

# How unrealistic optimism is maintained in the face of reality

Tali Sharot<sup>1,4</sup>, Christoph W Korn<sup>2-4</sup> & Raymond J Dolan<sup>1</sup>

Unrealistic optimism is a pervasive human trait that influences domains ranging from personal relationships to politics and finance. How people maintain unrealistic optimism, despite frequently encountering information that challenges those biased beliefs, is unknown. We examined this question and found a marked asymmetry in belief updating. Participants updated their beliefs more in response to information that was better than expected than to information that was worse. This selectivity was mediated by a relative failure to code for errors that should reduce optimism. Distinct regions of the prefrontal cortex tracked estimation errors when those called for positive update, both in individuals who scored high and low on trait optimism. However, highly optimistic individuals exhibited reduced tracking of estimation errors that called for negative update in right inferior prefrontal gyrus. These findings indicate that optimism is tied to a selective update failure and diminished neural coding of undesirable information regarding the future.

People's estimations of the future are often unrealistically optimistic<sup>1-5</sup>. A problem that has puzzled scientists for decades is why human optimism is so pervasive when reality continuously confronts us with information that challenges these biased beliefs<sup>5</sup>. According to influential learning theories, agents should adjust their expectations when faced with disconfirming information<sup>6,7</sup>. However, this normative account is challenged by observations that providing people with evidence that disconfirms their positive outlook often fails to engender realistic expectations. For example, highlighting previously unknown risk factors for diseases is surprisingly ineffective at altering an individual's optimistic perception of their medical vulnerability<sup>8,9</sup>. Even experts show worrying optimistic biases. For instance, financial analysts expect improbably high profits<sup>3</sup> and family law attorneys underestimate the negative consequences of divorce<sup>2</sup>.

The wider societal importance of these errors derives from the fact that they reduce precautionary actions, such as practicing safe sex or saving for retirement<sup>8,9</sup>. On the upside, optimistic expectations can lower stress and anxiety, thereby promoting health and well-being<sup>10-12</sup>. Although the existence of unrealistic optimism has been extensively documented<sup>1-5,13,14</sup>, the biological and computational principles that help to maintain optimistically biased predictions in the face of reality are unknown<sup>5</sup>. Notably, such biases are not explained by theories assuming equal learning across outcome valence<sup>6,7</sup>.

There is a large body of evidence suggesting that learning from information that disconfirms one's expectations is mediated by regions involved in error processing and conflict detection, including anterior cingulate cortex, adjacent medio-frontal regions and lateral prefrontal regions<sup>15-20</sup>. Here we tested whether a failure to alter optimistic predictions when presented with disconfirming data was mediated by differential error processing in response to information that was better, or worse, than expected.

We combined a learning procedure with functional brain imaging. Our task enabled us to quantify how much participants adjusted their beliefs in response to new information. Although optimism is defined both as the overestimation of positive future events and the underestimation of future negative events<sup>1</sup>, we focused on the latter, as this is strongly related to a concern that people do not take necessary action to protect themselves against hazards<sup>9</sup>. We obtained functional magnetic resonance imaging (fMRI) data while participants estimated their likelihood of experiencing 80 different adverse life events (for example, Alzheimer's disease, robbery; see **Supplementary List of Stimuli**). After each trial, participants were presented with the average probability of that event occurring to an individual living in the same socio-cultural environment as the participants (**Fig. 1a**). We then assessed whether participants used this information to update their predictions by subsequently asking them in a second session to provide estimates of their likelihood of encountering all 80 events.

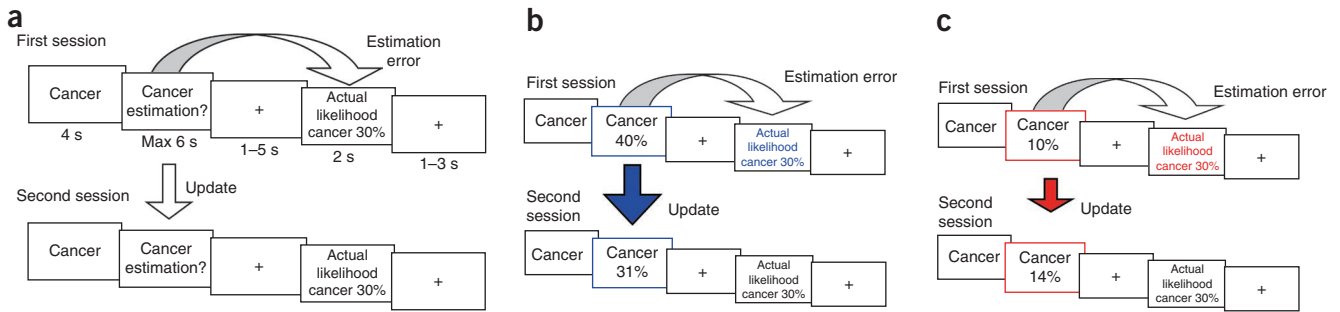
## RESULTS

### Behavior: selective updating

Our behavioral results revealed a marked asymmetry in belief updating. Participants learned to a greater extent from information offering an opportunity to adopt more optimistic expectations (referred to as desirable information; **Fig. 1b**) than from information that challenged participants' rosy outlook (referred to as undesirable information; **Fig. 1c**). Specifically, participants were more likely to update their beliefs when the average probability of experiencing a negative life event was lower than the participants' own probability estimate than in a situation in which the average probability was higher ( $t_{18} = 4.4, P < 0.001$ ; **Supplementary Fig. 1a**). This robust effect was observed in 79% of the participants (**Fig. 2a**) and replicated across

<sup>1</sup>Wellcome Trust Centre for Neuroimaging, University College London, London, UK. <sup>2</sup>Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany. <sup>3</sup>School of Mind and Brain, Humboldt Universität zu Berlin, Berlin, Germany. <sup>4</sup>These authors contributed equally to this work. Correspondence should be addressed to T.S. (t.sharot@ucl.ac.uk).

Received 20 June; accepted 7 September; published online 9 October 2011; doi:10.1038/nn.2949



**Figure 1** Task design. **(a)** On each trial, participants were presented with a short description of 1 of 80 adverse life events and asked to estimate how likely this event was to occur to them. They were then presented with the average probability of that event occurring to a person similar to themselves, living in the same socio-cultural environment. For each event, an estimation error term was calculated as the difference between the participant's estimation and the information provided. The second session was the same as the first session. For each event, an update term was calculated as the difference between the participant's first and second estimations. **(b,c)** Examples of trials for which the participant's estimate was higher **(b)** or lower **(c)** than the average probability. Here, for illustration purposes, the blue and red frames denote the participant's response (either an overestimation or underestimation, respectively). The blue and red text denote information that calls for an adjustment in an optimistic (desirable, **b**) or pessimistic (undesirable, **c**) direction, respectively.

additional behavioral studies (**Supplementary Behavioral Studies 1 and 2**). The difference was also significant when updates were calculated as the percentages of the initial estimate, ( $t_{18} = 4.7$ ,  $P < 0.001$ ; **Supplementary Fig. 1b**).

