Edinburgh Research Explorer

Trust in information, political identity and the brain

OPEN ACCESS

Author for correspondence:
Laura Cram
e-mail: laura.cram@ed.ac.uk

THE ROYAL SOCIETY
PUBLISHING

Trust in information, political identity and the brain: an interdisciplinary fMRI study

Adam Moore[1], Sujin Hong[2] and Laura Cram[2]

[1]Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK
[2]Neuropolitics Research Lab, School of Social and Political Science, University of Edinburgh, 18 Buccleuch Place, Edinburgh EH8 9LN, UK

AM, 0000-0002-7002-5460; SH, 0000-0001-6957-7746; LC, 0000-0003-4775-171X

Misinformation has triggered government inquiries and threatens the perceived legitimacy of campaign processes and electoral outcomes. A new identity polarization has arisen between Remain and Leave sympathizers in the UK Brexit debate, with associated accusations of misinformation use. Competing psychological accounts of how people come to accept and defend misinformation pit self-reinforcing motivated cognition against lack of systematic reasoning as possible explanations. We harness insights from political science, cognitive neuroscience and psychology to examine the impact of trust and identity on information processing regarding Brexit in a group of Remain identifiers. Behaviourally, participants' affective responses to Brexit-related information are affected by whether the emotional valence of the message is compatible with their beliefs on Brexit (positive/negative) but not by their trust in the source of information. However, belief in the information is significantly affected by both (dis)trust in information source and by belief compatibility with the valence of the information. Neuroimaging results confirm this pattern, identifying areas involved in judgements of the self, others and automatic processing of affectively threatening stimuli, ultimately supporting motivated cognition accounts of misinformation endorsement.

This article is part of the theme issue 'The political brain: neurocognitive and computational mechanisms'.

1. Introduction

There is increasing concern about social media's magnification of misinformation, misleading information and conspiracy theories (hereafter misinformation) and its impact on democracy (e.g. [1–3]; for reviews, see [4,5]). Psychological accounts of misinformation acceptance differ as to whether effortful cognition is primarily involved in the acceptance or the rejection of misinformation. Accounts on the acceptance end of the spectrum (self-reinforcing (SR) accounts) propose that misinformation acceptance is largely the product of alignment with an individual's preferred worldview/ideology/prior beliefs. When such alignment occurs, individuals may selectively attend to and process information compatible with what they already believe [6–8]. Alternatively, they may selectively search for perceived weaknesses in belief-incompatible information to dismiss it, even at the expense of truth ([9]; cf. [10]). Effortful rejection (ER) accounts propose that misinformation is typically accepted on the basis of superficial characteristics and shallow processing (cf. [11]), but when reasoning is engaged, people can reliably discriminate fake from real news regardless of ideological compatibility ([12]; though right-wing conservatives show greater susceptibility to misinformation than left-wing liberals—[13,14]; but see [15]). A recent identity-based hybrid (I-bH) account proposes that contextually salient goals control where effort is deployed, with self/identity/ideology-reinforcement weighed against desires for epistemic accuracy [16]. This model posits key roles for orbitofrontal cortex (OFC) and

ventromedial prefrontal cortex (vmPFC) in computing competing goals and the value of differing beliefs, respectively, which can then bias downstream processing.

Neuroimaging evidence is broadly compatible with these accounts. Areas associated with affect/emotion, value computation and conflict processing, such as lateral and medial OFC and anterior cingulate cortex (ACC), show increased activity when political partisans are confronted with information potentially damaging to their preferred candidate [17] and when trying to control affective reactions while explicitly evaluating socially relevant stimuli (e.g. abortion, murder; [18]). Similarly, vmPFC is more active when pre-existing beliefs bias reasoning performance [19] and such activity predicts increasing belief in given information [20]. This is consistent with the I-bH model, and to some extent with SR frameworks. By contrast, and supporting both ER and I-bH accounts, disbelieving information activates distributed brain regions involved in effortful cognition, including the left inferior frontal gyrus (IFG), bilateral anterior insula (AI), dorsal ACC extending to superior frontal gyrus (SFG), and superior parietal lobe [20], while right lateral PFC activation corresponds to correct logical inference in the face of competing beliefs [19].

