Article:

Garrett, N., González-Garzón, A.M., Foulkes, L. et al. (2 more authors) (2018) Updating beliefs under perceived threat. Journal of Neuroscience, 38 (36). pp. 7901-7911. ISSN 0270-6474

https://doi.org/10.1523/JNEUROSCI.0716-18.2018

JNeurosci THE JOURNAL OF NEUROSCIENCE

Research Articles: Behavioral/Cognitive

Updating Beliefs Under Perceived Threat

Neil Garrett[a,c], Ana María González-Garzón[a], Lucy Foulkes[a], Liat Levita[b] and Tali Sharot[a]

[a] *Affective Brain Lab, Experimental Psychology, University College London, WC1H 0AP*
[b] *Liat Levita, Department of Psychology, University of Sheffield, S10 2TP*
[c] *Princeton Neuroscience Institute, Princeton University, Princeton, New Jersey*

DOI: 10.1523/JNEUROSCI.0716-18.2018

Received: 19 March 2018

Revised: 20 May 2018

Accepted: 15 June 2018

Published: 6 August 2018

Author contributions: N.G., A.M.G.-G., and T.S. designed research; N.G., A.M.G.-G., and L.F. performed research; N.G., A.M.G.-G., L.F., and T.S. analyzed data; N.G. and T.S. wrote the first draft of the paper; N.G. and T.S. edited the paper; N.G. and T.S. wrote the paper; L.L. contributed unpublished reagents/analytic tools.

Conflict of Interest: The authors declare no competing financial interests.

We thank Einar Jensen and the South Metro Fire Rescue Authority for providing resources and assistance for Experiment II; Jon Roiser, Oliver Robinson, Stephanie Lazzaro, Andreas Kappes, Caroline Charpentier and Sebastian Bobadilla-Suarez for helpful comments on earlier editions of this manuscript; Dominik Bach for advice on SCR collection and analysis.

This research was supported by a Wellcome Trust Career Development Fellowship to T. Sharot and a UCL Impact Award to N. Garrett.

Corresponding authors: Neil Garrett: Email: ngarrett@princeton.edu; Tali Sharot, Email: t.sharot@ucl.ac.uk, Affective Brain Lab, Experimental Psychology, University College London, WC1H 0AP.

Cite as: J. Neurosci ; 10.1523/JNEUROSCI.0716-18.2018

Updating Beliefs Under Perceived Threat

Authors: Neil Garrett[a,c*], Ana María González-Garzón[a], Lucy Foulkes[a], Liat Levita[b] & Tali Sharot[a*]

Affiliations: [a]Affective Brain Lab, Experimental Psychology, University College London, WC1H 0AP; [b]Liat Levita, Department of Psychology, University of Sheffield, S10 2TP; [c]Princeton Neuroscience Institute, Princeton University, Princeton, New Jersey, 08540.

***Corresponding authors:** Neil Garrett: Email: ngarrett@princeton.edu; Tali Sharot, Email: t.sharot@ucl.ac.uk, Affective Brain Lab, Experimental Psychology, University College London, WC1H 0AP.

Number of Pages: 34
Number of Figures: 5
Number of Tables: 4
Number of words: Abstract (156), Introduction (617), Discussion (905)

Acknowledgements: We thank Einar Jensen and the South Metro Fire Rescue Authority for providing resources and assistance for Experiment II; Jon Roiser, Oliver Robinson, Stephanie Lazzaro, Andreas Kappes, Caroline Charpentier and Sebastian Bobadilla-Suarez for helpful comments on earlier editions of this manuscript; Dominik Bach for advice on SCR collection and analysis.

Competing Financial Interests Statement: The authors declare no competing financial interests with respect to their authorship or the publication of this article.

Funding: This research was supported by a Wellcome Trust Career Development Fellowship to T. Sharot and a UCL Impact Award to N. Garrett.

Abstract

Humans are better at integrating desirable information into their beliefs than undesirable. This asymmetry poses an evolutionary puzzle, as it can lead to an underestimation of risk and thus failure to take precautionary action. Here, we suggest a mechanism that can speak to this conundrum. In particular, we show that the bias vanishes in response to perceived threat in the environment. We report that an improvement in participants' tendency to incorporate bad news into their beliefs is associated with physiological arousal in response to threat indexed by galvanic skin response and self-reported anxiety. This pattern of results was observed in a controlled laboratory setting (Experiment I), where perceived threat was manipulated, and in firefighters on duty (Experiment II), where it naturally varied. Such flexibility in how individuals integrate information may enhance the likelihood of responding to warnings with caution in environments rife with threat, while maintaining a positivity bias otherwise, a strategy that can increase well-being.

Significance Statement

The human tendency to be overly optimistic has mystified scholars and lay people for decades: how could biased beliefs have been selected for over unbiased beliefs? Scholars have suggested that while the optimism bias can lead to negative outcomes, including financial collapse and war, it can also facilitate health and productivity. Here, we demonstrate that a mechanism generating the optimism bias, namely asymmetric information integration, evaporates under threat. Such flexibility could result in enhanced caution in dangerous environments while supporting an optimism bias otherwise, potentially increasing well-being.

Introduction

Whether a piece of news is good or bad is critical in determining whether it will alter our beliefs. In particular, people readily incorporate favorable news into their existing beliefs, yet tend to underweight the strength of unfavorable information (Eil and Rao, 2011; Kuzmanovic and Rigoux, 2017; Kuzmanovic et al., 2015, 2016; Lefebvre et al., 2017; Mobius et al., 2012; Sharot et al., 2011; Wiswall and Zafar, 2015). For example, when learning that their risk of experiencing future aversive events, such as robbery, is higher than they had expected, people are less likely to integrate these data into prior beliefs relative to a situation in which they learn that their risk is lower than expected (Sharot et al., 2011). The same pattern emerges when people receive desirable and undesirable information about their financial prospects (Wiswall and Zafar, 2015), or feedback about their intellectual abilities (Eil and Rao, 2011; Mobius et al., 2012), personality (Korn et al., 2012) and physical traits (Eil and Rao, 2011). This is known as a **valence-dependent learning asymmetry** (Sharot and Garrett, 2016).

Incorporating desirable information about the self at a higher rate than undesirable (Korn et al., 2012) will subsequently lead to overconfidence and optimistically biased predictions (Sharot et al., 2011). On the upside an optimistic outlook, even when biased, can improve physical and mental health (Taylor and Brown, 1988), boost motivation (Bandura, 1989), exploration (Tiger, 1979) and persistence (Sherman, 1980), thus enhancing success and well-being (for a review, see (Chang, 2001). However, ignoring negative information can result in faulty assessment and lack of precautionary action leading to, for example, ill preparedness in the face of natural disasters, and financial market bubbles (Shefrin, 2009) .

These apparent costs present a conundrum; why have humans evolved a bias in learning that leads to systematic errors in judgement? The common answer is that people make errors that are costly in certain situations, because those errors are advantageous in other situations, and on balance the benefits outweigh the costs (McKay and Dennett, 2010). There is another possibility though - that the asymmetry fluctuates in response to environmental demands. For example, in relatively safe surroundings, where potential harm is low, an asymmetry in information integration may be prominent leading to biased expectations. Yet in environments rife with threats, a physiological/psychological response may trigger changes to how information is integrated leading to more balanced information integration which may be adaptive in environments where potential costs are high (see Johnson and Fowler, 2011).

Because affect provides an internal signal about the external context, it could potentially be used to adaptively modulate cognitive biases. Specifically, we suggest that the key is a learning mechanism that is modulated by the two core aspects of affect: valence and arousal.

A valence-dependent learning mechanism biases judgements and an arousal-dependent switch controls the degree and perhaps sign of the bias.

To test this prediction, we exposed participants to an acute threat manipulation in the lab (Experiment I) or tested participants in a real-life environment (firefighters tested on call, Experiment II). After measuring indicators of arousal, stress and anxiety, participants completed the belief update task (Chowdhury et al., 2014; Garrett and Sharot, 2014; Garrett et al., 2014; Kappes et al., 2018; Korn et al., 2013; Kuzmanovic et al., 2015, 2016; Moutsiana et al., 2013, 2015; Sharot et al., 2011, 2012a, 2012b) (**Fig. 1**). Past studies have shown that participants put more weight on good news (i.e. that a negative life event is less likely to occur than expected, **Fig. 1a**) compared to bad news (i.e. that a negative event is more likely to occur than expected, **Fig. 1b**) in altering beliefs in this task. Here we test whether heightened response to threat abolishes this bias.

Materials & Methods

Experimental Design and Statistical Analysis: Experiment I

Participants. Thirty-six participants recruited via the UCL participant pool participated in the study. Participants gave informed consent and were paid for their participation. The study was approved by the Research Ethics Committee of the University College London. One participant's responses resulted in only two good news trials (out of a possible 40), which prevented us from calculating a meaningful information integration parameter (we define how we calculate information integration parameters below), thus this participant's data had to be excluded. Two participant's cortisol samples were insufficient for analysis, and samples of six participants who were suspected to have depression (BDI score greater than 10) were never sent to be analyzed. Thus, analysis that includes cortisol scores is given for n = 27. Note, however, that either excluding those participants all together from all analysis or including them as done here generated similar results. Each participant was randomly assigned to either the threat manipulation condition (13 females, 6 males, mean age = 26.37 years, SD = 6.58) or the control condition (10 females, 6 males, mean age = 24.94 years, SD = 3.82).

Manipulation Procedure. We designed the experiment such that the perceived threat was unrelated to the information presented in the task. Thus, we could test whether the effect of perceived threat on information integration was general rather than specific to the source of the threat itself.

Participants assigned to the threat manipulation group were told that they would be exposed to an uncomfortable, stressful, event at the end of the study. Specifically, they were informed that at the end of the experiment they would be required to deliver a speech on a surprise topic, which would be recorded on video and judged live by a panel of staff members. They were shown an adjacent room across a double mirror window where chairs and tables were already organized for the panel. In addition, participants were presented with six difficult mathematical problems which they were asked to try and solve in 30 seconds. This manipulation is a variation of the Trier Social Stress Test (TSST) (Birkett, 2011) with the main difference between the typical TSST procedure and the one used here being that participants were threatened by the possibility of a stressful social event, and completed the main task under threat, but the threat was never executed. Having the participants believe the stressful event will take place at the end of the task, rather than before, increased the likelihood that participants' arousal levels remained high throughout the task. Participants assigned to the control condition were informed that at the end of the experiment they would be required to write a short essay on a surprise topic, which would not be judged. They were then presented with six elementary mathematical problems to solve in 30 seconds.

Manipulation Check. We examine if the threat manipulation resulted in the following psychological and physiological changes, which are typically observed in studies using variations of TSST (Birkett, 2011).