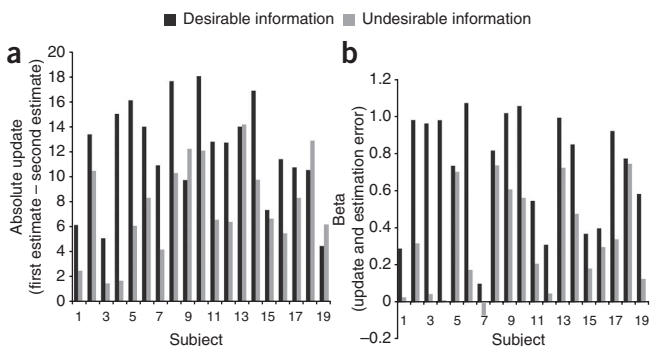
One obvious explanation for these results is that a greater weighting given to desirable information simply reflects differential memory for desirable compared with undesirable information. This was not the case. After the scanning session, participants were asked to indicate the actual probability (as previously presented) of each event occurring to an average person in the same socio-cultural environment. Memory errors were calculated as the absolute difference between the actual probability previously presented and the participants' recollection of that statistical number. Participants remembered information presented to them equally well, irrespective of whether it was desirable or undesirable ( $t_{18} = 0.75$ ,  $P > 0.4$ ; **Supplementary Fig. 1c** and **Supplementary Behavioral Studies 1 and 2**).

Furthermore, post-scanning questionnaire scores showed that any differential updating across valence was not explained by differences in emotional arousal, extent of negative valence, familiarity or past experience with the adverse life event (**Supplementary Results**). Specifically, the difference in absolute update for events for which participants received desirable and undesirable information remained significant after entering all of these scores as covariates ( $F_{1,13} = 9.7$ ,  $P < 0.01$ ). Thus, differential update could not be explained by differences in the degree of adversity of the events, by familiarity or by past experience with the events.

Notably, positively biased updating was not a result of differences in the underlying true probabilities (base rates) of the events. The difference

in update for events in which participants received desirable and undesirable information remained significant even after entering the true probabilities of the events as covariates ( $F_{1,17} = 6.04$ ,  $P < 0.05$ ). In other words, how common or rare the occurrence of an event is had no bearing on selective updating. Neither were there differences in the number of trials ( $t_{18} = 0.02$ ,  $P > 0.9$ ), the magnitude of the initial estimation errors ( $t_{18} = 1.85$ ,  $P > 0.05$ ) or reaction times ( $t_{18} = 1.04$ ,  $P > 0.3$ ) when participants received desirable or undesirable information. Finally, asymmetric updating was not explained by a differential processing of high and low percentages, as this was controlled for by asking participants to estimate their likelihood of encountering the adverse event on half of the trials and to estimate their likelihood of not encountering the adverse event on the other half of the trials (**Supplementary Results**).

Formal models suggest that learning is mediated by a prediction error signal that quantifies a difference between expectation and outcome<sup>7,21,22</sup>. We hypothesized that an analogous mechanism underpinned belief update in our task. We formulated the difference between participants' initial estimations and the information provided in terms of estimation errors (that is, estimation error = estimation – probability presented). Indeed, estimation errors predicted subsequent updates on an individual level (mean beta from individual linear regressions relating estimation errors to update = 0.53,  $P < 0.0005$ ). However, the strength of this association was valence dependent, being greater for information that offered an opportunity to adopt a more optimistic outlook (beta = 0.72) than for information that called for a more pessimistic outlook (beta = 0.33,  $t_{18} = 5.8$ ,  $P < 0.0005$ ; **Fig. 2b** and **Supplementary Fig. 1d**). Notably, this difference remained significant after controlling for the true underlying probabilities of the events by including them as a factor to the regression analysis ( $t_{18} = 2.6$ ,  $P < 0.02$ ).



**Figure 2** Behaviorally observed bias. **(a)** After receiving (desirable) information that presented an opportunity to adopt a more optimistic outlook, participants updated their estimations to a greater extent than after receiving (undesirable) information that called for a more pessimistic estimate. This asymmetric updates was observed in 15 out of 19 participants. For group means, see **Supplementary Figure 1a**. **(b)** Betas indicating the association between updates and estimation errors on an individual basis revealed that estimation errors predicted updates to a greater extent when participants received desirable information than when they received undesirable information. This asymmetry was observed in all 19 participants. For group means, see **Supplementary Figure 1d**.

**Figure 3** Brain activity tracking estimation errors. (a,b) Regions in which BOLD signal tracked participants' estimation errors on a trial by trial basis in response to desirable information regarding future likelihoods included the left IFG (a) and bilateral MFC/SFG (b) ( $P < 0.05$ , cluster level corrected). (c) BOLD signal tracking participants' estimation errors in response to undesirable information was found in the right IFG/precentral gyrus ( $P < 0.05$ , cluster level corrected). (d) Parameter estimates of the parametric regressors in both the left IFG and bilateral MFC/SFG did not differ between individuals with high or low scores on trait optimism. In contrast, in the right IFG, a stronger correlation between BOLD activity and undesirable errors was found for individuals with low scores on trait optimism relative to those with high scores. Error bars represent s.e.m. \* $P < 0.05$ , two-tailed independent sample  $t$  test.

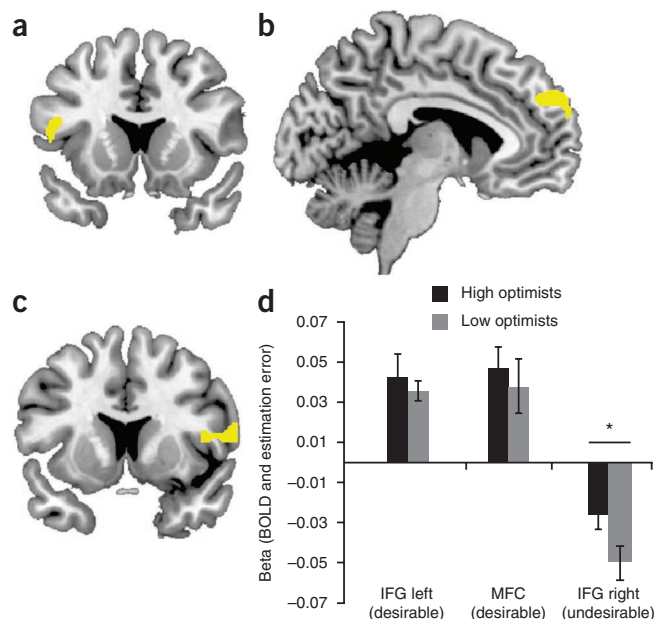
These behavioral findings suggest a likely computational principle that mediates unrealistic optimism. Specifically, they point to estimation errors as providing a learning signal whose impact depends on whether new information calls for an update in an optimistic or pessimistic direction. Consequently, we next examined our fMRI data to identify how blood oxygen level-dependent (BOLD) signal tracks estimation errors in response to information that entails either a belief adjustment in an optimistic or in a pessimistic direction.

