Selectively altering belief formation in the human brain

Tali Sharot[a,1,2], Ryota Kanai[b,1], David Marston[c], Christoph W. Korn[d,e], Geraint Rees[b,f], and Raymond J. Dolan[f]

[a]Department of Cognitive, Perceptual, and Brain Sciences, [b]Institute of Cognitive Neuroscience, and [f]Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, London WC1 H0AP, United Kingdom; [c]Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104-6241; [d]Department of Education and Psychology, Freie Universität, 14195 Berlin, Germany; and [e]Berlin School of Mind and Brain, Humboldt Universität zu Berlin, 10099 Berlin, Germany

Humans form beliefs asymmetrically; we tend to discount bad news but embrace good news. This reduced impact of unfavorable information on belief updating may have important societal implications, including the generation of financial market bubbles, ill preparedness in the face of natural disasters, and overly aggressive medical decisions. Here, we selectively improved people's tendency to incorporate bad news into their beliefs by disrupting the function of the left (but not right) inferior frontal gyrus using transcranial magnetic stimulation, thereby eliminating the engrained "good news/bad news effect." Our results provide an instance of how selective disruption of regional human brain function paradoxically enhances the ability to incorporate unfavorable information into beliefs of vulnerability.

cognition | learning | optimism

We are constantly flooded with information that helps form our beliefs about reality (e.g., via the Web, advertising, colleagues, and friends). Understanding how beliefs are formed is critical, as beliefs drive our actions and decisions. Normative theories assume beliefs are adjusted according to Bayes' Rule (1). Indeed, this assumption often holds when people incorporate favorable news into their existing beliefs (2). However, for unfavorable news people show an aversion to incorporating new information (3). Specifically, they discount the strength of the new information leading to noisy posterior beliefs (2). This tendency to selectively ignore negative information is known as the "good news/bad news effect" (2).

For example, people adjust their beliefs regarding their level of intelligence and physical attractiveness when they receive information indicating they are more intelligent and attractive than they had assumed. However, they relatively fail to adjust their beliefs in response to information suggesting they rate lower on these attributes than they had previously thought (3). In addition, when learning that their risk of experiencing future negative events, such as cancer, is higher than they had expected, people are less likely to integrate these data into their prior beliefs relative to a situation when they learn that their risk is lower than expected (4).

The consequences of readily integrating good news into our beliefs while underweighting bad news are likely to be considerable for an individual and for society. The bias is thought to help explain financial market bubbles (5), ill preparedness in the face of natural disasters (6), overly aggressive medical decisions (7), overconfidence (8), and optimism (4).

Here, we tested whether the tendency to discount bad news could be reversed by altering brain functions thought to mediate such discounting, thereby eliminating the good news/bad news effect. A candidate brain region is the inferior frontal gyrus (IFG), suggested to play a key role in two mechanisms likely to be involved in generating the good news/bad news effect: updating beliefs and inhibition.

First, the IFG is important for flexibly altering beliefs. For example, it plays an important role in reversal learning (9) and has been shown to track and integrate information into prior beliefs (4). Specifically, as the left IFG tracks desirable errors in estimation triggered by information that is better than expected, the right IFG tracks undesirable errors (4). In optimistic individuals the left

IFG shows greater fidelity in tracking errors relative to the right, such that adjustment of beliefs in response to bad news is less likely than in response to good news (4). Previous studies have suggested similar hemispheric asymmetry in processing positive and negative information (10). If relatively efficient function of left IFG (compared with the right) leads to selective updating, than interfering with its function should reduce the bias.

Second, the IFG is thought to mediate different forms of inhibition (for review, see ref. 11). These forms include response/action inhibition (12), task-set switching (13), inhibition of unwanted memories (14), and inhibition of working memory to resolve interference from previous trials (15). Although response inhibition has been suggested to involve primarily the right IFG (11, but see refs. 13 and 16), other forms of inhibition are thought to be lateralized to the left IFG (14) or bilateral IFG (15). If reduced learning from unfavorable news involves active inhibition mediated by the IFG (either left, right, or both), then disruption of its function will release this inhibition, improve learning from bad news, and thus eliminate the bias.

