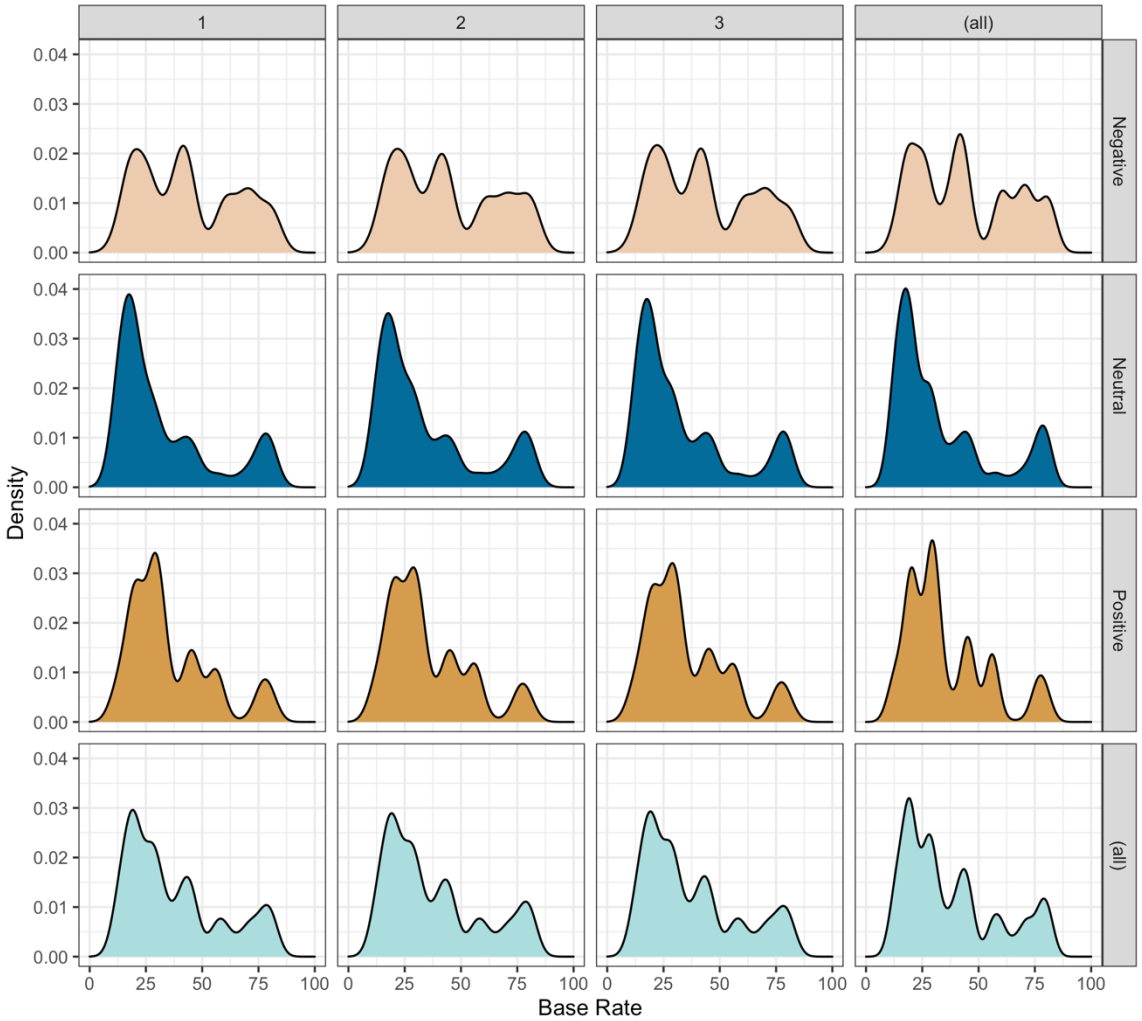


Figure S3: Density plots displaying the distributions of log transformed likelihood ratios across studies (top labels) and event valence (right labels). Implied likelihood ratios were calculated for each trial as: $\frac{E1}{1-E1} \div \frac{eBR}{1-eBR}$.