### fMRI: tracking of estimation errors

We first identified voxels in which activity tracked desirable or undesirable estimation errors. Trials were partitioned on the basis of whether participants received desirable or undesirable information regarding the future after their first rating. We entered the absolute estimation error on each trial as a parametric regressor modulating the time point at which participants were presented with information regarding the average probability of events (Fig. 1). From this analysis, we identified regions in which BOLD signal correlated with estimation errors for either desirable or undesirable information on a trial by trial basis ( $P < 0.05$ , cluster level corrected after voxel-wise thresholding at  $P < 0.001$ ).

BOLD signal correlated positively with desirable estimation errors in three distinct regions: left inferior frontal gyrus (IFG, peak voxel in Talairach coordinates:  $-58, 21, -1$ ;  $k = 274$ ; Fig. 3a), left and right medial frontal cortex/superior frontal gyrus (MFC/SFG,  $-10, 62, 28$ ;  $k = 412$ ; Fig. 3b) and right cerebellum ( $33, -79, -28$ ;  $k = 185$ ). In addition, BOLD signal correlated negatively with undesirable estimation errors in the right IFG extending into precentral gyrus and insula ( $46, 12, 9$ ;  $60, 10, 9$ ;  $k = 190$ ; Fig. 3c). There were no voxels in which activity correlated negatively with desirable errors or positively with undesirable errors. In other words, brain activity increased when the average probability was more desirable than participants' estimates (in left IFG, bilateral MFC/SFG and right cerebellum) and decreased when it was more undesirable (in right IFG). All of these regions have been implicated in error processing in various tasks<sup>15–20</sup>.

As our behavioral findings tied estimation errors to update, it follows that activity that tracks estimation errors should also predict subsequent update. To test this, we entered the level of subsequent update as a parametric regressor modulating the time point at which participants were first presented with information regarding the average probability of adverse events. We then retrieved the parameter estimates for each participant and each condition in the four functional regions of interest (ROIs) identified above, averaged across all voxels. Indeed, activity in the right IFG predicted update in response to undesirable information (parameter estimates significantly greater than zero,  $t_{18} = 2.2$ ,  $P < 0.05$ ). Activity in MFC/SFG ( $t_{18} = 2.1$ ,  $P < 0.05$ ) and right cerebellum ( $t_{18} = 2.4$ ,  $P < 0.05$ ) predicted update



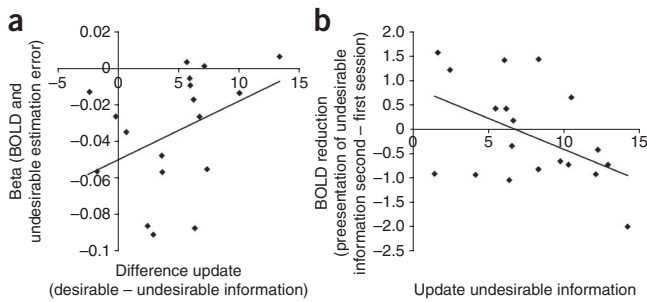
in response to desirable information (in the left IFG, parameter estimates were not significantly different from zero).

### Trait optimism, biased updating and BOLD signal

Next, we examined whether the extent to which brain activity tracked estimation errors was related to optimism. We compared betas from the parametric analysis (relating BOLD signal to estimation errors) in each functional ROI (averaged across all voxels) of participants with high scores for trait optimism with those who with low scores (according to the Life Orientation Test Revised scale, LOT-R<sup>23</sup>). There was no difference between high and low optimists on betas tracking desirable errors in the left IFG (Fig. 3d), bilateral MFC/SFG (Fig. 3d) or cerebellum. However, participants with high scores for trait optimism showed a weaker correlation of right IFG activity with undesirable errors relative to those with low scores (Fig. 3d). Specifically, betas indicating a correlation between BOLD signal and undesirable estimation errors in the right IFG showed a weaker inverse relation in high compared with low optimists ( $t_{17} = 2.1$ ,  $P < 0.05$ ). This difference could not be explained by the magnitude of the initial undesirable estimation errors ( $P > 0.8$ ) or memory for undesirable information ( $P > 0.5$ ), as neither differed between groups.

The same finding was obtained when we considered all four ROIs together in a single linear regression with group membership (high or low optimism) as the dependent measure, controlling for differential scores (desirable minus undesirable) of memory, familiarity, past experience, vividness and arousal, and extent of initial estimation errors. A stepwise procedure revealed that the best model predicting trait optimism was one that included only betas from the right IFG tracking undesirable estimation errors (beta = 0.46,  $P < 0.05$ ). This suggests a specific relationship between trait optimism and a reduced neural coding of undesirable errors regarding the future.

To examine this further we determined whether the optimistic update bias in our task related to how well BOLD activity tracked undesirable errors. Across individuals, BOLD signal tracking undesirable estimation errors predicted the observed update bias (desirable update minus undesirable update) in right IFG (partial correlation controlling for behavioral association between estimation errors and update,  $r_{12,3} = 0.45$ ,  $P = 0.05$ ). Participants with the largest optimistic update bias failed to track undesirable errors (the correlation between BOLD signal in



**Figure 4** Optimism and brain activity tracking undesirable estimation errors. (a) Across individuals, the extent of the update bias (difference in update in response to desirable errors minus undesirable errors) correlated with how strong the peak voxel in the right IFG ROI correlated with undesirable errors ( $r_{12} = 0.45$ ). Participants showing the greatest optimistic bias in updating showed the weakest tracking of undesirable estimation errors. (b) Participants who updated their estimates more in response to undesirable information showed a greater reduction in activity in the right IFG ROI the second time desirable information was presented relative to the first time the information was presented ( $r = -0.47$ ).

the right IFG and undesirable errors in these participants was close to zero; **Fig. 4a**). Participants who did not show selective updating had the strongest relationship between BOLD signal and undesirable estimation errors. Thus, these results suggest that an asymmetry in belief updating is driven by a reduced expression of an error signal in a region implicated in processing undesirable errors regarding the future.

A final prediction arising from our data is that those participants who updated their estimates the most in response to undesirable information should show greater BOLD reduction the second time undesirable information is presented. The logic here is that a greater update should correspond to smaller estimation errors in response to repeated presentation of information and thus to less overall activity in regions tracking estimation errors. To test this, we examined the correlation, across participants, between suppression of activity in the right IFG (BOLD signal at second presentation of undesirable information minus BOLD signal at first presentation of undesirable information, averaged across all voxels in the ROI) and the degree of update on undesirable trials. We found that participants who showed greater update in response to undesirable information had a greater reduction in activity at the second time point when undesirable information was presented ( $r = -0.47$ ,  $P < 0.05$ ; **Fig. 4b**); for a description of differential activation during the period when participants thought about encountering the adverse events, see **Supplementary Table 1**).