To test these hypotheses we used a causal intervention, using off-line repetitive transcranial magnetic stimulation (TMS) to disrupt the function of either left IFG, right IFG, or a control region (Oz) in three different groups of human participants who then completed a belief formation updating task (4). Participants estimated their likelihood of experiencing 40 adverse life events [e.g., Alzheimer's disease, robbery (4)]. After each trial, participants were presented with the actual frequency of that event in their representative population (Fig. 1A). We assessed whether participants used this factual information to revise their predictions by subsequently asking them, in a second session, to re-estimate their likelihoods for all life events. Participants then rated all stimuli on vividness, familiarity, prior experience, arousal, controllability, and negativity, and were tested for memory of the information presented.

Results

Trials were subsequently divided into ones in which participants received good news (i.e., the probability presented was better than the estimate of their own probability) (Fig. 1B) or bad news (i.e., the probability presented was worse) (Fig. 1C). For each participant, update scores corresponding to the difference between the first and second estimates were calculated for each trial. Then two average update scores were calculated for each participant: one for the good news condition and one for the bad news condition. We entered these average score values into a three (TMS group: right IFG, left IFG, control site) by two (valence: good news/bad news) ANOVA.

Fig. 1. (*A*) In each trial participants were presented with a short description of one of 40 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 living in the same sociocultural environment. The second session was the same as the first session, except that the average probability of the event to occur was not presented again. (*B* and *C*) Examples of trials for which the participant's estimate was (*B*) higher or (*C*) lower than the average probability. Here, for illustration purposes only, the blue and red frames denote the participant's response (either an overestimation or underestimation, respectively) and the blue and red filled boxes denote information that calls for an adjustment in a (*B*) favorable (good news) or (*C*) unfavorable (bad news) direction.

Our analysis revealed a significant interaction between group and valence [$F(2,27) = 6.5$, $P < 0.005$]. Participants who received TMS to the control site or right IFG showed the typical good news/bad news effect; they updated their beliefs more when presented with good news relative to bad news [right IFG: $t(9) = 3.4$, $P < 0.01$, 100% of participants showed the effect; control: $t(9) = 4.2$, $P < 0.01$, 90% of participants showed the effect]. In contrast, the good news/bad news effect was eliminated in participants who received TMS to the left IFG [$t(9) = 0.9$, $P > 0.3$, 40% of participants showed the effect] (Fig. 2), and was significantly reduced compared with right IFG [$F(1,18) = 3.2$, $P < 0.005$] and control group [$F(1,18) = 2.5$, $P < 0.02$]. The result suggests a causal link between left IFG function and the good news/bad news effect.

Note that the critical group by valence interaction was not affected by whether the participant was asked to decide the likelihood of the event happening to them in the future or not happening to them in the future. Specifically, entering average update values into a two (frame: happen, not happen) by three (TMS group: right IFG, left IFG, control site) by two (valence: good news/bad news) ANOVA did not reveal a significant three-way interaction [$F(2,27) = 0.58$, $P > 0.5$].

Next, we examined whether elimination of the good news/bad news effect following TMS to the left IFG was the result of a reduction of belief updating in response to good news, an enhancement in response to bad news, or both. We found that the good news/bad news effect was abolished following TMS to the left IFG because a disposition to adjust beliefs when receiving bad news was enhanced relative to controls [$t(18) = 2.6$, $P < 0.05$]. In contrast, updating in response to good news was preserved [$t(18) = 0.068$ $P > 0.9$]. This finding is consistent with the hypothesis that the left IFG inhibits updating in response to bad news.

Importantly, there were no significant effects on any other measures elicited from the participants. Specifically, no group by valence interaction or group effect emerged when testing participants' ratings of familiarity with stimuli, prior experience, vividness, arousal, negativity, controllability, and reaction times (see Table 1 for scores and statistics). Participants were also asked at the end of the session 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. These errors did not differ between groups and there was no interaction with valence (Table 1). We also calculated the Pearson correlation coefficients between update and the participants' initial estimate for all good news trials and all bad news trials, to test whether TMS was effecting these correlations. These data did not differ between groups and there was no interaction with valence (all $P > 0.1$). Nonetheless, we entered all differential scores (average rating for good news trials minus average rating for bad news trials) of all of the measures above as covariates into an ANCOVA testing update scores with TMS group (right IFG, left IFG, control site) as random between subject effects and valence (good news, bad news) as within-subject effect. This analysis showed that the interaction between group and valence on update scores remained significant even after controlling for all these additional variables [$F(2,17) = 6.43$, $P < 0.01$].