1. *Self-Report*. Before and after the induction procedure participants filled out a short-form of the State scale of the Spielberger State Trait Anxiety Inventory developed by Marteau and Bekker (Marteau and Bekker, 1992). Participants reported their current anxiety state according to 6 statements (e.g. I am worried) on a 4-point Likert scale (1 = *not at all* to 4 = *very much*). Possible scores range from 6 to 24 with high scores indicating high levels of state anxiety.
2. *Skin Conductance Level (SCL).* SCL is an index of sympathetic tone which reflects changes in autonomic arousal. Skin conductance was recorded for 2 minutes pre- and post-induction whilst participants stared at a fixation cross using disposable electrodermal gel electrodes (Biopac, EL507) attached to the distal phalanx of the pointer and middle fingers of the participants' non-dominant hand. Skin conductance responses were monitored using a MP36R system (BIOPAC Systems, Inc., Goleta, CA) and analyzed with BIOPAC software Acq*Knowledge.* The difference in mean SCL in each period were taken as a change in participants' autonomic arousal levels.
3. *Cortisol Level.* To measure changes in participants' cortisol levels, saliva samples were collected using Salivette collection devices, (Salimetrics, UK). Four samples

were taken at different time points: before the induction procedure (baseline: t0); immediately after the induction procedure but prior to undertaking the task (10 min after the threat/control manipulation: t1); halfway through the task (30min after the threat/control manipulation: t2); after the task and completion of post experiment questionnaires (+1hr after the threat/control manipulation: t3). The experiment was conducted between 2pm and 4pm, restricted to these times to control for the diurnal cycle of cortisol. Samples were stored at -80°C before being assayed. Analysis of salivary cortisol was completed by Salimetrics. Intra-assay and inter-assay coefficients of variation were all below 6.1% (*M* = 1.5%, *SD* = 1.2). Cortisol values were measured in µg/dL. Shapiro-Wilk (SW) tests on cortisol levels at each sample period revealed that these were not normally distributed (one sample SW < .01 for all four sample intervals). As a result, cortisol values were log transformed. Since cortisol stress response has a temporal delay (mediated by the slower time scale HPA axis), it is difficult to precisely align the time of the cortisol response to perceived levels of threat at different points in the task. Because of this, the main cortisol measure we use in the manuscript was calculate as the mean difference between cortisol levels at time periods t1, t2 and t3 from baseline cortisol levels at t0, as done previously (Lenow et al., 2017; Lighthall et al., 2013; Otto et al., 2013). This measure represents the average cortisol response throughout the duration of task performance. Below is the formula we used to derive this index where log cort is the natural log-transformed cortisol concentrations:

$$\log cort\ \Delta = \frac{\log cort_{t1} + \log cort_{t2} + \log cort_{t3}}{3} - \log cort_{t0}$$

Behavioral Task. The task was adopted from past studies (Chowdhury et al., 2014; Garrett and Sharot, 2014; Garrett et al., 2014; Korn et al., 2013; Moutsiana et al., 2013, 2015; Sharot et al., 2011, 2012a, 2012b).

Stimuli. Stimuli (80 short descriptions of different negative life events, for example: domestic burglary, card fraud) were separated into two lists, each containing 40 events. Participants were randomly assigned one of the two lists of 40 evens at the start of the experiment. For each event the average probability of that event occurring at least once to someone from the UK within the same age range as the participants was calculated from data compiled from online resources (including the Office for National Statistics and 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 was equal to the range of possible underestimation, participants were told that the range of probabilities lay between 3% and 77% and they were only permitted to enter estimates within this range. Note that differences between the average probabilities provided to participants and the actual probabilities for the sample of participants tested cannot explain differences between the two groups, as we randomly assign participants to either the threat manipulation condition of the control condition.

Behavioral Task (Fig. 1). Participants completed a practice session comprising 3 trials before beginning the main experiment. The main experiment comprised 40 trials. On each trial one of 40 adverse life events were presented for 3s, and participants were asked to estimate how likely the event was to happen to them in the future. Participants had up to 5s to respond. If participants had already experienced an event in their lifetime they were instructed to estimate the likelihood of that event happening to them again in the future. If the participant failed to respond, that trial was excluded from all subsequent analyses (M = 1.31, SD = 1.39). Following presentation of a fixation cross (5-10s jittered) participants were then presented with the base rate of the event in a demographically similar population for 2s followed by a fixation cross (5-10s jittered). In a second session, immediately after the first, participants were asked again to provide estimates of their likelihood of encountering the same events so that we could assess how they updated their estimate in response to the information presented.

Note, that studies have shown that the update bias exists both when classifying trials according to participants' estimates of self-risk and when trials are classified according to estimates of base rates (Garrett and Sharot, 2014; Kuzmanovic et al., 2015). Thus, we used the traditional design and analysis here (Sharot et al., 2011) . Moreover, multiple past studies have shown that the amount of update bias does not alter whether participants are asked to estimate the likelihood of the event happening in the future or the likelihood of the event not happening in the future (Garrett and Sharot, 2014; Garrett et al., 2014; Sharot et al., 2011). Thus, scores are not driven by response to high and low numbers, but rather by valence per se. As this has been established in the past we used the standard version of the task here (i.e. eliciting estimation of an event happening).

Memory control. To test for memory effects participants were asked at the end of the experiment to provide the actual probability previously presented of each event. Memory errors were calculated as the absolute difference between the probability previously presented and the participants' recollection of that statistic:

***Memory Error* = | *Probability Presented* − *Recollection of Probability Presented* |**

Other controls. At the end of experiment, participants also rated stimuli on 6-point scales for vividness [for the question "How vividly could you imagine this event?" (1 = *not at all vivid* to 6 = *very vividly*)], familiarity [for the question "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?" (1 = *not at all familiar* to 6 = *very familiar*)], prior experience [for the question "Has this event happened to you before?" (1 = *never* to 6 = *very often*)], emotional arousal [for the question "When you imagine this event, how emotionally arousing do you find the image in your mind?" (1 = *not at all arousing* to 6 = *very arousing*)] and negativity [for the question "How negative would this event be/is this event for you?" 1 = *not negative at all* to 6 = *very negative*)].

Statistical analysis. Trials were partitioned according to participants' first estimates into ones in which participants received good news [i.e., the probability presented was lower than the first estimate of their own probability (**Fig. 1a**)] or bad news [i.e., the probability presented was higher (**Fig. 1b**)]. While information can be better or worse than expected, all stimuli are negative (i.e. robbery, card fraud), thus comparison is never between positive and negative stimuli, but between information that is better or worse than expected.

Trials for which the estimation error was zero were excluded from subsequent analyses as these could not be categorized into either condition (*M* = 0.89 trials, *SD* = 0.92).

For each trial an estimation error term was calculated as the difference between the probability presented and participants' first estimate on that trial:

Estimation Error = Probability Presented - First Estimate

Update was calculated for each trial such that positive updates indicate a change toward the probability presented and negative updates a change away from the probability presented:

Update (Good News) = First Estimate – Second Estimate
Update (Bad News) = Second Estimate – First Estimate

Formal models suggest that learning from information that disconfirms one's expectations is mediated by a prediction error signal that quantifies a difference between expectation and outcome (Sutton and Barto, 1998). We have previously shown that an analogous mechanism underpins belief updating in this task (Sharot et al., 2011). Specifically, the difference between participants' initial estimations and the information provided (that is, estimation error = probability presented – first estimate) predicts subsequent updates, as would be expected from learning models (Sutton and Barto, 1998). Hence, similar to our previous papers (Garrett et al., 2014; Moutsiana et al., 2013; Sharot et al., 2011), we estimated the extent to which participants integrated new information into their beliefs by correlating estimation errors and update scores with one another separately for good and bad news trials for each participant. This resulted in two pearson correlation values for each participant: one for good news trials and one for bad news trials. We denote these Pearson correlation scores as good news (α_G) and bad news (α_B) information integration parameters. Shapiro-Wilk tests were applied to check the values of α_G and α_B were normally distributed. To check the values of α_G and α_B were not at floor or ceiling, we conducted one sample t-tests (separately α_G and α_B) against values of 0 (to test for floor effects) and 1 (to test for ceiling effects).

To determine whether information integration from good and/or bad news was altered by the threat manipulation, the resulting information integration parameters were submitted to a 2 by 2 ANOVA with valence (good/bad news) as a repeated-measure and group (threat manipulation/control) as a between-subjects factor.

We identified possible confounds to add as covariates to our analysis as follows; first, for factors that were not task related and therefore did not have a valence component (specifically: initial self-reported anxiety, initial SCL, initial cortisol and BDI) we conducted independent sample t-tests (control vs threat manipulation group) for each factor separately to determine if a group difference existed (**Table 1**). For task related variables that could be divided by valence (specifically; number of trials, memory scores, ratings on familiarity, vividness, past experience, negativity, emotional arousal and mean first estimates) we calculated the difference between mean good news and mean bad news for each participant for each of these factors. This gives a bias score for each factor for each subject whereby positive scores indicate a bias towards good news and negative scores indicate a bias towards bad news. We then conducted a one sample t-test (versus 0) on each of these scores for each group separately to isolate those factors which had valence effects in either set of participants. Next we conducted a series of independent sample t-tests to compare the control groups difference scores to the threat manipulation groups scores for each factor (this is equivalent to testing for an interaction between valence and group). For all of these tests we applied a threshold of $p<0.05$ and deliberately did not correct for multiple comparisons. This is because the purpose was to identify all potential confounds; by not correcting we are being more stringent. Any factor which showed a group effect or a valence effect was added as a covariate. These were: mean first estimates, ratings of vividness, familiarity, past experience and emotional arousal (**Table 1**).

To explore whether differences in information integration related to any of the specific physiological and psychological changes, we constructed a general linear model (GLM) with α entered as the dependent variable and changes in SCL, self-report anxiety and cortisol as independent variables. This was done separately for information integration parameters for good (α_G) and bad (α_B) news. To control for general changes in information integration and allow us to detect valence-specific effects, we entered information integration parameters for good news (α_G) as a covariate when estimating information integration parameters for bad news (α_B) and vice versa (Moutsiana et al., 2013). In addition, following the same selection procedure outlined above we controlled for any variable where there was a significant ($p<0.05$) difference between groups, between types of information (i.e. valence) or a group*valence interaction, by including these in the GLM as covariates.

For α_B the formula for the regression in full therefore is as follows:

α_B = β0 + β1*Change in SCL + β2*Change in Self-report + β3*Change in Cortisol + β4*Mean Initial Estimate + β5*Initial Self-report anxiety+ β6*Mean Bad News Vividness Rating + β7*Mean Bad News Familiarity Rating + β8*Mean Prior Experience Bad News Rating + β9*Mean Emotional Arousal Bad News Rating + β10*α_G

For (α_G) the formula for this was as follows:

α_G = β0 + β1*Change in SCL + β2*Change in Self-report + β3*Change in Cortisol + β4*Mean Initial Estimate + β5*Initial Self-report anxiety + β6*Mean Good News Vividness Rating + β7*Mean Good News Familiarity Rating + β8*Mean Prior Experience Good News Rating + β9*Mean Emotional Arousal Good News Rating + β10*α_B

Finally, we reran the analysis above this time controlling for within-subject covariates at the within-subject level and between-subject factors at the between-subject level. Specifically, for each participant we computed an alternative set of information integration parameters – one for good news ($\alpha_{G\_partial}$) and one for bad news ($\alpha_{B\_partial}$) - by carrying out a series of partial correlations in which absolute estimation error and update were the two variables of interest. Within-subject covariates - identified as above (first estimate, vividness, familiarity, past experience and emotional arousal) - were controlled for on a trial by trial basis. We examined whether these alternative information integration parameters for bad news ($\alpha_{B\_partial}$) related to change in self report and/or change in SCL controlling for any additional between subject confounds as above (initial self-report anxiety ratings and information integration for good news). This was done by entering alternative information integration parameters for bad news ($\alpha_{B\_partial}$) as the dependent variable into 2 GLMs as follows:

$$\alpha_{B\_partial} = \beta0 + \beta1\text{*Change in Self-report} + \beta2\text{*Initial Self-report anxiety} + \beta3\text{*}\alpha_{G\_partial}$$

$$\alpha_{B\_partial} = \beta0 + \beta1\text{*Change in SCL} + \beta2\text{*Initial Self-report} + \beta3\text{*}\alpha_{G\_partial}$$

We then examined the significance of the regression weights in each GLM for change in Self Report and change in SCL. To visualize the effect of each of these (Fig. 4) we generated two partial regression plots. These are scatterplots of the residuals of the dependent variable ($\alpha_{B\_partial}$) and the independent variable (either Change in Self-report or Change in SCL) when these are regressed on the rest of the independent variables (Initial Self report and $\alpha_{G\_partial}$).