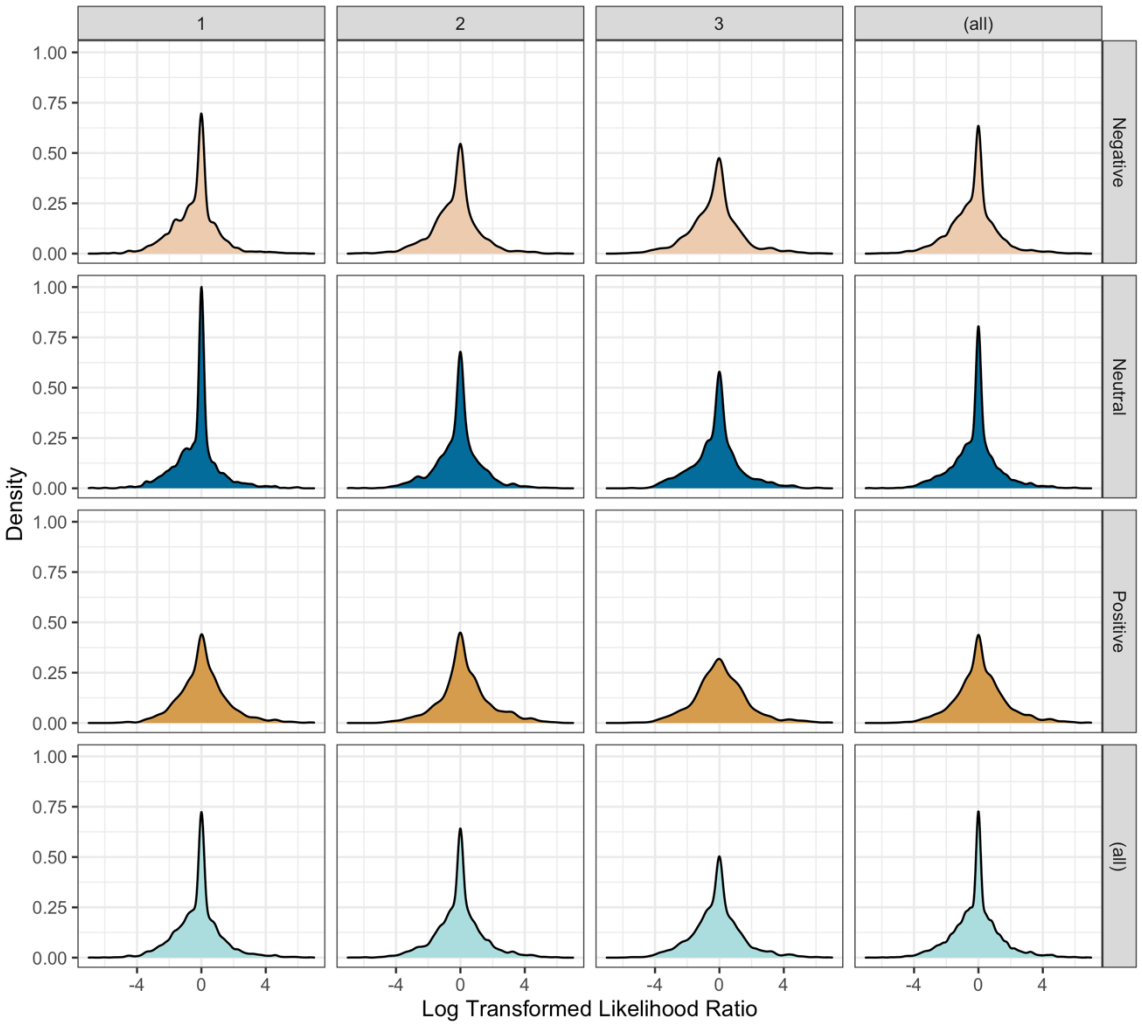


Figure S4: Density plots displaying the distributions of base rate error across studies (top labels) and event valence (right labels). Base rate error was calculated for each trial as: $|\boldsymbol{eBR} - \boldsymbol{BR}|$.