Finally, we examined if initial estimates for either good news trials or bad news trials were at ceiling or floor. One-sample t tests revealed that the average ratings in each group for each valence condition were significantly different from both the minimum and maximum probabilities (see methods, all $P < 0.0001$). Thus, initial estimates were neither at floor nor ceiling. There were no significant difference between groups on initial estimates and no interaction of group and valence (Table 1).

Discussion

Disruption of the left IFG abolished a deep-rooted bias in belief formation known as the good news/bad news effect. This effect refers to the reduced tendency to alter beliefs in response to unfavorable information compared with favorable information. Specifically, interfering with activity of the left IFG using TMS
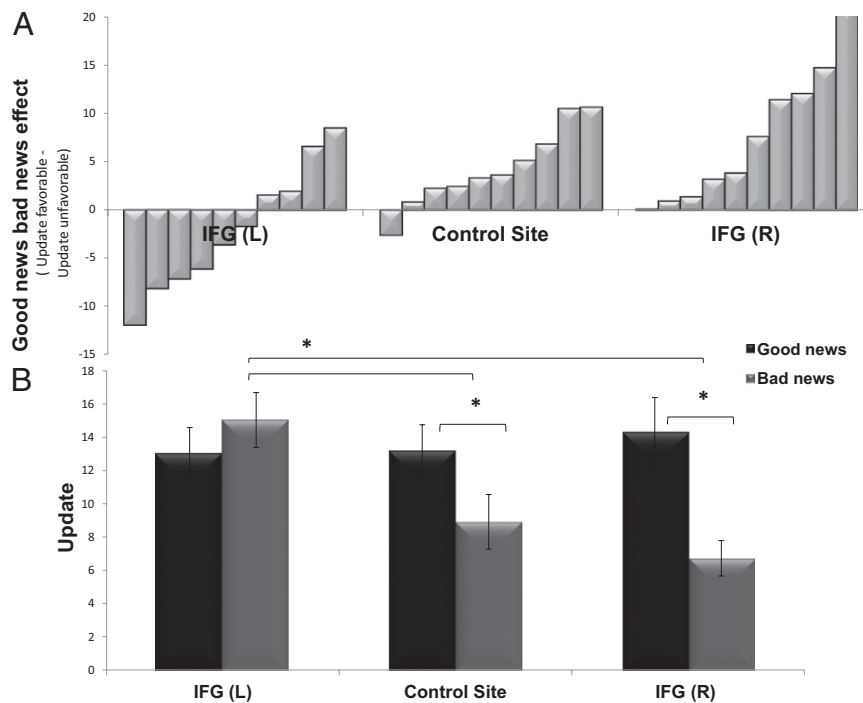
Fig. 2. (*A*) Bar graphs plotting the good news/bad news effect (= update in response to good news − update in response to bad news) for every participant in every group, rank-ordered by effect size. Positive numbers indicate greater belief update in response to good news relative to bad news, negative numbers indicate the reverse. The effect was observed in only 40% of participants who received TMS to the left IFG, in 90% of those who received TMS to the control site, and in 100% of participants who received TMS to the right IFG. (*B*) Mean update for all participants in each TMS condition when receiving good or bad news. Error bars are SEM; *indicates statistical significance at a threshold of $P < 0.05$, two tailed. Participants whose left IFG is disrupted with TMS update their estimates more after receiving bad news. As a consequence, they do not show the good news/bad news effect observed in the other two groups (this reduction is significant relative to the other two groups).

paradoxically enhanced participants' ability to alter beliefs in response to unfavorable information regarding the risk of encountering negative events. The effect on belief formation following TMS to the left IFG was highly selective and did not affect the evaluation of the stimuli presented (including emotional evaluation of the stimuli, sense of familiarity, and vividness) or memory for the information presented.