We ran the equivalent analysis for good news ($\alpha_{G\_partial}$) as follows:

$$\alpha_{G\_partial} = \beta0 + \beta1\text{*Change in Self-report} + \beta2\text{*Initial Self-report} + \beta3\text{*}\alpha_{B\_partial}$$

$$\alpha_{G\_partial} = \beta0 + \beta1\text{*Change in SCL} + \beta2\text{*Initial Self-report} + \beta3\text{*}\alpha_{B\_partial}$$

Experimental Design and Statistical Analysis: Experiment II

Participants. Thirty-three operational staff stationed across seventeen fire stations within the South Metro Fire and Rescue Authority of the State of Colorado in the United States participated in the study. Five of these participants failed to complete the study leaving 28 participants (1 female, 27 males, mean age = 43.15 years, *SD* = 9.87). A link to an online version of the experiment was sent by email to operational staff inviting them to participate in the study whilst on duty. Employees were given 18 days to attempt the experiment. They were permitted to take the experiment once in this time period and were explicitly requested to do so whilst on shift (i.e. in the station between calls). Participation in the experiment was anonymous, voluntary and unpaid.

Task, stimuli and control variables. An online version of the task used in Experiment I was designed using Qualtrics Survey Software (Qualtrics, Provo, UT). The task began by asking basic demographic questions (age, gender, marital status, level of education and number of children) and some questions pertaining to their work (including how long they had worked in the service, how many people they supervised, number of emergency they went on, what their rank in the service was) and social environment (social support at work and outside, and stress experienced at home).

After providing this information, participants read task instructions on screen at their own pace and then undertook a practice session comprising 3 practice trials. As in Experiment I, stimuli (80 short descriptions of different negative life events; the majority of these were the same as those used in Experiment I but 18 events were exchanged with alternative negative life events) were separated into two lists, each containing 40 negative life events. Participants were randomly assigned one of the two lists of 40 events at the start of the experiment. The task was the same as in Experiment I, except that there was only one fixation cross displayed in each session (for 1s) after participants submitted estimates (i.e. in the first session, unlike in Experiment I, a second fixation cross was not displayed after base rate presentation). Furthermore, mindful of the firefighters' unpredictable time constraints, memory for the information given and subjective ratings (past experience with the event and negativity) were elicited for half the stimuli and participants completed a short version of the state scale of the self-report at the beginning of the study (Chlan et al., 2003), without providing physiological measures of autonomic arousal.

Statistical analysis: Linear regressions were performed using ordinary least squares implemented using SPSS version 25 for bad news and good news separately, with α entered as the dependent variable and self-reported state anxiety as the independent variable. To rule out potential confounds we followed a similar procedure as in Experiment I. Specifically, we separately tested whether a range of potential confounding factors had valence effects. These factors were: mean first estimates, memory scores, ratings of negativity, ratings of past experience and number of trials. We did this by calculating the difference between mean good news and mean bad news for each participant for each of these factors. This gives a bias score for each factor for each subject whereby positive scores indicate a bias towards good news and negative scores indicate a bias towards bad news. We then conducted a one sample t-test (versus 0) on each of these scores to identify factors which had valence effects. We used a threshold of $p<0.05$ and deliberately did not correct for multiple comparisons. This is because the purpose was to identify all potential confounds; by not correcting we are being more stringent. Any factor which showed a valence effect was then added as a covariate. These were mean first estimates, ratings of past experience and number of trials (**Table 3**).

To test for a relationship between anxiety and the asymmetry within the firefighters (i.e. preferential updating for bad news over good) we calculated an information integration bias score for each participant. This is simply the difference between α_G and α_B. A score of 0 indicates no bias in information integration in either direction whilst positive scores indicate greater information integration for good news relative to bad news and negative scores the opposite. We then examined whether the information integration bias related to self-reported anxiety as follows:

Information Integration Bias Score (α_G - α_B) = β0 + β1*Self-Reported Anxiety + β2*Mean Initial Estimate + β3*Mean Prior Experience Bias Score (Mean Prior Experience Bad News Rating - Mean Prior Experience Good News Rating) + β4*Number of Trials Bias Score (Number of Good News Trials - Number of Bad News Trials)

Next we ran a GLM for each of the two sets of information integration parameters (α_G and α_B) separately. To ensure effects were valence specific rather than reflecting general changes in information integration, good news (α_G) was also added as a covariate when examining information integration parameters for bad news (α_B) and vice versa when examining information integration for good news.

For bad news information integration parameter (α_B), the formula for the regression in full therefore is as follows:

α_B = β0 + β1*Self-Reported Anxiety + β2*Mean Initial Estimate + β3*Mean Prior Experience Bad News Rating + β4*Number of Bad News Trials + β5*α_G

For good news information integration parameter (α_G), the formula for the regression in full therefore is as follows:

α_G = β0 + β1*Self-Reported Anxiety + β2*Mean Initial Estimate + β3*Mean Prior Experience Good News Rating + β4*Number of Good News Trials + β5*α_B

Finally, we reran the analysis above this time controlling for within-subject covariates at the within-subject level and between-subject factors at the between-subject level. Specifically, for each participant we computed an alternative set of information integration parameters – one for good news ($\alpha_{G\_partial}$) and one for bad news ($\alpha_{B\_partial}$) - by carrying out a series of partial correlations in which absolute estimation error and update were the two variables of interest. Within-subject covariates (first estimates), were controlled for on a trial by trial basis (note it was not possible to control for past experience on a trial by trial basis here because participants in this study completed ratings only for a subset of events). We then examined whether these alternative information integration parameters for bad news ($\alpha_{B\_partial}$) related to self-reported anxiety, controlling for additional between subject covariates (number of bad news trials and information integration for good news) at the between subject level. This was done by entering alternative information integration parameters for bad news ($\alpha_{B\_partial}$) as the dependent variable into a GLM as follows:

$$\alpha_{B\_partial} = \beta 0 + \beta 1 * \text{Self-Reported Anxiety} + \beta 2 * \text{Number of Bad News Trials} + \beta 3 * \alpha_{G\_partial}$$

We then examined the significance of the regression weight for Self-Reported Anxiety.

We ran the same analysis for information integration parameters for good news ($\alpha_{G\_partial}$) as follows:

$$\alpha_{G\_partial} = \beta 0 + \beta 1 * \text{Self-Reported Anxiety} + \beta 2 * \text{Number of Good News Trials} + \beta 3 * \alpha_{B\_partial}$$

To visualize the effect of each of these (Fig. 5) we generated two partial regression plots. These are scatterplots of the residuals of the dependent variable ($\alpha_{B\_partial}$ or $\alpha_{G\_partial}$) and the independent variable of interest (Self-Reported Anxiety) when these are regressed on the rest of the independent variables (Number of Bad News Trials and $\alpha_{G\_partial}$ when examining $\alpha_{B\_partial}$, Number of Good News Trials and $\alpha_{B\_partial}$ when examining $\alpha_{G\_partial}$).

Results

Experiment I

Threat manipulation was successful. Subjective self-reports of anxiety and physiological measures of skin conductance level (SCL) and cortisol showed that the manipulation was effective. Specifically, following the manipulation, self-report anxiety (**Fig. 2a**) and SCL (**Fig. 2b**) showed an increase relative to before (baseline), which was greater in the threat manipulation group relative to controls (self-reported anxiety: $t(33) = 4.16$, $p < .001$; SCL: $t(33) = 3.32$, $p = .002$, independent sample t-test). There were no baseline (t0) differences in cortisol levels between the two groups ($t(25) = -.89$, p = 0.38). Mean cortisol levels (averaged across t1, t2 and t3) relative to baseline (t0) showed a trend towards being higher in the threat manipulation group relative to controls ($t(25) = 1.90$, $p = .07$). This effect was driven by a reduction in cortisol levels over time in the control group (main effect of time at t1, t2 and t3 relative to baseline: $F(2,26) = 17.19$, $p < .001$, repeated measures ANOVA) - an effect previously observed when participants become familiar with a novel experiment context (Stones et al., 1999) - but an absence of this common reduction in the threat manipulation group (main effect of time: $F(2,22) = 1.00$, $p > .25$; **Fig. 2c**). Across participants, these measures were correlated with each other (self-report & SCL: $r(33) = .39$, $p = .02$; SCL &

cortisol: $r(25) = .47$, $p = .01$; trend for cortisol & self-report: $r(25) = .33$, $p = .09$). To control for the diurnal cycle of cortisol, each participant undertook the experiment between 2pm and 4pm.

Threat eliminates asymmetric information integration. Our results show that the acute threat manipulation eliminated the well-established asymmetry in information integration (Garrett et al., 2014; Moutsiana et al., 2013; Sharot et al., 2011). Specifically, the two sets of information integration parameters (α_G, α_B) were entered into a group (control/threat) by valence (good news/bad news) ANOVA controlling for possible confounds (see **Methods**). The analysis revealed a group by valence interaction ($F(1,27) = 7.56$, $p = .01$, $\eta_p^2 = .22$), which also remained if estimation errors were controlled for ($F(1,26) = 7.88$, $p = .01$) (Garrett and Sharot, 2017) and if the difference between number of good and bad news trials are controlled for ($F(1,26) = 6.97$, $p = .01$).

Post hoc tests revealed that the group by valence interaction was the result of asymmetric information integration in the control group, such that the information integration parameter was larger for good news than bad ($t(15) = 3.34$, $p = .004$, paired sample t-test), but absent in the threat manipulation group ($t(18) = .92$, $p > .25$, paired sample t-test; **Fig. 3**). Participants in the threat manipulation group were more likely to effectively integrate bad news into their beliefs relative to those in the control group (significant difference in bad news information integration parameters α_B: $t(33) = 2.44$, $p = .02$, independent sample t-test), whilst information integration parameters for good news (α_G) did not differ between groups ($t(33) = .611$, $p > .250$, independent sample t-test). There were no floor or ceiling effects for α_G and α_B in the threat manipulation or control group (all at $p < .001$, one sample t-tests versus 0 and 1 respectively) and participants first estimates were not significantly different from the information provided ($t(34) = -0.45$, $p = .65$, one sample t-test versus 0 on the difference between participants' first estimates and the information provided).

Past studies show that asymmetric information integration in this task is not associated with an asymmetry in memory (Moutsiana et al., 2013; Sharot et al., 2011, 2012a, 2012b). In fact, asymmetry in information integration is observed even when the second estimate is elicited immediately after information is on screen (Kuzmanovic et al., 2015, 2016; Kuzmanovic and Rigoux, 2017). Here, we submitted memory scores to a group (threat manipulation/control) by valence (good news/bad news) ANOVA (see **Methods** for details). This did not reveal a main effect of valence ($F(1,33) = 1.24$, $p > .25$), or a main effect of group ($F(1,33) = 1.03$, $p > .25$) or an interaction ($F(1,33) = .62$, $p > .25$). This suggests that valence dependent changes in information integration across groups cannot be attributed to memory or encoding/attention.

Conducting an ANOVA on participants' first estimates with valence (good/bad news) as a repeated factor and group (threat/control) as a between participant factor revealed no main effect of group ($F(1,33) = 1.18$, $p > .25$), the obvious main effect of valence (as trials are binned into good and bad according to first estimates, $F(1,33) = 278.08$, $p < .001$) and a group by valence interaction ($F(1,33) = 6.71$, $p = .014$). The interaction was characterized by the threat group providing lower first estimates than controls for stimuli which will subsequently be categorized as good news ($t(33) = -2.30$, $p = .028$) but no significant difference for trials that will be subsequently categorized as bad news ($t(33) = 1.59$, $p = .123$). Controlling for the difference between first estimates on good and bad news trials in the main ANOVA looking at information integration parameters did not alter the results ($F(1,26) = 5.43$, $p = .028$).