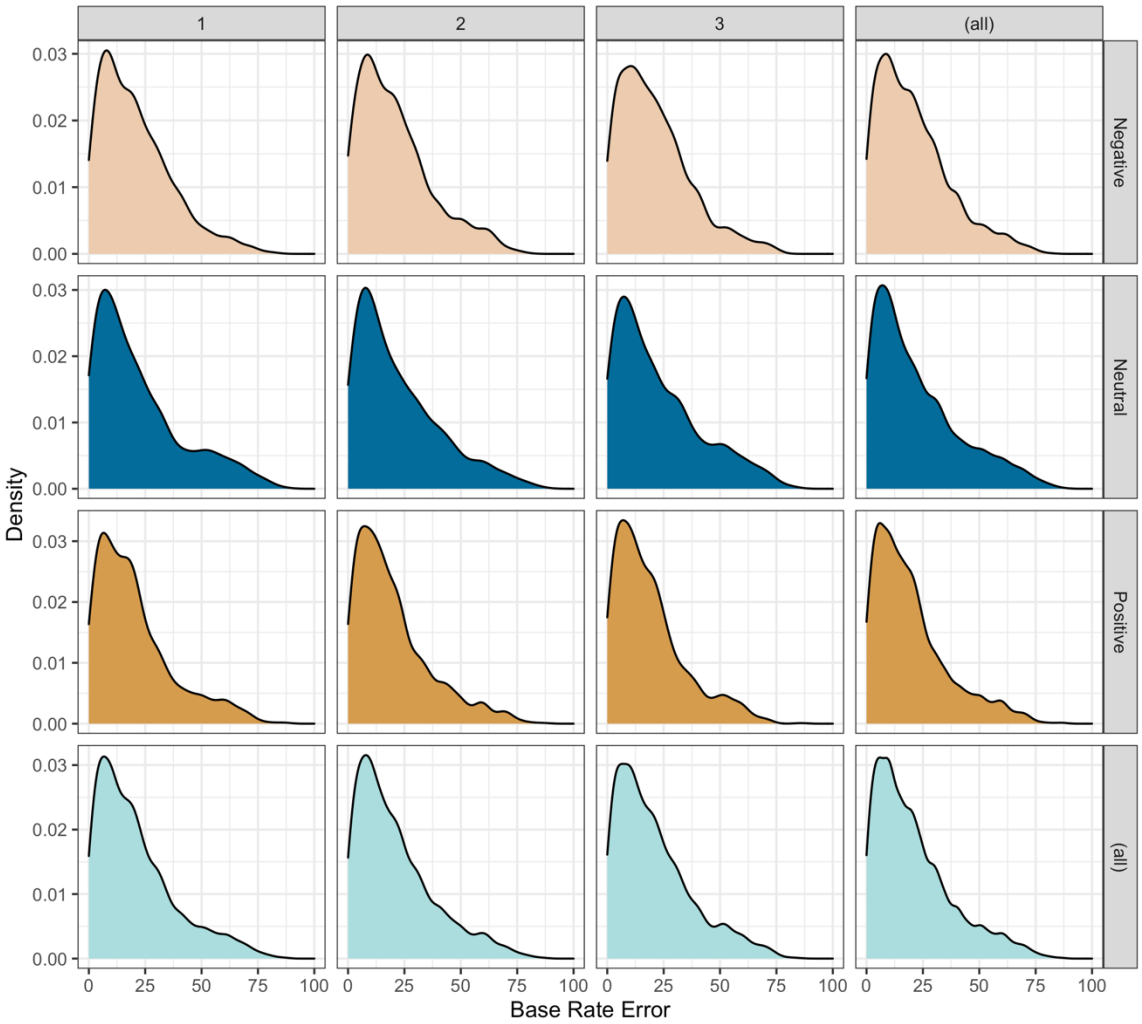


Figure S5: Density plots displaying the distributions of "estimation error" across studies (top labels) and event valence (right labels). Estimation error was calculated for each trial as: $|\boldsymbol{E1} - \boldsymbol{BR}|$.