The enhancement in the inclination to adjust one's beliefs when confronted with bad news following TMS of the left IFG, suggests that this region normally inhibits this ability and that its disruption releases such inhibition. This finding is consistent with the known inhibitory role to the IFG (11). The left IFG has been previously reported to mediate inhibition of unwanted memories (14), inhibition of working memory to reduce interference (14), and inhibition essential for task switching (13) and shifting (17). The present results suggest that the left IFG also plays a role in inhibiting unwanted news from altering beliefs. Suppression of the left IFG may also conceivably render other regions task-active; we speculate that this may include the contralateral right IFG by releasing it from interhemispheric inhibition (18).

Belief change following TMS to the left IFG was specific to adjustments in response to bad news in so far as updating in response to good news was unaltered. A previous functional MRI study using the same update task used here reported error-related activity in response to good news in the left IFG as well as in several other brain regions, including the right and left medial frontal cortex/superior frontal gyrus, and right cerebellum (4). Activity of these latter regions predicted the amount of update in response to good news (4). Thus, although the left IFG plays a role in tracking positive information (4), the present

results suggest that disruption of just this one region is insufficient to significantly influence favorable updates.

Our results should not be interpreted as suggesting that the disruption of left IFG function improves learning or decision-making in general. Rather, our study provides an interesting instance of how a selective disruption to a specific brain region may enhance the tendency to integrate negative information into beliefs regarding vulnerability. Paradoxically, this enhancement may be suboptimal, as underestimating our susceptibility to negative events has adaptive benefits that include increasing explorative behavior and reducing stress and anxiety, a factor that has links with physical and mental well-being (19). However, any advantage arising out of a reduced tendency to learn from bad news is likely to come at a cost. A pertinent example is the discounting of warning signs regarding financial risk, which is widely perceived as a contributing factor to the 2008 global economic collapse (20).

Materials and Methods

Participants. Thirty healthy right-handed participants were recruited via the University College London psychology subject pool (age range, 20–35 y). All subjects gave written informed consent and were paid for their participation. The study was approved by the University College London Research Ethics Committee. Participants were randomly assigned to one of three groups: right IFG TMS, left IFG TMS, or Control Site TMS (10 subjects in each group).

Stimuli. Forty short descriptions of negative life events (e.g., car theft, Alzheimer's disease) were presented in random order. Stimuli were split into two lists of 20 events each.

For each adverse event the average probability of that event occurring at least once to a person living in the same sociocultural environment as the participants was determined from on-line resources (Office for National Statistics, Eurostat, PubMed). Very rare or very common events were not

Table 1. Participants' ratings of familiarity with stimuli, prior experience, vividness, arousal, negativity, controllability, memory, update, initial estimates, and reaction times

Questionnaire and variables	IFG (L) mean (SD)		Control site mean (SD)		IFG (R) mean (SD)		Interaction statistics (group × valence ANOVA)
	Good news	Bad news	Good news	Bad news	Good news	Bad news	F, P
Subjective Scales Questionnaire: All scales 1 = low to 6 = high							
Familiarity	3.54 (0.69)	3.21 (0.59)	4.01 (1.04)	3.67 (0.86)	3.21 (0.87)	2.79 (0.66)	0.06, 0.94
Prior experience	1.72 (0.49)	1.55 (0.25)	1.56 (0.26)	1.37 (0.27)	1.68 (0.35)	1.54 (0.45)	0.04, 0.95
Vividness	3.44 (0.70)	3.26 (0.50)	3.79 (1.00)	3.14 (1.02)	3.12 (0.66)	2.98 (0.55)	3.30, 0.05
Emotional arousal	2.57 (0.81)	2.64 (0.78)	3.20 (1.06)	3.02 (0.89)	3.12 (0.93)	3.21(1.17)	1.13,0.34
Negativity	4.16 (0.94)	4.22 (0.93)	4.44 (0.80)	4.26 (0.50)	4.55 (0.92)	4.52 (0.91)	0.76, 0.47
Controllability	3.02 (0.68)	3.11 (0.65)	2.83 (0.59)	3.23 (0.70)	3.14 (0.51)	3.13 (0.52)	1.07, 0.35
Task-related variables							
Memory errors	13.00 (5.35)	12.06 (3.42)	12.04 (6.88)	11.34 (5.01)	15.08 (9.08)	12.15 (3.72)	0.45, 0.63
Initial estimates	44.47 (4.47)	21.71 (4.05)	46.62 (7.40)	19.88 (2.93)	46.09 (4.90)	23.86 (5.93)	1.07, 0.35
Update	13.04 (4.87)	15.04 (5.22)	13.19 (4.92)	8.91 (5.19)	14.30 (6.59)	6.71 (3.37)	6.5, 0.005
Update in the "not happen" condition	13.11 (6.26)	12.57 (7.08)	15.51 (7.68)	9.36 (8.48)	17.51 (7.86)	3.5 (7.35)	3.32, 0.05
Update in the "happen" condition	12.89 (5.35)	16.36 (7.38)	9.72 (6.46)	8.94 (5.75)	9.85 (8.28)	8.08 (7.18)	0.74, 0.48
Reaction time first estimate (ms)	2,802 (805.20)	2,844 (1063.61)	2,490 (937.60)	2,545 (820.48)	2,516 (1108.44)	2,655 (994.31)	0.10, 0.90
Reaction time second estimate (ms)	2,371 (629.10)	2,431 (651.50)	2,156 (1,150.41)	2,020 (1,123.60)	2,127 (986.81)	2,141 (1,168.45)	0.41, 0.66