What therefore could account for the selective fluctuations in information integration of bad news? To examine which of the changes to the psychological and physiological measures (SCL, cortisol level, self-report) could *independently* explain alterations in information integration of bad news, we ran a General Linear Model (GLM) in which information integration parameters for bad news (α_B) were entered as the dependent variable and changes in self report, SCL, and cortisol as independent variables (all entered together in one regression). To ensure that effects were valance-specific and could not be accounted for by general changes to information integration, information integration parameters for good news (α_G) were added as a covariate as done before (Moutsiana et al., 2013) [note that the same pattern of results pertains if we omit this covariate (self-reported anxiety: $F(1,17) = 4.75$, $p = 0.04$, SCL: $F(1,17) = 8.81$, $p = .009$)]. We also controlled for all other possible confounds (see **Methods**). The analysis revealed that changes in self-reported anxiety ($F(1,16) = 6.90$, $p = .02$, $b_i = .03$, $\eta_p^2 = .0.30$) and change in physiological stress indicated by SCL ($F(1,16) = 4.99$, $p = .04$, $b_i = .05$, $\eta_p^2 = .24$) explained the variance in information integration parameters for bad news, each of which remained significant if estimation errors were also controlled for (self-reported stress: $F(1,15) = 4.61$, $p = .048$, SCL: $F(1,15) = 4.67$, $p = .047$) (Garrett and Sharot, 2017). In other words, participants who showed the greatest increase in SCL (which reflects the sympathetic component of the autonomic nervous system stress response (Bechara et al., 1996; Figner and Murphy, 2011) and self-reported anxiety were most likely to change their beliefs in proportion to the difference between their first estimates and the bad news received. Change in cortisol (which is suggested to reflect the hypothalamic-pituitary-adrenal (HPA) axis (Gunnar and Quevedo, 2007) component of the stress response) did not relate to information integration for bad news ($F(1,16) = .46$, p > .25,

b_i = -.04, η_p^2 = .03). The null result for cortisol may indicate either that the increase in bad news information integration is not associated specifically with cortisol level increase, or a Type II error. Ratings of emotional arousal, familiarity and information integration parameters for good news (α_G) were also significant predictors in the regression (see **Table 2** for parameter estimates of covariates).

For completeness we repeated the analysis on information integration parameters for good news, α_G (including information integration parameters for bad news, α_B, and all possible covariates mentioned above) and found no significant effects (change in self report: $F(1,16)$ = .47, p > .25, b_i = -.01; change in SCL: $F(1,16)$ = .61, p > .25, b_i = .03; change in cortisol: $F(1,16)$ = .72, p > .25, b_i = .07).

Finally, we examined whether the same results are observed when controlling for within-subject covariates at the within-subject level and between-subject factors at the between-subject level. Specifically, for each participant we computed an alternative set of information integration parameters by correlating absolute estimation error and update controlling for the same within-subject covariates as above (first estimate, vividness, familiarity, past experience and emotional arousal) but controlling for them on a trial by trial basis. We then examined whether these alternative information integration parameters for bad news related to changes in self-reported anxiety and/or changes in SCL (additional between subject factors - initial self-report and the alternative information integration parameters for good news - were also entered as control variables). Indeed, both effects were significant using this approach (change in self report: $F(1,31)$ = 10.57, p = .003, b_i = .05; change in SCL: $F(1,31)$ = 4.51, p = .04, b_i = .08, **Fig. 4a, b**), while the equivalent analysis on information integration parameters from good news was not (change in self report: $F(1,31)$ = .001, p = > .25, b_i = -.001; change in SCL: $F(1,31)$ = .55, p = > .25, b_i = .036).

The results of Experiment I suggested that inducing threat abolishes valence dependent asymmetry in information integration. Thus, the previously observed bias in information integration (Garrett et al., 2014; Korn et al., 2013; Kuzmanovic et al., 2015; Moutsiana et al., 2013, 2015; Sharot et al., 2011, 2012a, 2012b) is not constant but changes with perceived threat in the environment.

Experiment II

Next we set out to extend our findings from Experiment I in a natural setting. Here, we did not fashion a perceived threat, but instead measured anxiety in an environment in which perceived threats would be naturally volatile. Specifically, firefighters from the state of

Colorado performed the belief update task whilst on duty at their respective fire stations. We targeted this group of participants because they would have a naturally large range of anxiety levels owing to the volatile nature of their profession. Changes in cortisol levels were not found to be a significant predictor of information integration parameters for bad news in Experiment I. Therefore, we ruled out collecting this as a measure in Experiment II. Whilst changes in self-reported anxiety and changes in SCR were both found to be significant predictors in Experiment I, these two measures were correlated with one another ($r(33) = .39$, $p = .02$). Since self-reported anxiety had the larger effect size and was easier to collect, we opted to make this our main measure.

Self-reported anxiety was significantly correlated ($r(26) = -.51$, $p < .01$) with the bias in information integration (that is α_G minus α_B). In particular, heightened anxiety was associated with a reduction in the bias. This result remained significant when controlling for possible confounds (see **Methods**), $F(1,23) = 6.67$, $p = .02$, $\eta_p^2 = .23$, $b_i = -.05$.

To examine whether the relationship between heightened anxiety and reduced bias was the result of increased sensitivity to bad news, reduced sensitivity to good news, or both we first constructed a GLM in which information integration parameters for bad news (α_B) was regressed on self-reported anxiety, controlling for possible confounds (mean first estimates, mean ratings of prior experience and number of bad news trials, see **Methods** for details). In addition, to ensure effects were valence-specific and could not be accounted for by general changes in information integration, information integration parameters for good news (α_G) were also added as a covariate (note however that the self-reported anxiety effect pertains if we omit this covariate: $F(1,23) = 9.77$, $p = .005$). This analysis revealed that self-reported anxiety significantly explained the variance in information integration parameters for bad news, α_B ($F(1,22) = 10.52$, $p = .004$, $\eta_p^2 = .32$, $b_i = .05$; **Table 4**), an effect which remained significant if estimation errors are also controlled for ($F(1,21) = 9.79$, $p = .005$) (Garrett and Sharot, 2017). The higher the acute anxiety reported by a firefighter, the more likely the firefighter was to integrate bad news into beliefs in proportion to the difference between their first estimations and the information provided. In this model, information integration from good news ($F(1,22) = 4.69$, $p = .04$) was also a significant predictor of information integration from bad news. There were no floor or ceiling effects for α_G or α_B (all at $p < .001$, one sample t-tests against values of 0 and 1).

We then conducted the same analysis on information integration parameters for good news (α_G) with information integration parameters for bad news (α_B), mean first estimates, mean ratings of prior experience and number of good news trials as covariates. This revealed a nonsignificant trend in the opposite direction than for information integration parameters for bad news, α_B ($F(1,22) = 3.86$, $p = .06$, $\eta_p^2 = .15$, $b_i = -.05$), such that greater self-reported anxiety was related to a trend for *less* information integration in response to good news. Information integration parameters for bad news (α_B) was also significant ($F(1,22) = 7.44$, $p = 0.01$, $\eta_p^2 = .25$, $b_i = 0.75$).

Finally, we examined whether the same results are observed when controlling for within-subject covariates at the within-subject level and between-subject factors at the between-subject level. Under this alternative approach higher self-reported anxiety was related to greater information integration in response to bad news ($F(1,24) = 8.34$, $p = .008$, $b_i = .03$, **Fig. 5a**). For good news the opposite effect was found such that higher self-reported anxiety was related to reduced information integration ($F(1,24) = 4.80$, $p = .038$, $b_i = -.045$, **Fig. 5b**). It is interesting that this latter effect was observed only in Experiment 2 and not Experiment 1, which may indicate that natural real-life threats could have an especially strong impact on information integration processes.

These results suggest that anxiety is related to a valence-dependent enhancement in the ability to adjust beliefs in response to new information. We highlight that whilst in Experiment I, threat was manipulated and thus causation could be inferred by comparing the threat manipulation and control groups, Experiment II was conducted to reveal an *association* in "real life". Together, the experiments suggest that under a perceived threat (whether manipulated or naturally occurring) positively biased integration of information is not observed.

Discussion

Our results provide evidence that the well-documented asymmetry in belief formation evaporates under perceived threat. Specifically, Experiment I shows that in a low threat environment individuals integrated information asymmetrically, faithfully incorporating good news into their existing beliefs while relatively disregarding bad news (Eil and Rao, 2011; Sharot et al., 2011). Under perceived threat however, this asymmetry disappeared; participants showed an increased capacity to integrate bad news into prior beliefs. Increased physiological arousal and self-reported anxiety were found to correlate with enhanced integration of unfavorable information into beliefs. In Experiment II, firefighters on duty who reported higher state anxiety also exhibited greater selective integration of bad news. Because the increase in information integration in both experiments was valence specific it cannot reflect a general improvement in learning, and because memory for the information presented was not affected, modulation of attention is an unlikely explanation.

The finding that the positivity bias in belief updating alters flexibly as a function of perceived threat reveals a potentially adaptive mechanism. In particular, the relative failure to incorporate bad news into prior beliefs leads to positively biased beliefs (also known as the optimism bias). This bias can lead to both positive effects – including increased exploration (Berger-Tal and Avgar, 2012) and motivation (Bandura, 1989) - and negative effects – including failure to take precautionary action. It has been suggested that overestimating the likelihood of attaining rewards and underestimating the likelihood of harm is adaptive in environments where potential gains are sufficiently greater than costs (Johnson and Fowler, 2011). This is because under uncertainty, optimistically biased individuals will claim resources (e.g., a spouse or a job) they could not otherwise attain, as better but less optimistic competitors may walk away from the fight. Moreover, overestimating the value of novel environments can lead to increased rate of exploration allowing the opportunity for the true value of an environment to be learned quicker (Berger-Tal and Avgar, 2012; Sutton and Barto, 1998), which is associated with superior performance in behaviours such as reproduction (Egas and Sabelis, 2001) and foraging (Rutz et al., 2006). However, in environments where potential harm is considerably greater than potential reward, computational models suggest the optimism bias to be disadvantageous (Johnson and Fowler, 2011). Thus, a valence dependent bias in information integration that disappears under threat could be optimal in enabling a more accurate assessment of risk.

In our experiments, the source of the threat was unrelated to the information content of the task. Thus, acute stress had a valence-specific, yet general, effect on how participants used information to alter their beliefs (i.e. in response to a social threat, participants did not selectively increase their response to information about social judgment, but to negative information in general). Indeed, many threat induction methods, including threat of electric shock, Cold Pressor Tasks and the Trier Social Stress Test, produce general changes to behavior and neural responses that are not confined to the source of the threat itself (Cavanagh et al., 2010; Lenow et al., 2017; Otto et al., 2013; Robinson et al., 2013; Youssef et al., 2012). Similar findings have been observed in non-human animals, where different stressors have been shown to alter the degree of positive biases in a range of decision-making tasks (Harding et al., 2004; Matheson et al., 2008; Rygula et al., 2013). This may be adaptive, as threat may signify a dangerous environment that requires a general enhancement of caution.

However, if perceived threat is prolonged or dissociated from reality, enhanced integration of negative information over long periods of time could lead to psychiatric problems. We have previously shown that patients suffering from Major Depressive Disorder (MDD) exhibit increased updating of beliefs in response to negative information relative to healthy controls (Garrett et al., 2014). MDD is often triggered by a stressful life event (Caspi et al., 2003; Roiser et al., 2012). In individuals predisposed to MDD such a stressful life event (or series of such events) could result in prolonged periods of perceived threat and thus increased sensitivity to negative information. This in turn can form pessimistic beliefs, a symptom of MDD (American Psychiatric Association, 2013; Strunk et al., 2006), leading to even greater perceived threat about one's environment. It is possible that a similar mechanism may contribute to symptoms observed in other clinical pathologies such as in clinical anxiety and phobia.

We speculate that stress in response to perceived threat may interfere with top down control mechanisms that may normally inhibit integration of unwanted information (for review see Yu, 2016). A second, not mutually exclusive, possibility is that the stress reaction directly boosts the neural representation of estimation errors generated from bad, but not good, news. Indeed, it has been shown that negative prediction errors in dopamine rich striatal nuclei are selectively amplified under threat (Robinson et al., 2013) - a modulation that could be mediated by stress-induced changes to dopamine release (Frank et al., 2004; Lemos et al., 2012; Schultz et al., 1997; Sharot et al., 2012a). Future studies are required to test these hypotheses.