## DISCUSSION

Our findings offer a mechanistic account of how unrealistic optimism persists in the face of challenging information. We found that optimism was related to diminished coding of undesirable information about the future in a region of the frontal cortex (right IFG) that has been identified as being sensitive to negative estimation errors. Participants with high scores on trait optimism were worse at tracking undesirable errors in this region than those with low scores. In contrast, tracking of desirable information in regions processing desirable estimation errors (MFC/SFG, left IFG and cerebellum) did not differ between high and low optimists.

Reduced BOLD tracking of undesirable errors in right IFG predicted the extent to which participants selectively updated their beliefs using information that enforced optimism, while (relatively) dismissing information that contradicted it. Notably, this effect was not

explained by how well participants recalled the information presented to them and it did not reflect specific characteristics of the adverse events (for example, familiarity, negativity, arousal, past experience, or how rare or common the event was). Thus, unlike predictions from learning theory, where both positive and negative information are given equal weight<sup>6,7</sup>, we found a valence-dependent asymmetry in how estimation errors affected beliefs about one's personal future.

Our fMRI data indicate that error-evoked activity differed in response to desirable and undesirable information regarding possible future outcomes. Segregated regions encoded error-related activity in response to new information that called for optimistic or pessimistic adjustments. Although the left IFG, left and right MFC/SFG and right cerebellum tracked desirable errors, the right IFG tracked undesirable errors. It is worth noting that previous studies have suggested hemispheric asymmetry in processing positive and negative information that is consistent with that our results in the IFG<sup>24</sup>. Furthermore, although BOLD signal in regions tracking desirable estimation errors increased when the average probability was better than the participant's estimate, it decreased when the average probability was worse than expected in regions tracking negative estimation errors. In other words, activity increased for a better than expected outcome and dipped for a worse than expected outcome, a pattern resembling that of neurons signaling prediction errors<sup>21,22</sup>.

All of the regions that we identified as coding estimation errors have previously been shown to track errors in different contexts, including errors resulting from incorrect responses<sup>15</sup>, errors in expectations in the absence of action<sup>16</sup>, reversal errors<sup>17,18</sup>, and prediction errors that code differences between expectations and outcomes<sup>19</sup>. Note that those errors emerge when outcomes are experienced, and capture differences in expected and real magnitudes. Here estimation errors captured the difference between participants' predictions of the likelihood of a possible future outcome and information about the average probability of experiencing these outcomes. This information was presented in an explicit form that engaged higher cognitive functions, and this is likely to explain the preferential engagement of cortical regions. Our findings suggest that these error signals are subject to motivational modulation (also see ref. 20). Specifically, a motivation to adopt the most rewarding (or least aversive) perspective on future outcomes is likely to modulate the effect of an error signal that subsequently influences update. As information regarding hypothetical outcomes is less constrained than actual experiences, the effect of such information may be more easily altered by motivation. It is possible that reduced coding of negative errors is limited to such cases, and we do not know whether it will generalize to instances where outcomes are in fact experienced.

We previously described a positivity bias in the imagination of future life events, where participants imagined positive future events as closer in time and more vivid than negative events, a bias mediated by activity in rostral anterior cingulate cortex and amygdala<sup>25</sup>. Although participants in that study showed an optimistic bias in unconstrained imagination, here we identified an optimistic learning bias when participants' beliefs were challenged by new information. Our results provide a powerful explanatory framework for how optimistic biases are maintained.

Underestimating susceptibility to negative events can serve an adaptive function by enhancing explorative behavior and reducing stress and anxiety associated with negative expectations<sup>11,12,26</sup>. This is consistent with the observation that mild depression is related to unbiased estimation of future outcomes and severe depression to pessimistic expectations<sup>27</sup>. However, any advantage arising out

of unrealistic optimism is likely to come at a cost. For example, an unrealistic assessment of financial risk is widely seen as contributing factor to the 2008 global economic collapse<sup>28,29</sup>. Our findings suggest that this human propensity toward optimism is facilitated by the brain's failure to code errors in estimation when those call for pessimistic updates. This failure results in selective updating, which supports unrealistic optimism that is resistant to change. Dismissing undesirable errors in estimation renders us peculiarly susceptible to view the future through rose-colored glasses.

## METHODS

Methods and any associated references are available in the online version of the paper at <http://www.nature.com/natureneuroscience/>.

*Note: Supplementary information is available on the Nature Neuroscience website.*

## ACKNOWLEDGMENTS

We thank S. Fleming for assistance with analysis, P. Dayan for discussion, and T. Behrens, D. Schiller, Q. Huys, J. Winston, B. De Martino, S. Fleming, S. Bengtsson, K. Wunderlich, H. Heekeren, S. Kennerley and M. Guitart-Masip for comments on a previous version of this manuscript. This study was supported by a Wellcome Trust Program grant to R.J.D., a British Academy Postdoctoral Fellowship to T.S., and a German Academic Exchange Service (DAAD) scholarship to C.W.K.

## AUTHOR CONTRIBUTIONS

T.S. conceived the study. T.S. and C.W.K. designed the study, developed stimuli, and gathered and analyzed behavioral and fMRI data. T.S., C.W.K. and R.J.D. interpreted the data and wrote the paper.

## COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at <http://www.nature.com/natureneuroscience/>.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Weinstein, N.D. Unrealistic optimism about future life events. *J. Pers. Soc. Psychol.* **39**, 806–820 (1980).
- Baker, L.A. & Emery, R.E. When every relationship is above average: perceptions and expectations of divorce at the time of marriage. *Law Hum. Behav.* **17**, 439–450 (1993).
- Calderon, T.G. Predictive properties of analysts' forecasts of corporate earnings. *The Mid-Atlantic J. Bus.* **29**, 41–58 (1993).
- Puri, M. & Robinson, D.T. Optimism and economic choice. *J. Financ. Econ.* **86**, 71–99 (2007).
- Armor, D.A. & Taylor, S.E. When predictions fail: the dilemma of unrealistic optimism in *Heuristics and Biases: The Psychology of Intuitive Judgment* (eds. Gilovich, T., Griffin, D.W. & Kahneman, D.) 334–438 (Cambridge University Press, New York, 2002).
- Pearce, J.M. & Hall, G. A model for pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol. Rev.* **87**, 532–552 (1980).
- Sutton, R.S. & Barto, A.G. *Reinforcement Learning: An Introduction* (MIT Press, Cambridge, Massachusetts, 1998).
- Gerrard, M., Gibbons, F.X. & Reis-Bergan, M. The effect of risk communication on risk perceptions: the significance of individual differences. *J. Natl. Cancer Inst. Monogr.* **25**, 94–100 (1999).
- Weinstein, N.D. & Klein, W.M. Resistance of personal risk perceptions to debiasing interventions. *Health Psychol.* **14**, 132–140 (1995).
- Taylor, S.E. & Brown, J.D. Illusion and well-being: a social psychological perspective on mental health. *Psychol. Bull.* **103**, 193–210 (1988).
- Scheier, M.F. *et al.* Dispositional optimism and recovery from coronary artery bypass surgery: the beneficial effects on physical and psychological wellbeing. *J. Pers. Soc. Psychol.* **57**, 1024–1040 (1989).
- Taylor, S.E., Kemeny, M.E., Reed, G.M., Bower, J.E. & Gruenewald, T.L. Psychological resources, positive illusions and health. *Am. Psychol.* **55**, 99–109 (2000).
- Sharot, T. *The Optimism Bias* (Pantheon Books, New York, 2011).
- Chang, E.C. *Optimism & Pessimism: Implications for Theory, Research, and Practice* 1st edn. (American Psychological Association, 2000).
- Greening, S.G., Finger, E.E. & Mitchell, D.G.V. Parsing decision making processes in prefrontal cortex: response inhibition, overcoming learned avoidance, and reversal learning. *Neuroimage* **54**, 1432–1441 (2011).
- Yeung, N., Holroyd, C.B. & Cohen, J.D. ERP correlates of feedback and reward processing in the presence and absence of response choice. *Cereb. Cortex* **15**, 535–544 (2005).
- Cools, R., Clark, L., Owen, A.M. & Robbins, T.W. Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. *J. Neurosci.* **22**, 4563–4567 (2002).
- Mitchell, D.G.V. *et al.* Adapting to dynamic stimulus-response values: differential contributions of inferior frontal, dorsomedial, and dorsolateral regions of prefrontal cortex to decision making. *J. Neurosci.* **29**, 10827–10834 (2009).
- Taylor, S.F., Stern, E.R. & Gehring, W.J. Neural systems for error monitoring: recent findings and theoretical perspectives. *Neuroscientist* **13**, 160–172 (2007).
- Bengtsson, S.L., Lau, H.C. & Passingham, R.E. Motivation to do well enhances responses to errors and self-monitoring. *Cereb. Cortex* **19**, 797–804 (2009).
- Schultz, W., Dayan, P. & Montague, R.R. A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).
- Schultz, W. Predictive reward signal of dopamine neurons. *J. Neurophysiol.* **80**, 1–27 (1998).
- Scheier, M.F., Carver, C.S. & Bridges, M.W. Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the Life Orientation Test. *J. Pers. Soc. Psychol.* **67**, 1063–1078 (1994).
- Davidson, R.J. Hemispheric asymmetry and emotion. in *Questions About Emotions* (eds. Davidson, R.J. & Ekman, P.) 39–57 (MIT Press, Cambridge, Massachusetts, 1983).
- Sharot, T., Riccardi, A.M., Raio, C.M. & Phelps, E.A. Neural mechanisms mediating optimism bias. *Nature* **450**, 102–105 (2007).
- Varki, A. Human uniqueness and the denial of death. *Nature* **460**, 684 (2009).
- Strunk, D.R., Lopez, H. & DeRubeis, R.J. Depressive symptoms are associated with unrealistic negative predictions of future life events. *Behav. Res. Ther.* **44**, 861–882 (2006).
- Shefrin, H. How psychological pitfalls generated the global financial crisis. in *Voices of Wisdom: Understanding the Global Financial Crisis* (ed. Siegel, L.B.) *SCU Leavey School of Business Research Paper No. 10-04* <[http://www.istfin.eco.usi.ch/h\\_shefrin09.pdf](http://www.istfin.eco.usi.ch/h_shefrin09.pdf)> (Research Foundation of CFA Institute, 2010).
- Ubel, P. Human nature and the financial crisis. *Forbes* (22 February 2009).

## ONLINE METHODS

**Participants.** We recruited 20 healthy right-handed participants via the University College London psychology subject pool. Data from one participant was lost due to a computer error leaving nineteen participants (age range 19–27, 8 females). All subjects gave informed consent and were paid for their participation. The study was approved by the Research Ethics Committee of the University College London.

**Stimuli.** We presented 80 short descriptions of negative life events (for example, car theft, Parkinson's disease; see **Supplementary List of Stimuli**) in random order via a mirror mounted on the scanner head coil. Stimuli were split into two lists of 40 events each.

For each adverse event, the average probability of that event occurring at least once to a person living in the same socio-cultural environment as the participants was determined from online resources (Office for National Statistics, Eurostat, Pubmed). Very rare or very common events were not included; all event probabilities lay between 10% and 70%. To ensure that the range of possible overestimation equaled the range of possible underestimation, participants were told that the range of probabilities lay between 3% and 77%.

**Scanning task.** First, participants went through three practice trials. The session began with a short structural scan, followed by four functional runs consisting of 40 trials each (that is, all 80 events were presented twice). Finally, an additional longer structural scan was performed.

The experimental task is depicted in **Figure 1**. On each trial a stimulus was presented on screen for 4 s. Participants were instructed to think of that event happening to them in the future. After 4 s, participants were to respond in the following manner: in half of the runs (either runs 1 and 2 or runs 3 and 4, counterbalanced across participants), the words "Estimation of happening?" appeared on screen and participants were to enter their estimated likelihood of the event happening to them in the future, and in the other two runs, the words "Estimation of NOT happening?" appeared on screen and participants were to enter their estimated likelihood of the event not happening to them in the future. We framed estimations in these two ways so that differential processing of undesirable and desirable information (that is, overestimation and underestimation of the likelihood of an event) could not be attributed to differential processing of high and low numbers. When participants had already experienced an event in their lifetime they were instructed to estimate the likelihood of that event happening (or not happening) to them again in the future.

Participants had up to 6 s to respond using a button box with four buttons in each hand. Each button corresponded to one digit. The digits 0 through 7 could be used to enter the estimated likelihoods in the 'happen' estimation and digits 2 through 9 in the 'not happen' estimation. If the participant failed to respond, then that trial was excluded from all consequent analyses (mean trials with no response = 2.5, s.d. = 4.0). A fixation cross then appeared for 1–5 s (jittered). Next, the event description appeared again for 2 s, together with the average probability of that event to occur (or not occur, depending on 'happen' or 'not happen' sessions). Finally, a fixation cross appeared for 1–3 s (jittered).

Participants estimated each event twice in two consecutive sessions, which immediately followed one another. One list of 40 stimuli (counterbalanced) was presented during scan 1 and then again during scan 2. The other list was presented during scan 3 and then again during scan 4.

**Main behavioral analysis.** All statistical percentages and all responses in the 'not happen' sessions were transformed into the corresponding numbers of the 'happen' sessions by subtracting the respective number from 100. For each event in each session, an estimation error term was calculated as the absolute difference between the participant's estimate and the corresponding statistical probability presented.

$$\text{estimation error} = |\text{first estimate} - \text{probability presented}| \quad (1)$$

For each participant, trials were classified according to whether the participant initially overestimated the probability of the event relative to the average probability (and thus received desirable information; **Fig. 1b**) or underestimated (thus received undesirable information; **Fig. 1c**). Trials for which the estimation error was zero were excluded from subsequent analyses (mean = 1.5 trials, s.d. = 1.6).