However, this research does not systematically disentangle the affective impact of the belief (in)compatibility of information itself versus the (dis)trust of the source of that information. In other words, how do people respond when (in)compatible information is delivered via a (dis)trusted source? This is critical because source identity can impact information acceptance regardless of actual content (e.g. [21]), and such identity cues can trigger motivated cognition ([8]; see also [22]). However, the affective impact of belief-incompatible information (without source identity) can drive a disconnection with external information and increased processing in default-mode networks [23]. Here we investigate how information source and belief compatibility *combine* to affect judgements of belief and feelings about politically salient information. Neuroimaging is useful in disentangling predictions made by different frameworks: SR accounts predict that trusted sources and compatible information should result in greater belief and greater activation in areas associated with effortful cognition (e.g. lateral PFC, ACC), and a mismatch between (dis)trusted source and (in)compatible information may elicit effortful cognition related to decreased belief (i.e. motivated rejection). ER accounts predict less activity in these areas in response to trusted/compatible stimuli, but possibly more in response to mismatch stimuli (which potentially provide cues triggering reasoning processes), though it is unclear what relation effortful reasoning may have to belief in this case. Under the I-bH account we can assume that our participants are motivated in large part, though not completely, by identity protection concerns related to their self-identification as Remain voters (on which they were pre-screened; see Methods). This suggests predictions largely in line with SR accounts, though vmPFC activity related to judging various beliefs (identity protection and truth/accuracy) against one another may emerge.

2. Methods

(a) Participants

Thirty-eight, right-handed, UK national, native English-speaking participants aged 18–42 (mean = 25.5, 20 females) underwent blood oxygenation level-dependent (BOLD) contrast fMRI scanning. Three were excluded owing to failure to record button responses. All were Brexit Remain supporters familiar with social media platforms. Pre-screening was conducted separately to avoid affecting behaviour. All participants provided informed consent and were paid £20. A debrief followed the fMRI scans. The study was approved by the Research Ethics Committees of the School of Social and Political Sciences and the Research Ethics procedures of the Edinburgh Imaging facility at The Queen's Medical Research Institute (EIF-QMRI) at the University of Edinburgh.

(b) Stimuli

We extracted text from tweets related to Brexit (posted between 1 January 2019 and 14 August 2019) gathered from various media sources, and separately we pre-screened all of our participants to determine a bespoke list of their three most-used and least-used media sources (based on the British social attitudes 2018 questionnaire, Q216, (https://www.bsa.natcen.ac.uk/media/39286/questionnaire_2018_v2.pdf), and an open-choice option; see electronic supplementary material, table S1). In the initial stage, two reviewers rated tweet texts as Negative, Positive or Ambiguous. From the Negative and Positive rated texts, we selected short texts, preferably one-line tweets and, where necessary, edited these to remove source information. The remaining tweets were validated in a pilot experiment by separate Remain-voting participants (i.e. not the fMRI sample; electronic supplementary material, Methods, figure S2; A full list of 120 tweet texts (60 negative, 60 positive) used in the present study is given in electronic supplementary material, table S4).

Negative valence tweets (i.e. regarding substantial costs/consequences) on Brexit are belief-compatible for Remain voters who oppose Brexit, while Positive valence tweets (i.e. regarding benefits/rewards) are belief-incompatible. These factors, information source (Trusted, Distrusted) and emotional valence/belief compatibility (Negative/Compatible, Positive/Incompatible), were fully crossed in a 2 × 2 ANOVA design (figure 1b,c).

(c) Task and fMRI design

Participants completed three practice trials before confirming that they understood the task. We used a rapid-event-related design. During the task a fixation cross was shown for the initial 20 s. The tweets and accompanying logos were presented for 2.5 s; two rating tasks followed, measuring belief and feeling, counterbalanced across participants. For the belief rating task, the statement 'I believe this is true' was presented with a 7-point scale from 'Not at all (1)' to 'Completely (7)' (reversed for 50% of participants). For feeling ratings, the phrase 'This makes me feel' was presented with a 7-point scale from 'Negative (1)' to 'Positive (7)' (reversed for 50% of participants). Cursor position on the scales always initiated in the middle. Inter-trial interval (ITI) was jittered 2–4 s. Participants used left and right thumb buttons to move the cursor to the left and to the right, respectively, on the response scales, and pressed the right index button to finalize their response (figure 1).