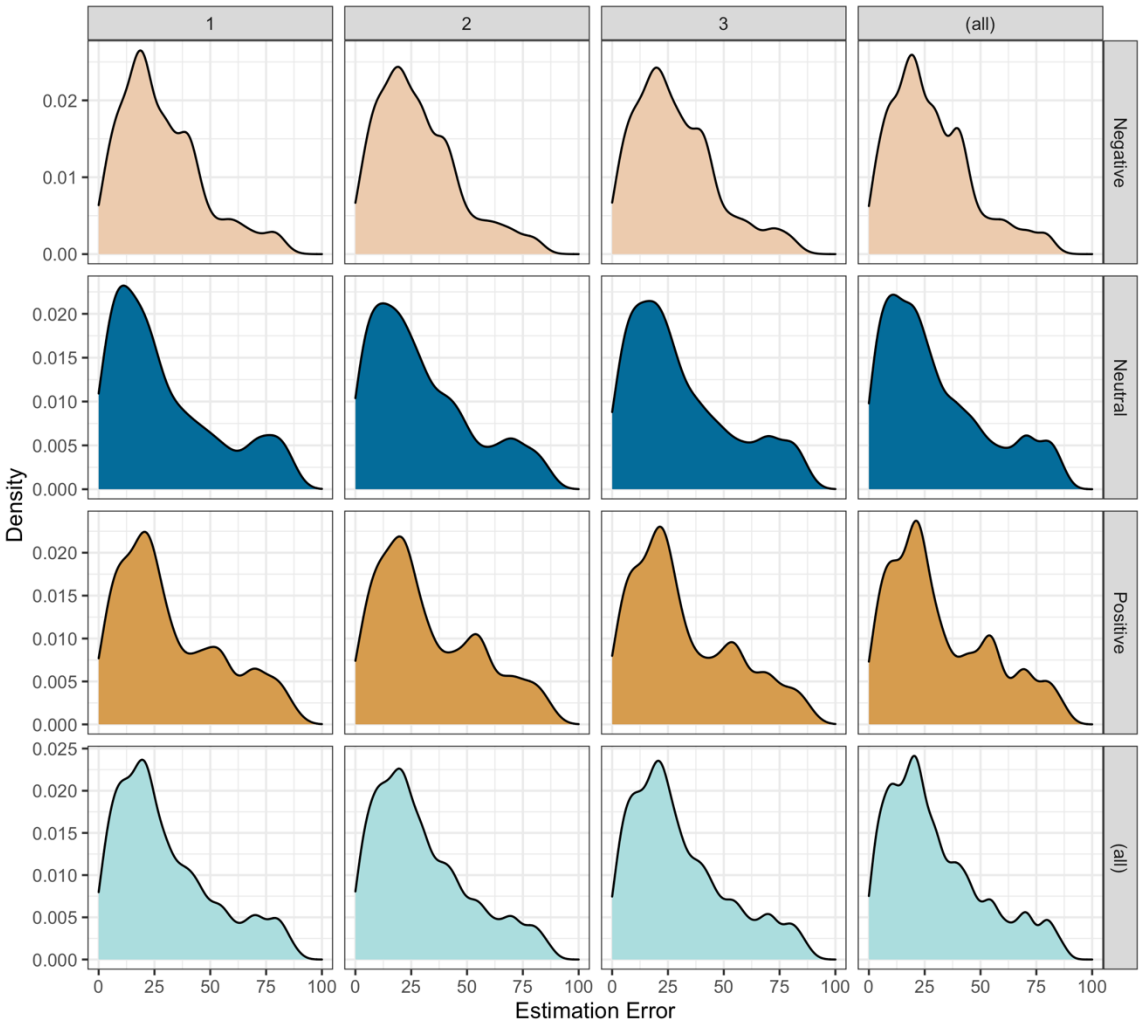


Figure S6: Probability scale compression and the relationship between base rate (BR) error, belief change, and implied likelihood ratios (LHR). **[A]** Ten simulated observations (x-axis) of paired base rates and posterior probabilities (y-axis). Across all ten pairs the LHR, $(\frac{E1}{1-E1} \div \frac{eBR}{1-eBR})$, 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).