We calculated the amount of update as the difference between the first and second estimates.

$$\text{update} = \text{first estimate} - \text{second estimate} \quad (2)$$

Average absolute update was compared for trials where participants received desirable and undesirable information using paired sample *t* tests across participants.

For each participant, a linear regression was conducted entering estimation errors as independent measures and updates as dependent measures. Linear regressions were then conducted (as above) separately for trials in which participants received desirable information and undesirable information. Further linear regression models with the true underlying probabilities of the event as an additional independent measure were calculated. Across subjects, betas were submitted to paired sample *t* tests and one sample *t* tests.

**Memory test and analysis.** After the scanning sessions, participants were asked to indicate the actual probability (previously presented) of each event occurring to an average person in the developed world. Memory errors were calculated as the absolute difference between the actual probability previously presented and the participant's recollection of that statistic.

$$\text{memory} = |\text{actual probability presented} - \text{recollection of probability presented}| \quad (3)$$

Memory errors for desirable and undesirable information were compared using a paired sample *t* test.

**Post-scanning questionnaire.** Participants were presented with the same trials again on a computer screen and were asked to rate events on five scales: vividness ("how vividly could you imagine this event, from 1 = not vivid to 6 = very vivid"), familiarity ("regardless if this event has happened to you before, how familiar do you feel it is to you from TV, friends, movies and so on, from 1 = not at all familiar to 6 = very familiar"), prior experience ("has this event happened to you before, from 1 = never to 6 = very often"), arousal ("when you imagine this event happening to you, how emotionally arousing is the image in your mind, from 1 = not arousing at all to 6 = very arousing") and negativity ("how negative would this event be for you, from 1 = not negative at all to 6 = very negative").

**Trait optimism (LOT-R).** Finally, participants completed the LOT-R scale<sup>23</sup> that measures trait optimism on a scale from 0 (pessimistic) to 24 (optimistic). Participants were generally optimistic; the mean optimism score was 15.68 (range = 7–21, s.d. = 3.98). We divided participants into those that scored higher than the mean (high optimism group, *n* = 11, mean = 18.54) and those that scored lower than the mean (low optimism group, *n* = 8, mean = 11.75).

**Scanning procedure.** Scanning was performed at the Wellcome Trust Centre for Neuroimaging at University College London using a 3T Siemens Allegra scanner with a Siemens head coil. Anatomical images were acquired using MPRage, which comprised 1-mm-thick axial slices parallel to the anterior commissure–posterior commissure plane. Functional images were acquired as echo-planar (EPI) T2\*-weighted images (time of repetition = 2.73 s, time of echo = 30 ms, flip angle = 90°, matrix = 64 × 64, field of view = 192 mm, slice thickness = 2 mm). A total of 42 axial slices (–30° tilt) were sampled for whole brain coverage (in-plane resolution = 3 mm × 3 mm).

**MRI data analysis.** Statistical Parametric Mapping (SPM5, Wellcome Trust Centre for Neuroimaging, <http://www.fil.ion.ucl.ac.uk/spm/>) was used for fMRI data analysis. After discarding the first six dummy volumes, images were realigned to the first volume, unwarped, normalized to a standard EPI template based on the Montreal Neurological Institute (MNI) reference brain, resampled to 2 mm × 2 mm × 2 mm voxels and spatially smoothed with an isotropic 8-mm full-width at half-maximum Gaussian kernel.

For each participant, we created a time series indicating the temporal position of event presentation, presentation of cue prompting response, motor response and presentation of information. These were modeled as time periods of 4, 0, 0 and 2 s, respectively. For all task components (except for motor response), regressors were subdivided into two conditions: trials of events for which participants first received desirable information and trials of events for which they first received undesirable information, resulting in seven regressors for each

session. Time periods were convolved with the canonical hemodynamic response function to create regressors of interest. Motion correction regressors estimated from the realignment procedure were entered as covariates of no interest.

To identify regions tracking estimation errors, we entered absolute estimation errors as parametric regressors modulating the time period of desirable and undesirable initial information presentation for each participant (Fig. 1b,c). For each condition (that is, for desirable trials and undesirable trials separately), we identified regions where parametric modulation estimates were significantly different from zero ( $P < 0.05$ , cluster level corrected; voxels first threshold at  $P < 0.001$ , uncorrected). These regions were defined as functional ROIs for subsequent analyses.

As estimation errors correlated with subsequent updates, one would expect that activity in regions tracking estimation errors would also predict subsequent update. To confirm this, we entered each participant's subsequent update on each trial as a covariate modulating the time period of initial information presentation. Estimates from this parametric modulation analysis were averaged for each participant and condition over all voxels in each ROI. The betas were then compared to zero using a one-sample  $t$  test ( $P < 0.05$ ).

To test whether trait optimism was related to how closely BOLD activity tracked desirable and undesirable estimation errors, we conducted the following analysis. Estimates from the parametric modulation analysis of estimation errors were averaged for each participant and condition over all voxels in each ROI. These betas were compared for participants in the high optimism group (those that scored above the mean on the LOT-R test of trait optimism), with those in the low optimism group (those that scored below the mean on the LOT-R) using an independent sample  $t$  test ( $P < 0.05$ ).

Furthermore, betas from all four ROIs were submitted in one linear regression analysis with optimism group membership (high or low) as the dependent measure, controlling for differential scores (desirable minus undesirable) of

memory, familiarity, past experience, vividness and arousal, and extent of initial estimation errors. We conducted a stepwise procedure to reveal the best model predicting trait optimism.

The above analysis revealed that the region tracking undesirable errors differentiated between participants who were high and those who were low on trait optimism. We tested whether activity in this region also predicted the extent of behavioral update bias across individuals. To that end, betas relating estimation errors with BOLD signal in the two peaks identified initially from the parametric analysis in this region were correlated with the behavioral update bias (desirable update minus undesirable update). To specifically isolate this relationship, we conducted a linear regression that controlled for betas from the behavioral analysis relating undesirable estimation errors and undesirable update.