In sum, our results provide evidence that asymmetric information integration is not set in stone, but changes acutely in response to the environment, decreasing under perceived threat. Such flexibility could be adaptive, potentially enhancing our likelihood to respond to warnings with caution in environments where future costs may be high, but enabling us to maintain positive beliefs otherwise, a strategy that has been suggested, on balance, to increase well-being (McKay and Dennett, 2010).

References

American Psychiatric Association (2013). Diagnostic and statistical manual of mental disorders (DSM-5®) (American Psychiatric Pub).

Bandura, A. (1989). Human agency in social cognitive theory. Am. Psychol. *44*, 1175–1184.

Bechara, A., Tranel, D., Damasio, H., and Damasio, A.R. (1996). Failure to respond autonomically to anticipated future outcomes following damage to prefrontal cortex. Cereb. Cortex N. Y. N 1991 *6*, 215–225.

Berger-Tal, O., and Avgar, T. (2012). The Glass is Half-Full: Overestimating the Quality of a Novel Environment is Advantageous. PLoS ONE *7*, e34578.

Birkett, M.A. (2011). The Trier Social Stress Test Protocol for Inducing Psychological Stress. J. Vis. Exp.

Caspi, A., Sugden, K., Moffitt, T.E., Taylor, A., Craig, I.W., Harrington, H., McClay, J., Mill, J., Martin, J., Braithwaite, A., et al. (2003). Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. Science *301*, 386–389.

Cavanagh, J.F., Frank, M.J., and Allen, J.J. (2010). Social stress reactivity alters reward and punishment learning. Soc. Cogn. Affect. Neurosci. *6*, 311–320.

Chang, E.C. (2001). Optimism and Pessimism: Implications for Theory, Research, and Practice (Washington, DC: American Psychological Association).

Chlan, L., Savik, K., and Weinert, C. (2003). Development of a Shortened State Anxiety Scale From the Spielberger State-Trait Anxiety Inventory (STAI) for Patients Receiving Mechanical Ventilatory Support. J. Nurs. Meas. *11*, 283–293.

Chowdhury, R., Sharot, T., Wolfe, T., Duzel, E., and Dolan, R.J. (2014). Optimistic update bias increases in older age. Psychol. Med. *44*, 2003–2012.

Egas, M., and Sabelis, M.W. (2001). Adaptive learning of host preference in a herbivorous arthropod. Ecol. Lett. *4*, 190–195.

Eil, D., and Rao, J.M. (2011). The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself. Am. Econ. J. Microecon. *3*, 114–138.

Figner, B., and Murphy, R.O. (2011). Using skin conductance in judgment and decision making research. Handb. Process Tracing Methods Decis. Res. 163–184.

Frank, M.J., Seeberger, L.C., and O'reilly, R.C. (2004). By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science *306*, 1940–1943.

Garrett, N., and Sharot, T. (2014). How Robust Is the Optimistic Update Bias for Estimating Self-Risk and Population Base Rates? PLoS ONE *9*, e98848.

Garrett, N., and Sharot, T. (2017). Optimistic update bias holds firm: Three tests of robustness following Shah et al. Conscious. Cogn. *50*, 12–22.

Garrett, N., Sharot, T., Faulkner, P., and Korn, C.W. (2014). Losing the rose tinted glasses: neural substrates of unbiased belief updating in depression. Front. Hum. Ldots.

Gunnar, M., and Quevedo, K. (2007). The neurobiology of stress and development. Annu. Rev. Psychol. *58*, 145–173.

Harding, E.J., Paul, E.S., and Mendl, M. (2004). Animal behaviour: Cognitive bias and affective state. Nature *427*, 312–312.

Johnson, D.D.P., and Fowler, J.H. (2011). The evolution of overconfidence. Nature *477*, 317–320.

Kappes, A., Faber, N.S., Kahane, G., Savulescu, J., and Crockett, M.J. (2018). Concern for Others Leads to Vicarious Optimism. Psychol. Sci. *29*, 379–389.

Korn, C.W., Prehn, K., Park, S.Q., Walter, H., and Heekeren, H.R. (2012). Positively Biased Processing of Self-Relevant Social Feedback. J. Neurosci. *32*, 16832–16844.

Korn, C.W., Sharot, T., Walter, H., Heekeren, H.R., and Dolan, R.J. (2013). Depression is related to an absence of optimistically biased belief updating about future life events. Psychol. Med. *44*, 579–592.

Kuzmanovic, B., and Rigoux, L. (2017). Valence-Dependent Belief Updating: Computational Validation. Front. Psychol. *8*.

Kuzmanovic, B., Jefferson, A., and Vogeley, K. (2015). Self-specific Optimism Bias in Belief Updating Is Associated with High Trait Optimism. J. Behav. Decis. Mak. *28*, 281–293.

Kuzmanovic, B., Jefferson, A., and Vogeley, K. (2016). The role of the neural reward circuitry in self-referential optimistic belief updates. NeuroImage *133*, 151–162.

Lefebvre, G., Lebreton, M., Meyniel, F., Bourgeois-Gironde, S., and Palminteri, S. (2017). Behavioural and neural characterization of optimistic reinforcement learning. Nat. Hum. Behav. *1*, 0067.

Lemos, J.C., Wanat, M.J., Smith, J.S., Reyes, B.A.S., Hollon, N.G., Van Bockstaele, E.J., Chavkin, C., and Phillips, P.E.M. (2012). Severe stress switches CRF action in the nucleus accumbens from appetitive to aversive. Nature *490*, 402–406.

Lenow, J.K., Constantino, S.M., Daw, N.D., and Phelps, E.A. (2017). Chronic and acute stress promote overexploitation in serial decision-making. J. Neurosci. *37*, 5681–5689.

Lighthall, N.R., Gorlick, M.A., Schoeke, A., Frank, M.J., and Mather, M. (2013). Stress modulates reinforcement learning in younger and older adults. Psychol. Aging *28*, 35.

Marteau, T.M., and Bekker, H. (1992). The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). Br. J. Clin. Psychol. Br. Psychol. Soc. *31 (Pt 3)*, 301–306.

Matheson, S.M., Asher, L., and Bateson, M. (2008). Larger, enriched cages are associated with 'optimistic' response biases in captive European starlings (Sturnus vulgaris). Appl. Anim. Behav. Sci. *109*, 374–383.

McKay, R.T., and Dennett, D.C. (2010). The evolution of misbelief. Behav. Brain Sci. *32*, 493.

Mobius, M.M., Niehaus, P., Niederle, M., and Rosenblat, T.S. (2012). Managing Self-Confidence: Theory and Experimental Evidence.

Moutsiana, C., Garrett, N., Clarke, R.C., Lotto, R.B., Blakemore, S.-J., and Sharot, T. (2013). Human development of the ability to learn from bad news. Proc. Natl. Acad. Sci. *110*, 16396–16401.

Moutsiana, C., Charpentier, C.J., Garrett, N., Cohen, M.X., and Sharot, T. (2015). Human Frontal–Subcortical Circuit and Asymmetric Belief Updating. J. Neurosci. *35*, 14077–14085.

Otto, A.R., Raio, C.M., Chiang, A., Phelps, E.A., and Daw, N.D. (2013). Working-memory capacity protects model-based learning from stress. Proc. Natl. Acad. Sci. *110*, 20941–20946.

Robinson, O.J., Overstreet, C., Charney, D.R., Vytal, K., and Grillon, C. (2013). Stress increases aversive prediction error signal in the ventral striatum. Proc. Natl. Acad. Sci. *110*, 4129–4133.

Roiser, J.P., Elliott, R., and Sahakian, B.J. (2012). Cognitive Mechanisms of Treatment in Depression. Neuropsychopharmacology *37*, 117–136.

Rutz, C., Whittingham, M.J., and Newton, I. (2006). Age-dependent diet choice in an avian top predator. Proc. R. Soc. Lond. B Biol. Sci. *273*, 579–586.

Rygula, R., Papciak, J., and Popik, P. (2013). Trait pessimism predicts vulnerability to stress-induced anhedonia in rats. Neuropsychopharmacology *38*, 2188–2196.

Schultz, W., Dayan, P., and Montague, P.R. (1997). A neural substrate of prediction and reward. Science *275*, 1593–1599.

Sharot, T., and Garrett, N. (2016). Forming Beliefs: Why Valence Matters. Trends Cogn. Sci. *20*, 25–33.

Sharot, T., Korn, C.W., and Dolan, R.J. (2011). How unrealistic optimism is maintained in the face of reality. Nat. Publ. Group *14*, 1475–1479.

Sharot, T., Guitart-Masip, M., Korn, C.W., Chowdhury, R., and Dolan, R.J. (2012a). How dopamine enhances an optimism bias in humans. Curr. Biol. CB *22*, 1477–1481.

Sharot, T., Kanai, R., Marston, D., Korn, C.W., Rees, G., and Dolan, R.J. (2012b). Selectively altering belief formation in the human brain. Proc. Natl. Acad. Sci. U. S. A. *109*, 17058–17062.

Shefrin, H. (2009). How Psychological Pitfalls Generated the Global Financial Crisis (Rochester, NY: Social Science Research Network).

Sherman, S. (1980). On the self-erasing nature of errors of prediction. J. Pers. Soc. Psychol. *39.21*, 1–22.

Stones, A., Groome, D., Perry, D., Hucklebridge, F., and Evans, P. (1999). The effect of stress on salivary cortisol in panic disorder patients. J. Affect. Disord. *52*, 197–201.

Strunk, D.R., Lopez, H., and DeRubeis, R.J. (2006). Depressive symptoms are associated with unrealistic negative predictions of future life events. Behav. Res. Ther. *44*, 861–882.

Sutton, R.S., and Barto, A.G. (1998). Reinforcement learning: An introduction.

Taylor, S.E., and Brown, J.D. (1988). Illusion and well-being: a social psychological perspective on mental health. Psychol. Bull. *103*, 193–210.

Tiger, L. (1979). Optimism: The Biology of Hope. (New York: Simon & Schuster).

Wiswall, M., and Zafar, B. (2015). How Do College Students Respond to Public Information about Earnings? J. Hum. Cap. *9*, 117–169.

Youssef, F.F., Dookeeram, K., Basdeo, V., Francis, E., Doman, M., Mamed, D., Maloo, S., Degannes, J., Dobo, L., Ditshotlo, P., et al. (2012). Stress alters personal moral decision making. Psychoneuroendocrinology *37*, 491–498.

Yu, R. (2016). Stress potentiates decision biases: A stress induced deliberation-to-intuition (SIDI) model. Neurobiol. Stress *3*, 83–95.

Figures and Tables

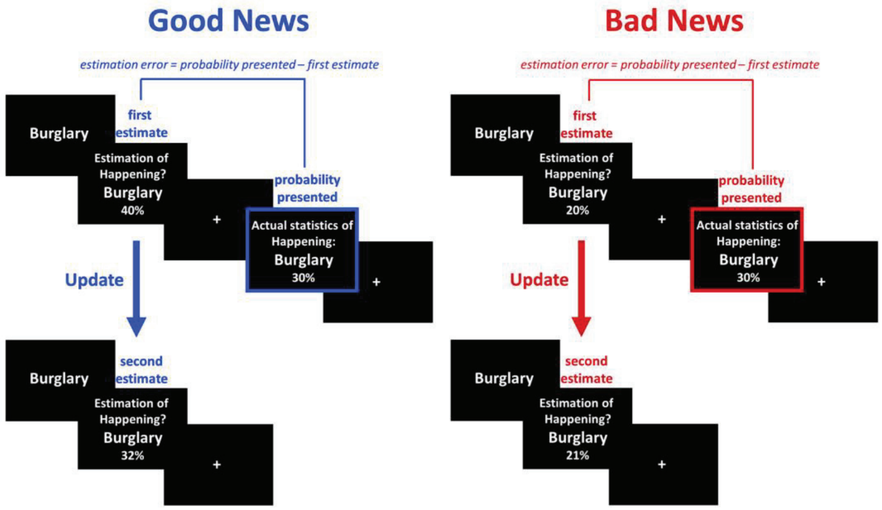


Figure 1. Behavioral Task.