included; all events 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%.

Procedure. Participants completed the task 5 min after the 40-s off-line continuous θ-burst stimulation (cTBS) protocol (see details below).

The paradigm was adapted from our previous study (4). On each trial participants were presented with 1 of 40 adverse life events for 4 s. Then participants estimated how likely the event was to happen to them in the future. Participants had up to 10 s to respond. After a fixation cross (1.2 s), they were presented with the actuarial frequency of the event, for a demographically similar population, for 3 s, followed by fixation for 1.2 s (Fig. 1). In a second session participants were asked to again provide estimates of their likelihood of encountering the events, so that we could assess how they used the information provided. Update was calculated for trials in which participants received good news (when a participant received an estimate that was better than expected; update = first estimate − second estimate) and trials in which they received bad news (when a participant received an estimate that was worse than expected; update = second estimate − first estimate). Thus, positive updates indicate a change toward the actuarial frequency and negative updates a change away from the actuarial frequency.

The extent of the good news/bad news effect is equal to the difference between update in respond to good news minus the amount of update in respond to bad news:

$$Good\ news\ bad\ news\ effect = update\ in\ response\ to\ good\ news \\ - update\ in\ response\ to\ bad\ news \quad [1]$$

The second session was the same as the first, except that the average probability of the event to occur was not presented again.

Trials were divided into two blocks. Each block consisted two sessions, with each session involving 20 trials. In one block (counterbalanced) participants estimated their likelihood as described above. In the other block they estimated their likelihood of not encountering the adverse events and received the frequency of the events not happening, such that results could not be attributed to differential processing of high or low numbers. 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.

To test memory for the information presented, subjects were asked at the end of the session 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.

$$memory\ error = |actual\ probability\ presented \\ - recollection\ of\ probability\ presented|. \quad [2]$$

Participants then rated all stimuli on: 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); negativity (How negative would this event be for you? From 1 = not negative at all to 6 = very negative); and controllability (Is this event under your control? From 1 = not at all to 6 = very much).

TMS. TMS pulses were delivered by a Magstim Rapid2 Stimulator (Magstim) at 40% of the maximum stimulator output using a small TMS coil (figure-of-eight shape, 50-mm diameter). We used an off-line continuous cTBS protocol, which consisted of three pulses at 50 Hz repeated at 200-ms intervals for a total duration of 40 s. A 5-min rest period was implemented after the termination of cTBS before participants started the task.

Target sites were selected based on a previous functional MRI paper (4). Specifically, we targeted the peak voxel in a region where blood oxygen-level dependent signals correlate with participant's estimation error in response to negative information about the future (right IFG) and the region in which activity correlated with estimation errors in response to positive information (left IFG). The target sites for individual participants were identified by converting the Talairach coordinates (x/y/z = ± 4 8/18/16) into real space for individual participants. In this procedure, structural MRI scans were first transformed to a standard Montreal Neurological Institute (MNI) template using the FMRIB Software Library. The coordinates for IFG sites are marked in the MNI space and then inverse transformed to the original image space. The markers for the target sites were used to guide frameless stereotaxy with the Brainsight system (Rogue Research). In addition to the two IFG sites, Oz position of the

International 10-20 system was used as a control site for nonspecific effects of TMS.