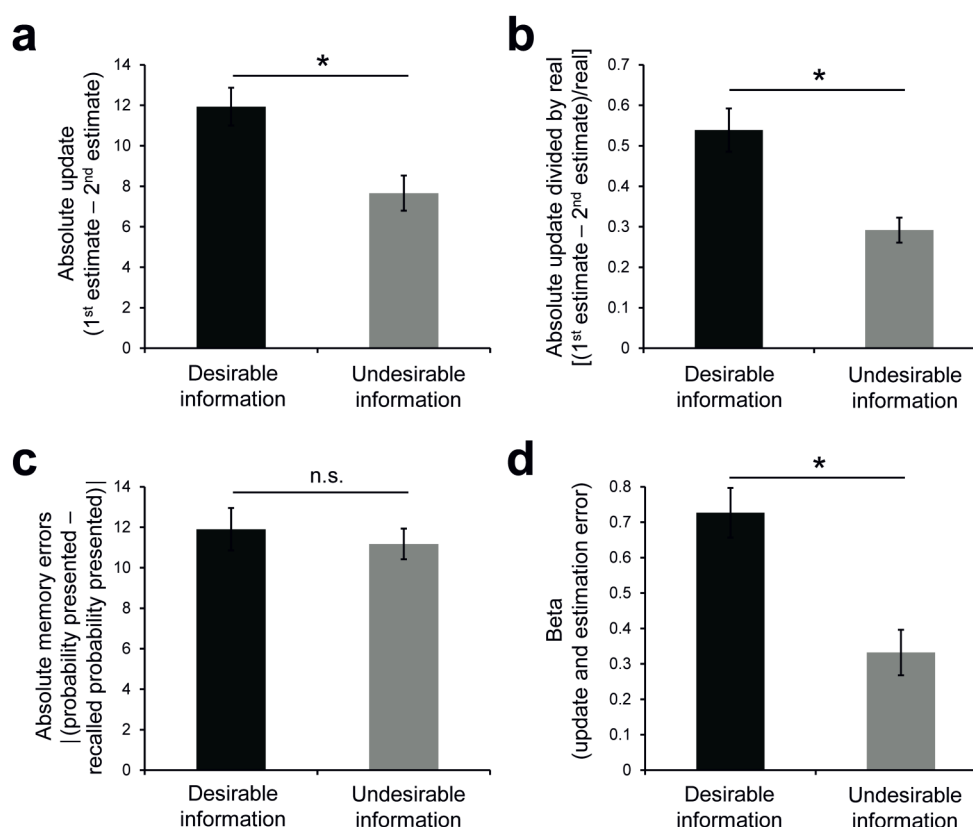
Lastly, we tested the hypothesis that greater undesirable update was related to reduced activity in the region tracking undesirable estimation errors in the second session (relative to the first). To that end, we correlated mean updates on undesirable trials across individuals with activity indicating repetition suppression (BOLD activity during second presentation of undesirable information minus BOLD activity during first presentation of that information). Note that this analysis examined the reduction in the magnitude of activity rather than the correlation of activity with a parametric regressor.

For completeness, we conducted a whole brain exploratory analysis comparing activity during the time participants thought about encountering the adverse event for desirable and undesirable trials (Supplementary Table 1). All activations are displayed on sections of the standard MNI reference brain. Anatomical labels were assigned using the Talairach daemon (University of Texas Health Science Center San Antonio; <http://www.talairach.org/>) according to peak voxels in Talairach and Tournoux coordinate space.

# How unrealistic optimism is maintained in the face of reality

Tali Sharot, Christoph W Korn & Raymond J Dolan

## Supplementary Online Material



### Supplementary Figure 1. Behaviorally observed bias.

All bars represent group means. (a) After receiving (desirable) information that presented an opportunity to adopt a more optimistic outlook, participants updated their estimations to a greater extent than after receiving (undesirable) information that called for a more pessimistic estimate. (b) This difference remained significant when updates (difference between first and second estimates) on each trial were divided by



the true probability of the respective event happening to an average person. (c) Memory performance tested post-scanning (calculated as the absolute difference between the statistical number presented for each event and the participants' recollection of that number) did not differ between desirable and undesirable information. (d) Betas indicating the association between update and estimation errors on an individual basis showed that estimation errors predicted update to a greater extent when participants received desirable information than when they received undesirable information.

Error bars (s.e.m.). \* =  $P < 0.05$ , two-tailed paired sample t-test.

**Supplementary Table 1** Regions throughout the brain that showed differential BOLD signal for different trial types during “thinking” period. Session 1& 3: contrasting activity during the time period participants were asked to think about encountering adverse events for the first time; Trials of stimuli for which participants will subsequently give an estimate lower than the average probability (i.e. will receive desirable information) versus trials of stimuli for which participants will subsequently give an estimate higher than the average probability (i.e. will receive undesirable information). Session 2& 4: contrasting activity during the time participants were asked to think about encountering adverse events for the second time; Trials of stimuli for which participants previously gave an estimate lower than the average probability (i.e. received desirable information) versus trials of stimuli for which participants gave an estimate higher than the average probability (i.e. received undesirable information) .  $P < 0.001$ , uncorrected; 10 > contiguous voxels.

Session 1&3					Session 2&4				
	X	Y	Z	R/L		X	Y	Z	R/L
Subsequent Underestimations>Subsequent Overestimations					Preivouse Underestimations>Preivouse Overestimations				
Precuneus, BA 7	28	-68	38	R	Superior Temporal Gyrus, BA 39	57	-63	22	R
Middle Temporal Gyrus, BA 21	-57	3	-12	L	Fusiform Gyrus, BA 20	-38	-40	-18	L
Precuneus, BA 31	-6	-73	26	L	Cerebellum	34	-63	-10	R
Inferior Occipital Gyrus, BA 19	44	-72	-5	R	Lingual Gyrus, BA 18	-14	-82	-8	L
Superior Frontal Gyrus, BA 11	16	56	-8	R	Parahippocampal Gyrus, BA 30	18	-46	6	R
Lingual Gyrus, BA 18	-16	-82	-9	L	Lingual Gyrus, BA 18	28	-76	-8	R
Inferior Occipital Gyrus, BA 18	-32	-84	-9	L	Middle Occipital Gyrus, BA 19	28	-82	21	R
Inferior Frontal Gyrus, BA 47	-44	23	-15	L	Cuneus, BA 19	12	-86	32	R
Lingual Gyrus, BA 19	-28	-74	0	L	Fusiform Gyrus, BA 37	-50	-55	-16	L
Inferior Frontal Gyrus, BA 46	36	33	8	R	Precuneus, BA 7	16	-44	43	R
Cerebellum	2	-52	-33	R	Inferior Occipital Gyrus, BA 18	38	-88	-4	R
Middle Occipital Gyrus, BA 19	-34	-83	4	L	Fusiform Gyrus, BA 20	34	-40	-17	R
Cerebellum	14	-80	-11	R					
Middle Temporal Gyrus, BA 19	38	-62	12	R					
Medial Frontal Gyrus, BA 8	-6	35	37	L					
Cuneus, BA 19	6	-84	34	R					
Cerebellum	-6	-71	-17	L					
Superior Temporal Gyrus, BA 22	-46	-26	-5	L					
Supramarginal Gyrus, BA 40	-55	-49	28	L					
Lateral Globus Pallidus	-20	0	-5	L					
Middle Frontal Gyrus, BA 8	-32	12	36	L					
Cuneus, BA 17	-8	-93	3	L					
Supramarginal Gyrus, BA 40	57	-45	30	R					
Paracentral Lobule, BA 5	14	-36	52	R					
Fusiform Gyrus, BA 37	-40	-59	-9	L					
Middle Temporal Gyrus, BA 21	51	-29	-2	R					
Middle Occipital Gyrus, BA 18	28	-91	8	R					
Inferior Frontal Gyrus, BA 9	36	11	23	R					
Cuneus, BA 17	14	-89	3	R					
Lingual Gyrus, BA 18	12	-54	5	R					
Subsequent Overestimations>Subsequent Underestimations					Preivouse Overestimations>Preivouse Underestimations				
					Caudate	-12	-11	19	L
					Medial Frontal Gyrus, BA 9	22	41	13	R
					Inferior Frontal Gyrus, BA 45	61	18	5	R
					Medial Frontal Gyrus, BA 9	-20	38	20	L