On each trial, participants were presented with a short description of an adverse event and asked to estimate how likely this event was to occur to them in the future. They were then presented with the probability of that event occurring to someone from the same age, location and socio-economic background as them. The second session was the same as the first except that the average probability of the event to occur was not presented. Examples of trials for which the participant's estimate was (a) higher or (b) lower than the statistical information provided leading to receipt of good and bad news respectively.

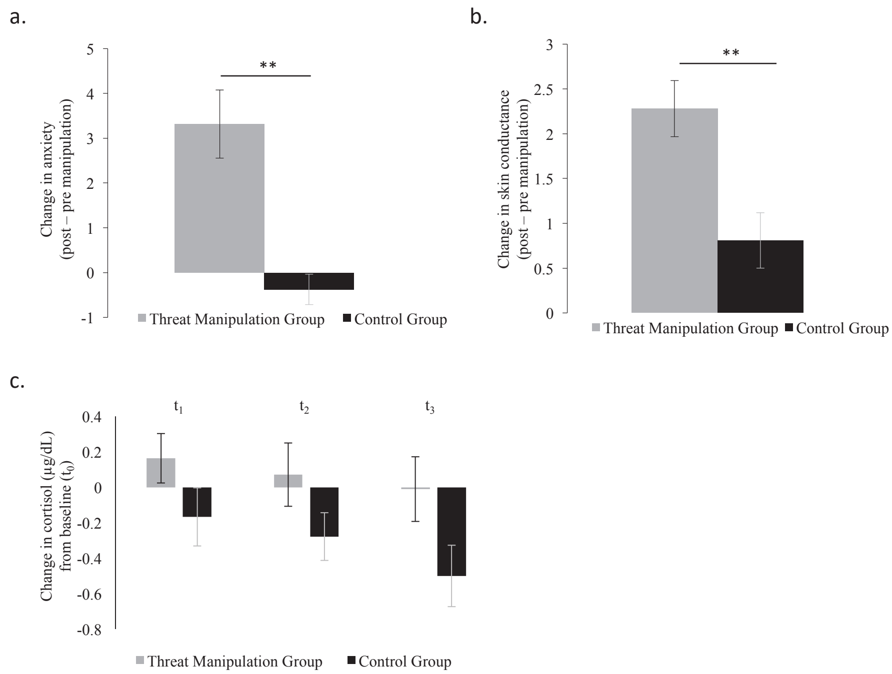


Figure 2. Manipulation Check.

Measures of (a) self-reported state anxiety, (b) skin conductance and (c) cortisol levels were greater after manipulation relative to before in the threat manipulation group compared to the control group. Time points for cortisol measurements are as follows: t0 = before threat/control manipulation procedure; t1 = immediately after threat/control manipulation procedure, prior to undertaking the task (+10 min from t0); t2 = halfway through the task (+30min from t0); t3 = after completion of task and post experiment questionnaires (+1hr from t0).

** *p* < .050; Error bars represent standard error of the mean.

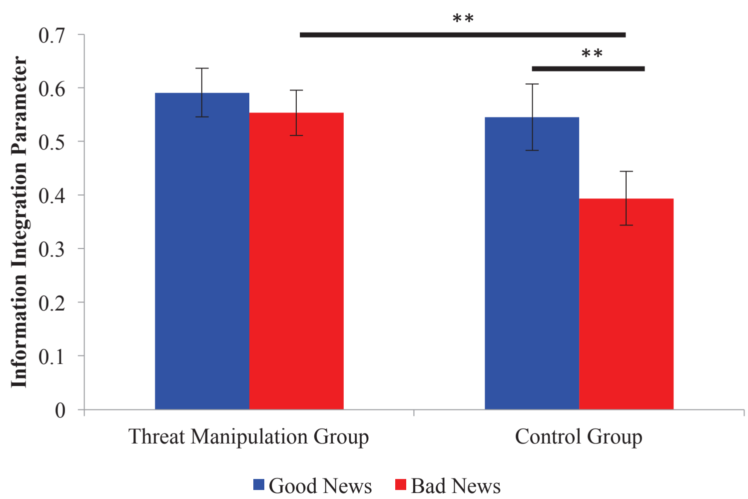


Figure 3. Bias in Information Integration Parameters Vanishes under Threat manipulation.

While the control group showed asymmetrical information integration parameters (***α***) in response to good and bad news, this bias vanished in the threat manipulation group, due to an increase in α_B (information integration parameter for bad news). The Group*Valence interaction was significant, controlling for all covariates identified in Table 1 (see Methods). ** $p < .05$ independent/paired sample t test as appropriate; Error bars represent standard error of the mean.

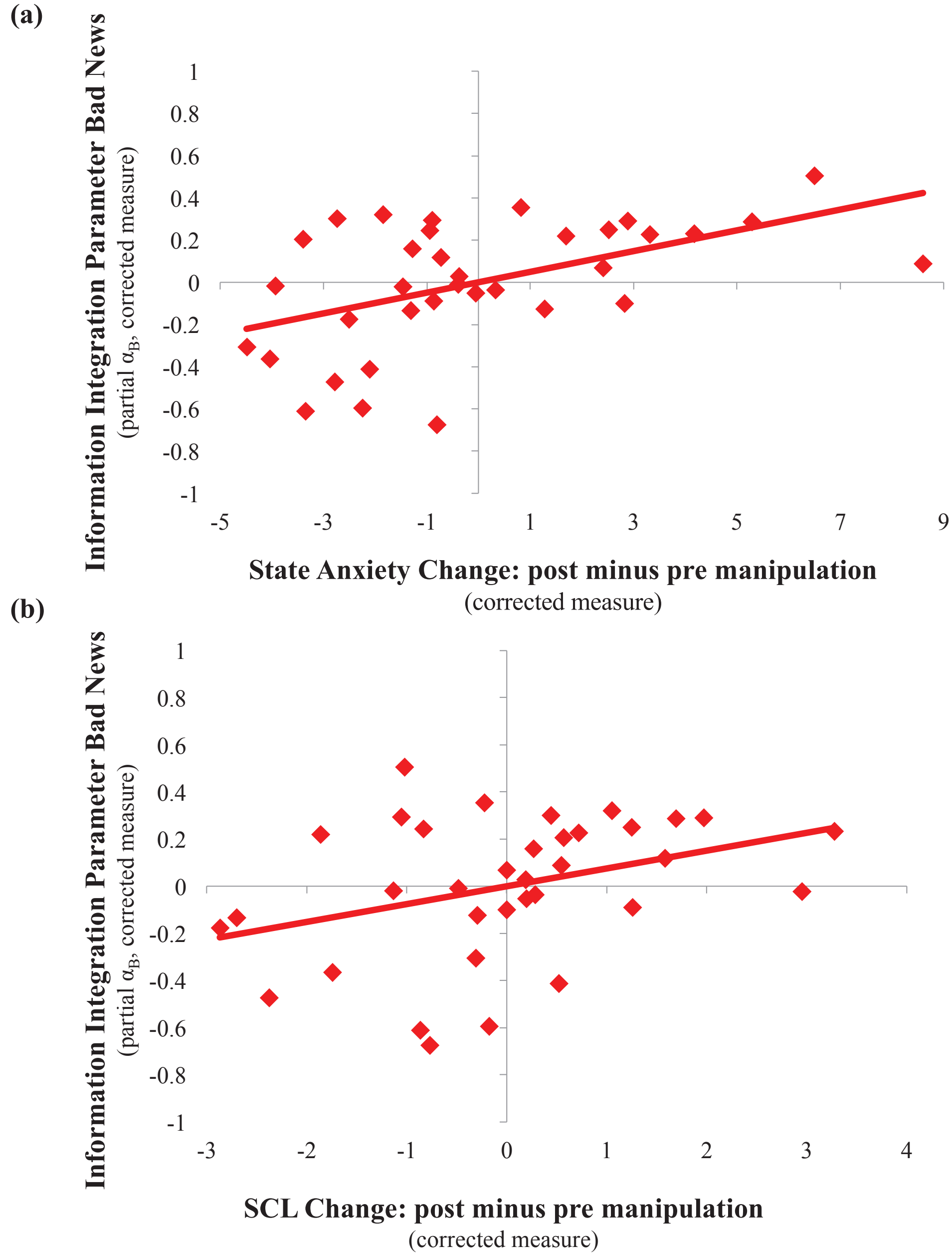


Figure 4. Greater integration of bad news related to state anxiety and SCL.
Following the manipulation, an increase in both **a.** self-reported anxiety (b_i = .049, p = .003, η_p^2 = .25) and **b.** skin conductance (SCL) (b_i = .076, p = .042, η_p^2 = .13) were related to larger information integration from bad news, correcting for possible confounds. Plotted are the partial regression plots from two linear models (one for self-report and one for SCL) that control for additional covariates.

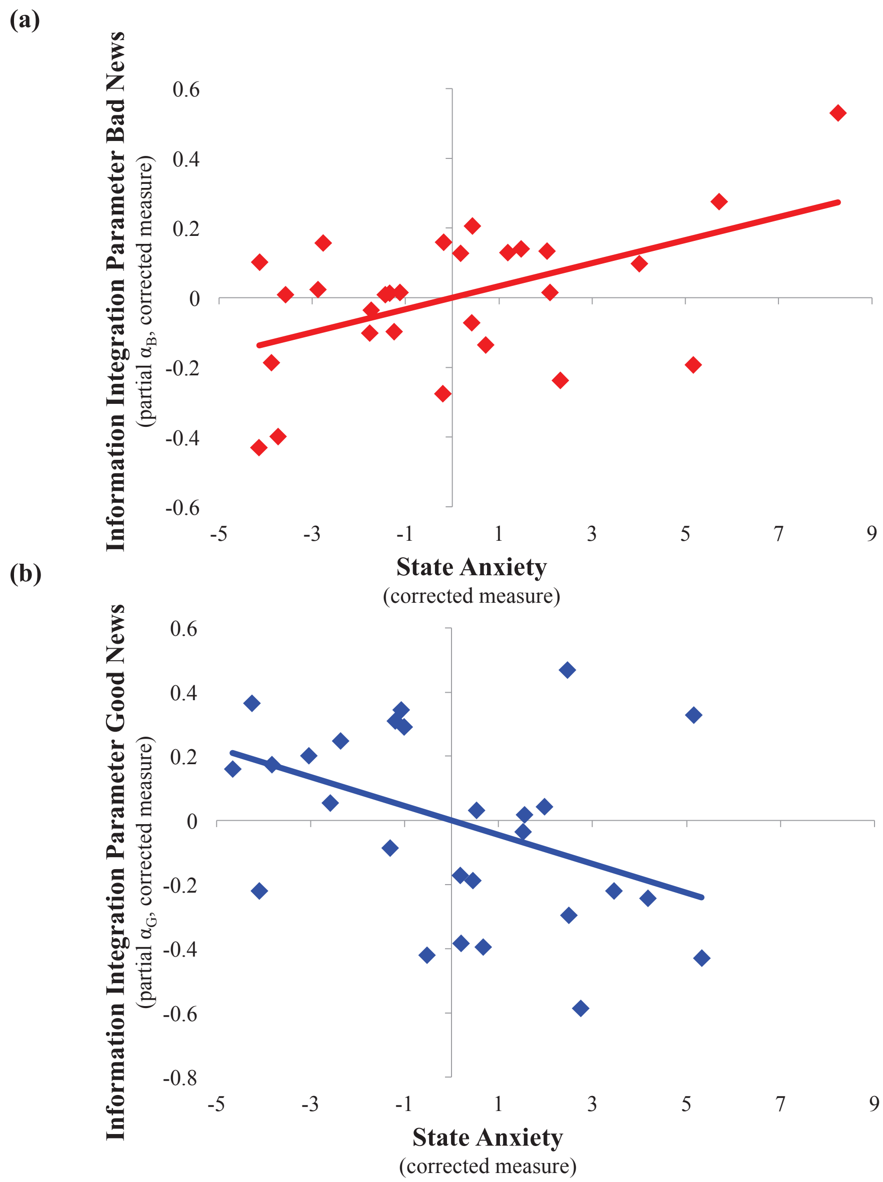


Figure 5. State anxiety in firefighters differentially relate to integration of good and bad news. Subjective state anxiety scores (STAI) of firefighters on shift were related to larger information integration from bad news (b_i = .03, p = .008, η_p^2 = .26) and lower information integration from good news (b_i = -0.045, p = .038, η_p^2 = .17), correcting for possible confounds. Plotted are the partial regression plots for **a.** bad news (partial α_B) and **b.** good news (partial α_G) from 2 separate linear models (one for bad news and one for good news) that control for additional covariates.