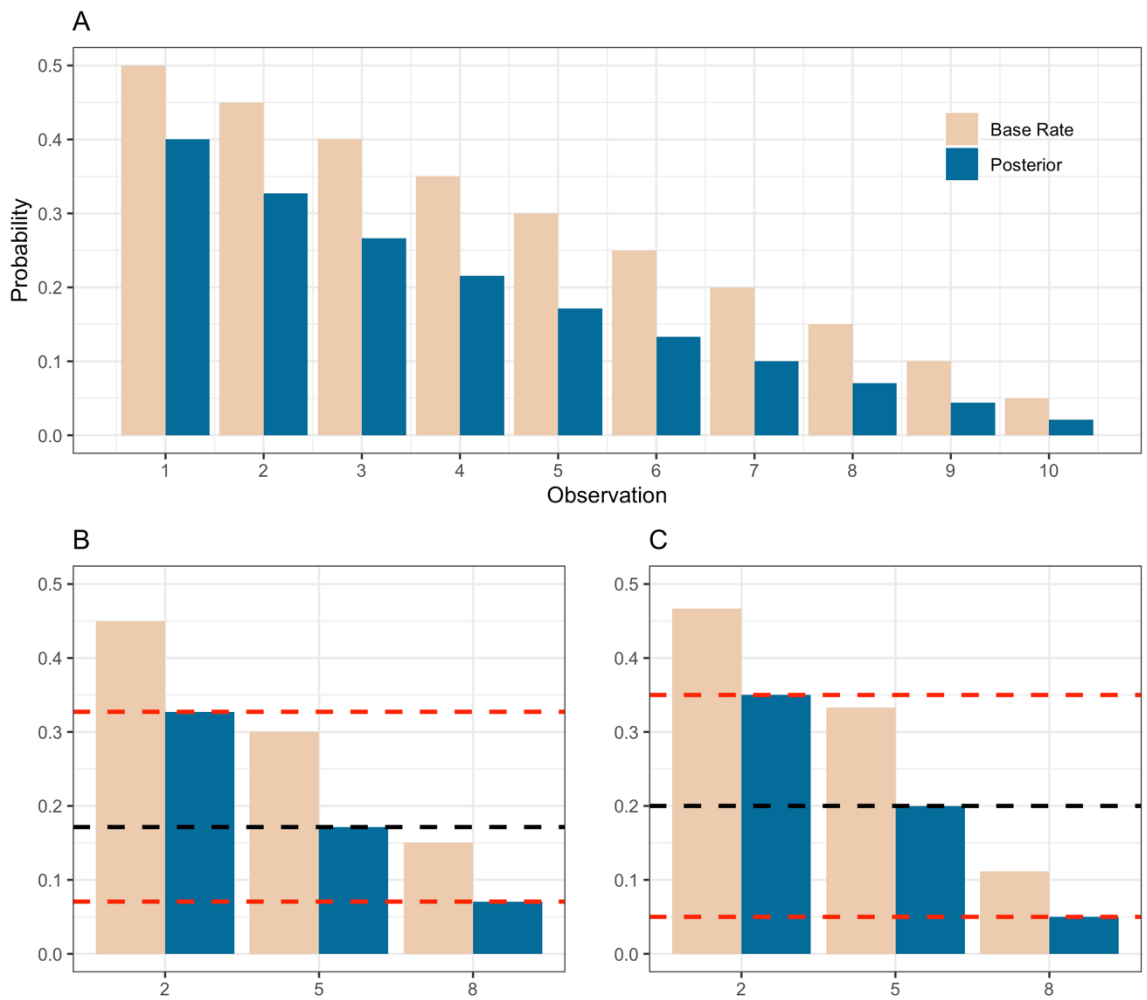


Figure S7: Results of the regression analysis using estimation error as the independent variable in 500 simulated "experiments." Each iteration, or experiment, involved sampling 200 participants and their responses to the 20 events with the lowest average base rate error ($|eBR - BR|$) from the aggregated 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.