## Supplementary List of Stimuli

fraud when buying something on the internet  
theft from vehicle  
card fraud  
sport related accident  
household accident  
mouse/rat in house  
knee osteoarthritis (causing knee pain and swelling)  
being cheated by husband/wife  
more than £30000 debts  
miss a flight  
hernia (rupture of internal tissue wall)  
death before 80  
witness a traumatising accident  
domestic burglary  
bone fracture  
depression  
heart failure  
obesity  
irritable bowel syndrome (disorder of the gut)  
chronic high blood pressure  
diabetes (type 2)  
victim of violence by stranger  
disease of spinal cord  
serious hearing problems  
infertility  
car stolen  
dementia  
drug abuse  
gallbladder stones  
being convicted of crime  
house vandalised

restless legs syndrome  
gluten intolerance  
appendicitis  
age related blindness  
genital warts  
chronic ringing sound in ear (tinnitus)  
death before 60  
alcoholism  
Parkinson's disease  
back pain  
computer crash with loss of important data  
being fired  
eye cataract (clouding of the lens of the eye)  
skin burn  
hospital stay longer than three weeks  
bicycle theft  
divorce  
victim of bullying at work (nonphysical)  
arteries hardening (narrowing of blood vessels)  
theft from person  
having fleas/lice  
sexual dysfunction  
hepatitis A or B  
victim of violence with need to go to A&E  
severe teeth problems when old  
cancer (of digestive system/lung/prostate/breast/skin)  
abnormal heart rhythm  
victim of violence by acquaintance  
herpes  
migraine  
having a stroke  
victim of violence at home  
severe insomnia

osteoporosis (reduced bone density)  
death before 70  
severe injury due to accident (traffic or house)  
autoimmune disease  
artificial joint  
victim of mugging  
asthma  
blood clot in vein  
ulcer  
kidney stones  
Alzheimer's disease  
anxiety disorder  
limb amputation  
epilepsy  
liver disease  
death by infection  
*Events used during the training sessions*  
dying before 90  
glaucoma  
post-traumatic stress disorder

**Supplementary Behavioural Study I – Testing memory after a single presentation of the actual probabilities**

Rationale: In our fMRI study (reported in the main text) participants' memory for the information presented did not differ for desirable and undesirable information. We concluded that the selective updating observed in response to desirable and undesirable information was not driven by differential memory. A limitation here was that participants were presented twice with the actual likelihoods before their memory was tested. It is possible that after the first presentation of the statistical likelihoods (before the second update) participants' memory was biased, but not after the second presentation of the likelihoods.

Method: To exclude this possibility we tested ten additional subjects (age range 20–36, 7 females) on a similar behavioural task as utilized in the fMRI study with one critical difference – participants were not presented with the statistical probabilities in the second session. For simplicity we only asked participants to estimate their likelihood of encountering the adverse events (rather than estimate their likelihood of not encountering the adverse events) as our previous results showed no effect of how the question was worded. Analysis was conducted as before.

Results: First, we replicated our main behavioural results, showing greater update on trials when participants received desirable information relative to undesirable information ( $t(9) = 3; P < 0.02$ ). Importantly, there was no difference in participants' memory for desirable and undesirable information after one presentation of the likelihoods ( $t(9) = 0.3, P > 0.7$ ), suggesting that selective update was not driven by differential memory for desirable and undesirable information.

## **Supplementary Behavioural Study II – Testing memory before the second estimation of personal probabilities**

Rationale: A limitation to the design employed above is that it had order confounding; memory was tested after receiving the second estimate from the participants. We thus tested whether the same pattern of results would emerge (i.e. a valence effect on update but not on memory) if memory was tested before second estimates were elicited.

Method: We tested twenty additional subjects (age range 20–30, 13 females) on a similar behavioural task as utilized in the study above with one critical difference – after completing the first session participants received the memory test and only then they were asked to estimate all probabilities again. For simplicity half of the participants were asked to estimate their likelihood of encountering the adverse events and half were asked to estimate their likelihood of not encountering the adverse event. Analysis was conducted as before.

Results: First, we replicated our main behavioural results, showing greater update on trials when participants received desirable information relative to undesirable information ( $t(19) = 2.4$ ;  $P < 0.05$ ). Importantly, there was no difference in participants' memory for desirable and undesirable information ( $t(19) = 0.99$ ,  $P > 0.3$ ), suggesting that the results reported in the main text were not due to memory being tested after the second estimates were elicited.

### **Supplementary Results**

To control for the possibility that differential updating following desirable and undesirable information could be explained by differential processing of high and low numbers, participants were asked to estimate their likelihood of encountering the

adverse event on half the trials, and to estimate their likelihood of not encountering the adverse event on the rest of the trials. A 2 by 2 ANOVA (wording of task: event happening / event not happening X information provided: desirable / undesirable) conducted on amount of update, did not reveal a significant interaction ( $F(1, 18) = 0.208, P > 0.6$ ). Furthermore, the wording of the task did not affect any of the fMRI results reported in the main text.

Post-scanning questionnaire scores showed that participants found events for which they received desirable information as emotionally arousing and as negative as the events for which they received undesirable information ( $t(18) = -0.3, P > 0.7$  and  $t(18) = 1.3, P > 0.2$ , respectively). Events for which participants received desirable information were imagined more vividly ( $t(18) = -2.3, P < 0.05$ ), were more familiar ( $t(18) = -2.2, P < 0.05$ ) and tended to be experienced more often in the past ( $t(18) = -2.1, P < 0.01$ ) than events for which participants received undesirable information. In other words, events for which a change in update was more likely were more familiar indicating more, rather than less, prior knowledge. Importantly, these factors could not explain differential updating; the difference in absolute update for events for which participants received desirable and undesirable information remained significant even after entering all post-scanning scores as covariates ( $F(1, 13) = 9.7, P < 0.01$ ).

As mentioned in the main text, the difference in updating for events for which participants received desirable and undesirable information remained significant even after entering the difference in true probabilities of the events as covariates ( $F(1, 17) = 6.04, P < 0.05$ ). Entering the true probabilities of the events as a regressor in the fMRI analysis did not reveal any differential effects for desirable and undesirable trials in any of the regions reported in the text.