	Threat Manipulation Group mean (SD)	Control Group mean (SD)
BDI and Baseline Stress Levels		
BDI	5.79 (5.23)	4.69 (3.22)
Initial Self Report STAI[G]	10.37 (2.65)	8.63 (1.36)
Initial SCL	6.27 (3.29)	5.90 (3.20)
Initial Cortisol (log transformed)	-1.99 (0.59)	-1.79 (0.53)
Task Variables		
First Estimates	29.82 (5.62)[V]	31.05 (5.89)[V]
Subjective Scales Questionnaire ***1 = low to 6 = high***	**Bias** ***(Good News – Bad News)***	
Vividness	0.41 (0.72)[V]	0.72 (0.65)[V]
Familiarity	0.30 (0.69)	0.49 (0.62)[V]
Prior experience	0.18 (0.61)	0.33 (0.41)[V]
Emotional arousal	0.33 (0.63)[V]	0.13 (0.86)
Negativity	0.20 (0.49)	-0.13 (0.58)
Other Task-related variables		
Number of Trials	-1.58 (8.99)	-1.56 (9.70)
Memory errors	-1.23 (3.16)	-0.21 (4.52)
Estimation errors (absolute)	-0.82 (5.27)	1.11 (5.84)
Update	2.60 (12.67)	4.21 (7.83)[V]

Table 1. BDI, Initial Self-Report STAI, Initial SCL, Initial Cortisol, Task-related variables, subjective scales, memory in Experiment I. Note that Estimation errors and Update (the final two rows) are the variables used to compute the information integration parameters (α_G and α_B) for each participant.

[G] Difference between Threat Manipulation and Control Groups, tested using independent sample t-tests ($p<0.05$).

[V] Significant effect of valence ($p<0.05$), tested using one sample t-test on the bias scores (difference between good and bad news) on each group separately.

	b_i	*Std. Error*	*t*	*p*	*95% Confidence Interval*		η_p^2
					Lower Bound	*Upper Bound*	
Initial Self Report STAI	-0.01	0.01	-1.10	0.29	-0.04	0.01	0.07
First estimates	0.01	0.01	0.89	0.39	-0.01	0.02	0.05
Vividness rating	-0.09	0.05	-1.84	0.09	-0.20	0.01	0.17
Familiarity rating	0.08	0.04	2.16	0.05	0.00	0.16	0.23
Prior experience rating	-0.04	0.06	-0.76	0.46	-0.17	0.08	0.04
Emotional arousal rating	-0.13	0.04	-3.03	0.01	-0.22	-0.04	0.37
Information integration parameter, good news (α_G)	0.39	0.15	2.60	0.02	0.07	0.71	0.30

Table 2. Parameter estimates of covariates in Experiment I.

First estimates (i.e. mean initial estimations), mean ratings on subjective scales (vividness, familiarity, past experience and emotional arousal) and α_G (information integration parameters for good news) were entered as covariates to account for fluctuations in α_B (information integration parameters for bad news).

	Mean (SD)
BDI	6.82 (7.45)
Task Variables	
First Estimates[V]	31.22 (6.96)
Subjective Scales Questionnaire ***1 = low to 6 = high***	**Bias** ***(Good News – Bad News)***
Prior experience[V]	0.54 (0.94)
Negativity	0.31 (0.90)
Other Task-related variables	
Number of Trials[V]	-10.89 (9.41)
Memory errors	-2.18 (6.51)
Estimation errors (absolute)[V]	-2.91 (5.16)
Update[V]	9.49 (12.04)

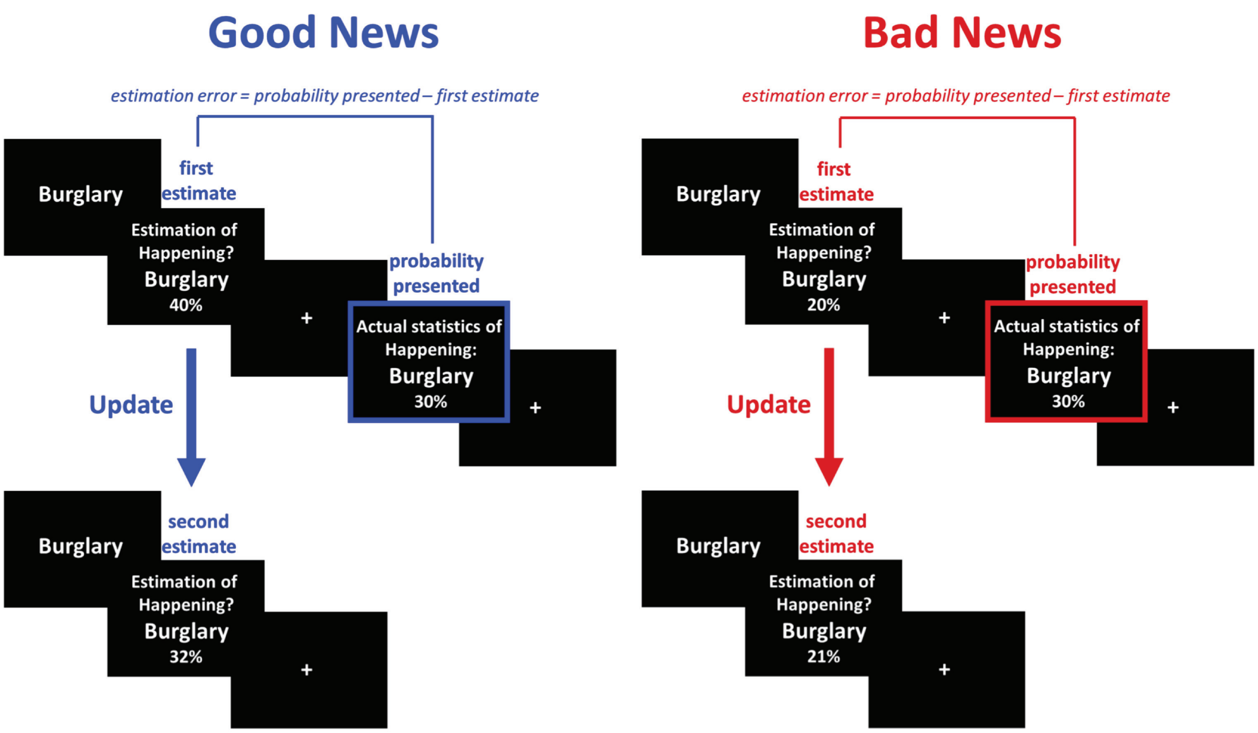
Table 3. Task-related variables, subjective scales and memory in Experiment II. Note that Estimation errors and Update (the final two rows) are the variables used to compute the information integration parameters (α_G and α_B) for each participant.

[V] Significant effect of valence ($p<0.05$), tested using one sample t-test on the mean bias scores (difference between good and bad news) for each participant.

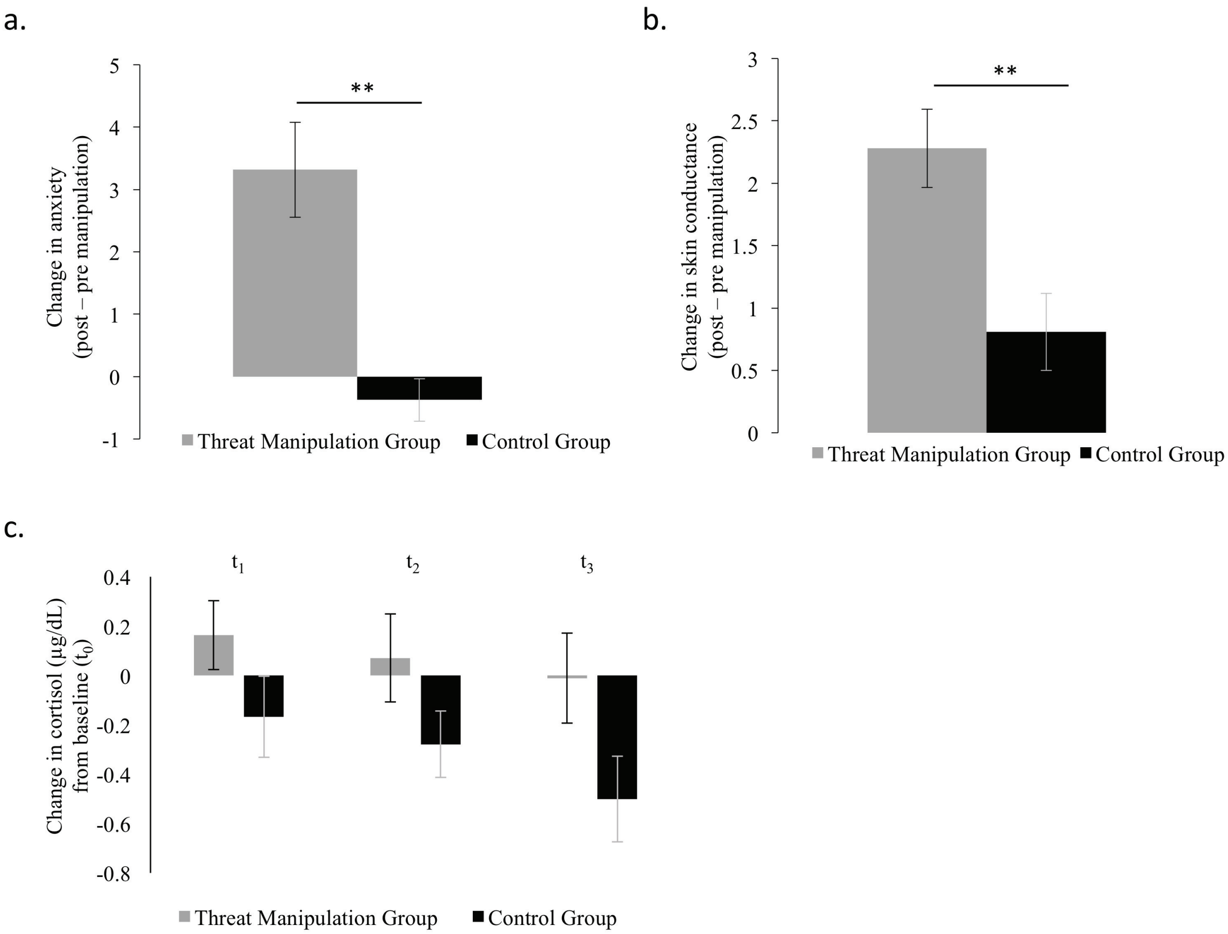
					95% Confidence Interval		
	b_i	*Std. Error*	*t*	*p*	*Lower Bound*	*Upper Bound*	η_p^2
First estimates	0.00	0.01	-0.03	0.97	-0.02	0.02	0.00
Prior experience rating	-0.03	0.09	-0.31	0.76	-0.21	0.15	0.00
Number of bad news trials	0.00	0.02	-0.01	0.99	-0.03	0.03	0.00
Information integration parameter, good news (α_G)	0.34	0.16	2.17	0.04	0.02	0.67	0.18

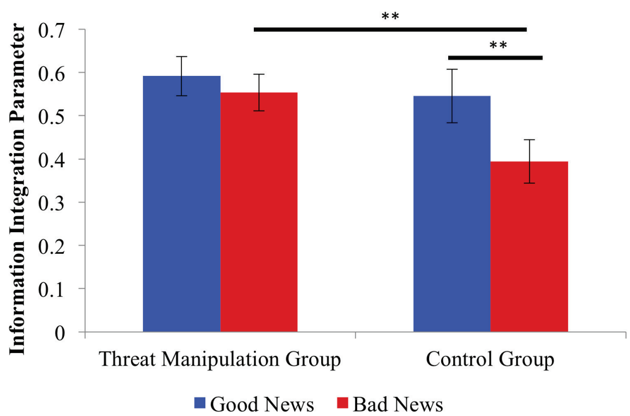
Table 4. Parameter estimates of covariates in Experiment II.