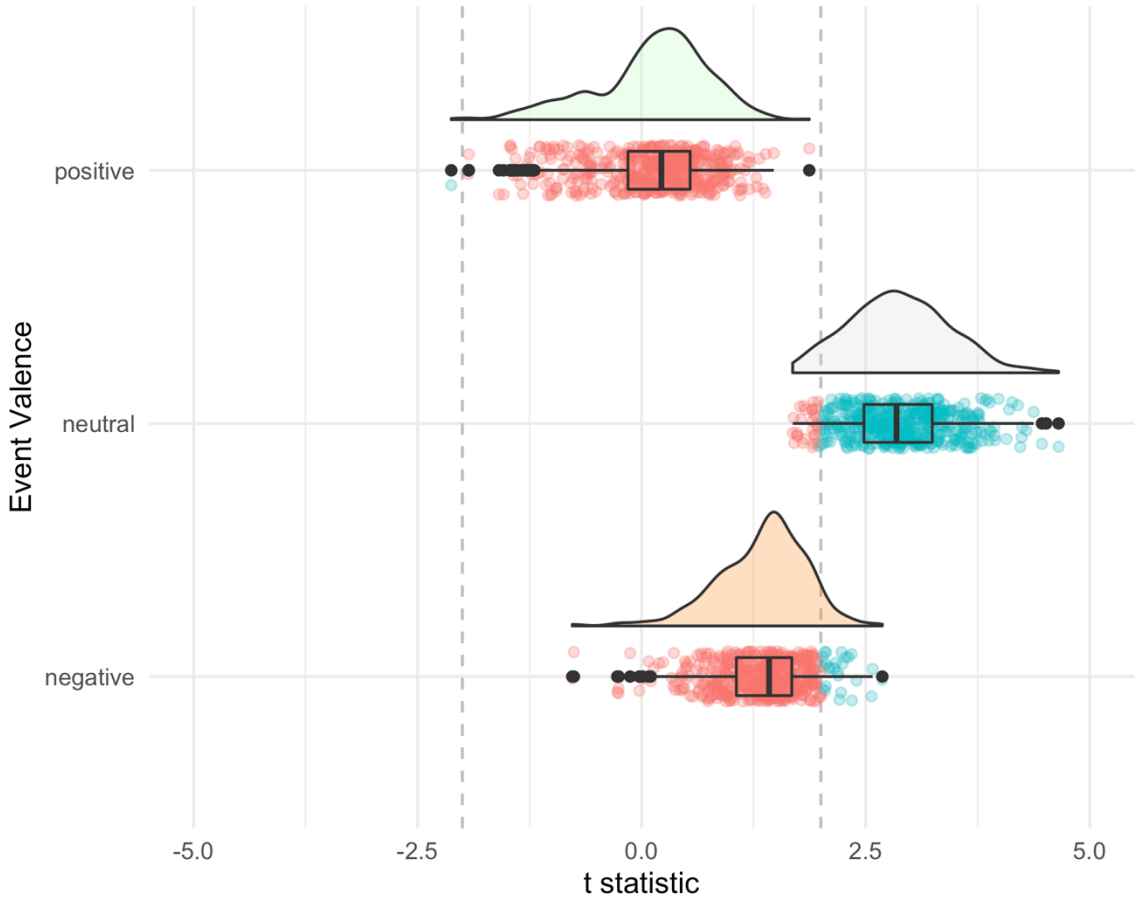
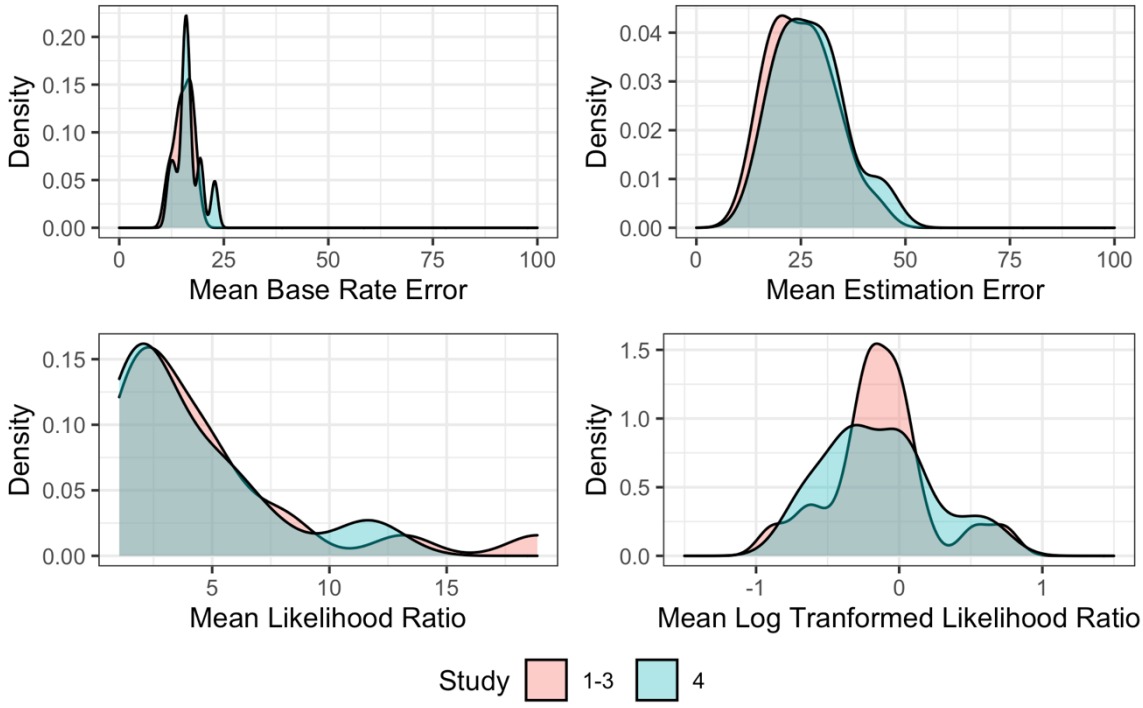


Figure S8: Density plots displaying the distributions of how participants estimated the statistical attributes of the 20 events used in Study 4 as compared to what was observed in Studies 1-3.



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