Behavioral Analysis. Two average update scores were calculated for each participant; one for the good news condition and one for the bad news condition. Those average scores were entered into a two (valence: good news, bad news) by three (group: left IFG, control, right IFG) ANOVA.

We then proceeded to conduct ANOVAs for each additional measure elicited (i.e., memory, reaction times, and the six stimuli ratings: vividness, familiarity, controllability, experience, arousal, negativity) to assure that they did not show an interaction. We also calculated for each subject the correlation between initial estimates and update for good news trials and bad news trials and entered those into an ANOVA.

Finally, to control for all these variables, we conducted an additional analysis in which the differential scores of all additional ratings described above were added as covariates in an ANCOVA entering update scores as a dependent variable, with valence as a within subject variable (valence: good news, bad news) and group as a between subject random effects variable (TMS group: IFG left, IFG right, control site).

1. Green D-M, Swets J-A (1966) *Signal Detection Theory and Psychophysics* (John Wiley and Sons, New York).
2. Eli D, Rao J-M (2012) The Good News-Bad News Effect: Asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics* 3:114138.
3. Koszegi B (2006) Ego utility, overconfidence, and task choice. *J Eur Econ Assoc* 4: 673–707.
4. Sharot T, Korn C-W, Dolan R-J (2011) How unrealistic optimism is maintained in the face of reality. *Nat Neurosci* 14:1475–1479.
5. Farmer R (1999) *The Macroeconomics of Self-fulfilling Prophecies* (MIT Press, Cambridge, MA).
6. Paton D (2003) Disaster preparedness: A social-cognitive perspective. *Disaster Prevention and Management* 12(Iss: 3):210–216.
7. Paling J (2003) Strategies to help patients understand risks. *BMJ* 327:745–748.
8. Johnson D-D, Fowler J-H (2011) The evolution of overconfidence. *Nature* 477:317–320.
9. Cools R, Clark L, Owen AM, Robbins TW (2002) Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. *J Neurosci* 22:4563–4567.
10. Davidson RJ (1995) Cerebral asymmetry, emotion and affective style. *Brain Asymmetry*, eds Davidson RJ, Hugdahl K (MIT Press, Cambridge, MA), pp 361–387.
11. Aron AR, Robbins TW, Poldrack RA (2004) Inhibition and the right inferior frontal cortex. *Trends Cogn Sci* 8:170–177.
12. Bunge SA, Dudukovic NM, Thomason ME, Vaidya CJ, Gabrieli JD (2002) Immature frontal lobe contributions to cognitive control in children: Evidence from fMRI. *Neuron* 33:301–311.
13. Dove A, Pollmann S, Schubert T, Wiggins CJ, von Cramon DY (2000) Prefrontal cortex activation in task switching: An event-related fMRI study. *Brain Res Cogn Brain Res* 9: 103–109.
14. Anderson MC, et al. (2004) Neural systems underlying the suppression of unwanted memories. *Science* 303:232–235.
15. D'Esposito M, Postle BR, Jonides J, Smith EE (1999) The neural substrate and temporal dynamics of interference effects in working memory as revealed by event-related functional MRI. *Proc Natl Acad Sci USA* 96:7514–7519.
16. Konishi S, et al. (1999) Common inhibitory mechanism in human inferior prefrontal cortex revealed by event-related functional MRI. *Brain* 122:981–991.
17. Nakahara K, Hayashi T, Konishi S, Miyashita Y (2002) Functional MRI of macaque monkeys performing a cognitive set-shifting task. *Science* 295:1532–1536.
18. Suppa A, et al. (2008) Theta burst stimulation induces after-effects on contralateral primary motor cortex excitability in humans. *J Physiol* 586:4489–4500.
19. Taylor S-E, Kemeny M-E, Reed G-M, Bower J-E, Gruenewald T-L (2000) Psychological resources, positive illusions, and health. *Am Psychol* 55:99–109.
20. Shefrin H (2010) *Voices of Wisdom: Understanding the Global Financial Crisis.* ed. Siegel LB pp. 224–256 (Research foundation of CFA Institute, Charlottesville, VA) Available at http://ssrn.com/abstract=1523931.