First estimates (i.e. mean initial estimations), mean ratings of past experience, number of bad news trials and α_G (information integration parameters for good news) were entered as covariates to account for fluctuations in α_B (information integration parameters for bad news).

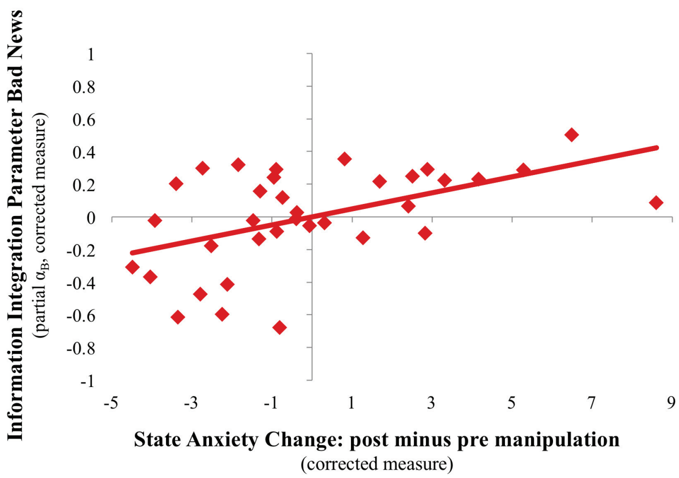


Good News
estimation error = probability presented – first estimate
Burglary
first estimate
Estimation of Happening?
Burglary
40%
+
probability presented
Actual statistics of Happening:
Burglary
30%
+
Update
Burglary
second estimate
Estimation of Happening?
Burglary
32%
+
Bad News
estimation error = probability presented – first estimate
Burglary
first estimate
Estimation of Happening?
Burglary
20%
+
probability presented
Actual statistics of Happening:
Burglary
30%
+
Update
Burglary
second estimate
Estimation of Happening?
Burglary
21%
+

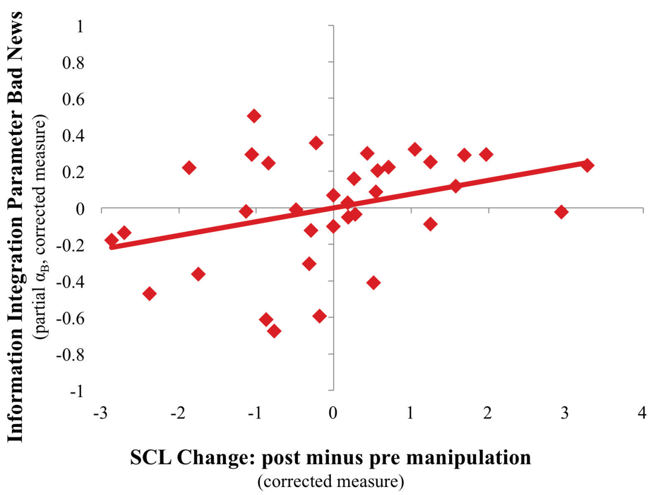

a.
Change in anxiety
(post – pre manipulation)
5
4
3
2
1
0
-1
**
Threat Manipulation Group
Control Group
b.
Change in skin conductance
(post – pre manipulation)
3
2.5
2
1.5
1
0.5
0
**
Threat Manipulation Group
Control Group
c.
Change in cortisol (µg/dL)
from baseline (t0)
t1
t2
t3
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
Threat Manipulation Group
Control Group


Information Integration Parameter
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
**
**
Threat Manipulation Group
Control Group
Good News
Bad News

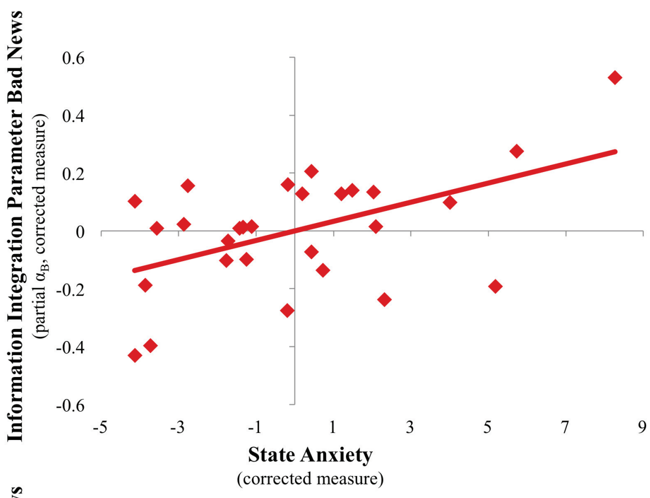
(a)


Information Integration Parameter Bad News
(partial α_B, corrected measure)
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-5
-3
-1
1
3
5
7
9
State Anxiety Change: post minus pre manipulation
(corrected measure)

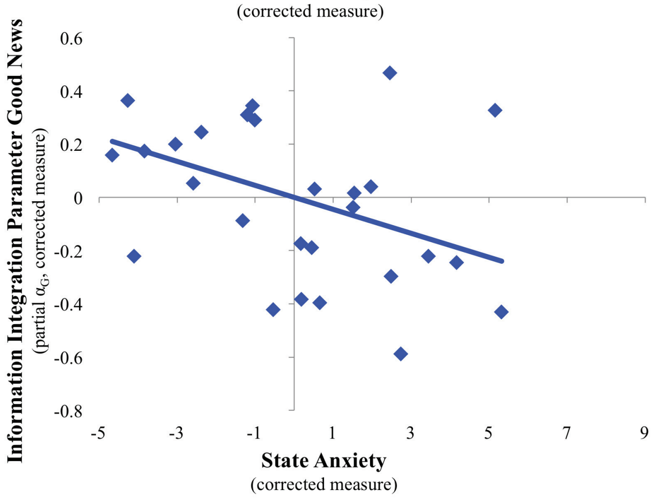
(b)


Information Integration Parameter Bad News
(partial α_B, corrected measure)
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-3
-2
-1
0
1
2
3
4
SCL Change: post minus pre manipulation
(corrected measure)

(a)


Information Integration Parameter Bad News
(partial α_B, corrected measure)
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-5
-3
-1
1
3
5
7
9
State Anxiety
(corrected measure)

(b)


Information Integration Parameter Good News
(partial α_G, corrected measure)
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-5
-3
-1
1
3
5
7
9
State Anxiety
(corrected measure)

	Threat Manipulation Group mean (SD)	Control Group mean (SD)
BDI and Baseline Stress Levels		
BDI	5.79 (5.23)	4.69 (3.22)
Initial Self Report STAI^{G}	10.37 (2.65)	8.63 (1.36)
Initial SCL	6.27 (3.29)	5.90 (3.20)
Initial Cortisol (log transformed)	-1.99 (0.59)	-1.79 (0.53)
Task Variables		
First Estimates	29.82 (5.62)^{V}	31.05 (5.89)^{V}
Subjective Scales Questionnaire 1 = low to 6 = high	**Bias** ***(Good News – Bad News)***	
Vividness	0.41 (0.72)^{V}	0.72 (0.65)^{V}
Familiarity	0.30 (0.69)	0.49 (0.62)^{V}
Prior experience	0.18 (0.61)	0.33 (0.41)^{V}
Emotional arousal	0.33 (0.63)^{V}	0.13 (0.86)
Negativity	0.20 (0.49)	-0.13 (0.58)
Other Task-related variables		
Number of Trials	-1.58 (8.99)	-1.56 (9.70)
Memory errors	-1.23 (3.16)	-0.21 (4.52)
Estimation errors (absolute)	-0.82 (5.27)	1.11 (5.84)
Update	2.60 (12.67)	4.21 (7.83)^{V}

Table 1. BDI, Initial Self-Report STAI, Initial SCL, Initial Cortisol, Task-related variables, subjective scales, memory in Experiment I. Note that Estimation errors and Update (the final two rows) are the variables used to compute the information integration parameters (α_G and α_B) for each participant.

^{G} Difference between Threat Manipulation and Control Groups, tested using independent sample t-tests ($p<0.05$).

^{V} Significant effect of valence ($p<0.05$), tested using one sample t-test on the bias scores (difference between good and bad news) on each group separately.

	b_i	*Std. Error*	*t*	*p*	*95% Confidence Interval* *Lower Bound*	*Upper Bound*	η_p^2
Initial Self Report STAI	-0.01	0.01	-1.10	0.29	-0.04	0.01	0.07
First estimates	0.01	0.01	0.89	0.39	-0.01	0.02	0.05
Vividness rating	-0.09	0.05	-1.84	0.09	-0.20	0.01	0.17
Familiarity rating	0.08	0.04	2.16	0.05	0.00	0.16	0.23
Prior experience rating	-0.04	0.06	-0.76	0.46	-0.17	0.08	0.04
Emotional arousal rating	-0.13	0.04	-3.03	0.01	-0.22	-0.04	0.37
Information integration parameter, good news (α_G)	0.39	0.15	2.60	0.02	0.07	0.71	0.30

Table 2. Parameter estimates of covariates in Experiment I.

First estimates (i.e. mean initial estimations), mean ratings on subjective scales (vividness, familiarity, past experience and emotional arousal) and α_G (information integration parameters for good news) were entered as covariates to account for fluctuations in α_B (information integration parameters for bad news).

	Mean (SD)
BDI	6.82 (7.45)
Task Variables	
First Estimates[V]	31.22 (6.96)
Subjective Scales Questionnaire ***1 = low to 6 = high***	**Bias** ***(Good News – Bad News)***
Prior experience[V]	0.54 (0.94)
Negativity	0.31 (0.90)
Other Task-related variables	
Number of Trials[V]	-10.89 (9.41)
Memory errors	-2.18 (6.51)
Estimation errors (absolute)[V]	-2.91 (5.16)
Update[V]	9.49 (12.04)

Table 3. Task-related variables, subjective scales and memory in Experiment II. Note that Estimation errors and Update (the final two rows) are the variables used to compute the information integration parameters (α_G and α_B) for each participant.

[V] Significant effect of valence ($p<0.05$), tested using one sample t-test on the mean bias scores (difference between good and bad news) for each participant.

	b_i	*Std. Error*	*t*	*p*	*95% Confidence Interval* Lower Bound	*95% Confidence Interval* Upper Bound	η_p^2
First estimates	0.00	0.01	-0.03	0.97	-0.02	0.02	0.00
Prior experience rating	-0.03	0.09	-0.31	0.76	-0.21	0.15	0.00
Number of bad news trials	0.00	0.02	-0.01	0.99	-0.03	0.03	0.00
Information integration parameter, good news (α_G)	0.34	0.16	2.17	0.04	0.02	0.67	0.18

Table 4. Parameter estimates of covariates in Experiment II.

First estimates (i.e. mean initial estimations), mean ratings of past experience, number of bad news trials and α_G (information integration parameters for good news) were entered as covariates to account for fluctuations in α_B (information integration parameters for bad news).