(d) fMRI apparatus and MRI data acquisition

Participants viewed the visual stimuli via a mirror attached on a head coil; we used PsychoPy (v. 3) to present stimuli [24]. Two response grips were used (NordicNeuroLab). We acquired imaging data of high resolution T1-weighted three-dimensional anatomical images (TR 2500 ms, TE 4.37 ms, 256 × 256 mm 100% field of view (FOV), 1 mm slice thickness), gradient-echo fieldmaps (same slice as the echo-planar imaging (EPI) images, resolution 3 × 3 mm, TR 599 ms, TE_1 5.19 ms, TE_2 7.65 ms, flip angle 60°, bandwidth 260, TA 1 m in 15 s), and gradient EPI functional images (TR 2260 ms, TE 27 ms, 192 mm 100% FOV, in-plane resolution
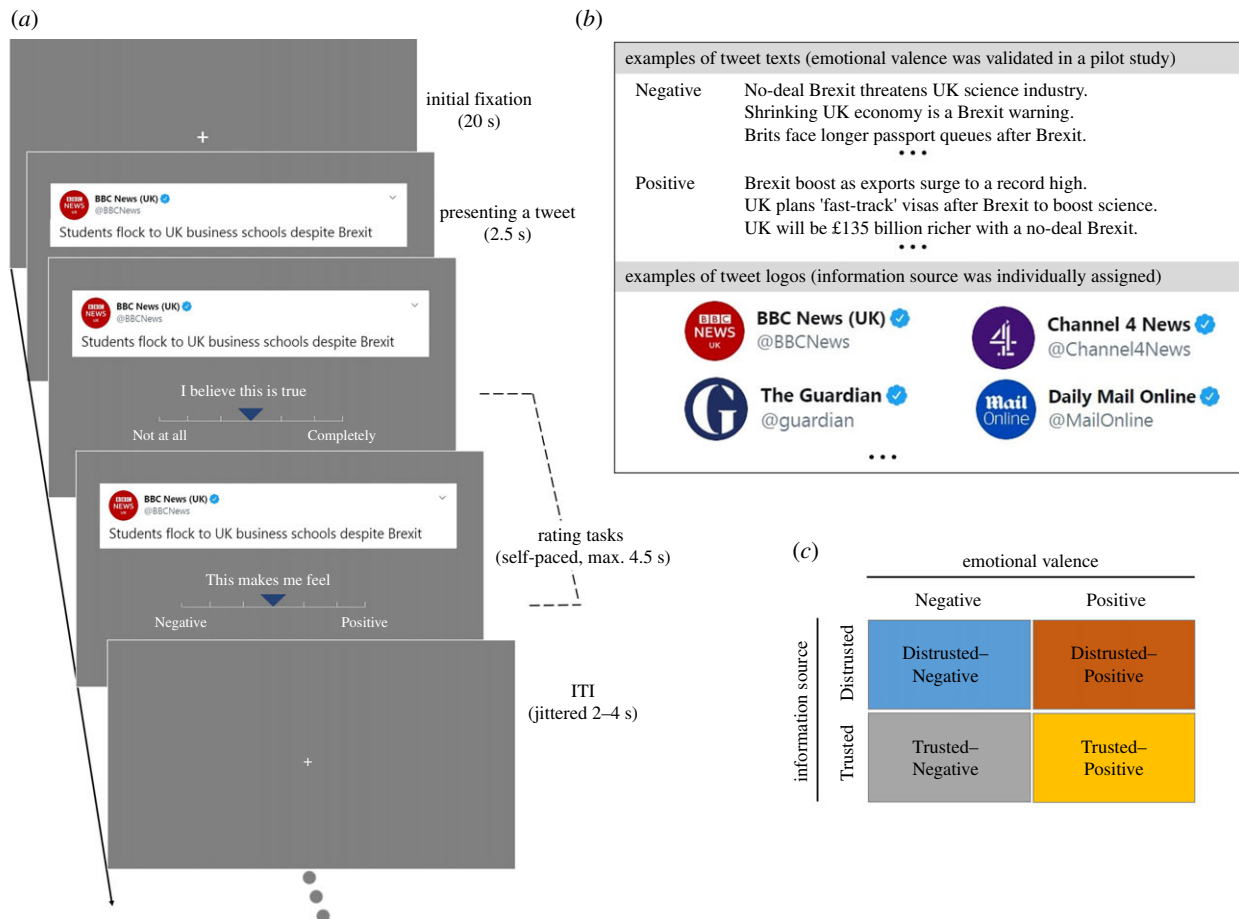
Figure 1. fMRI task and design. (*a*) Schematic of a single trial during fMRI scan. (*b*) Examples of tweet texts and logos. (*c*) A 2×2 repeated measures ANOVA design (emotional valence × information source).

3×3 mm, 3 mm slice thickness, no gap, flip angle 80°, slice order = Siemens ascending interleaved), using a 3 T Siemens Skyra MRI scanner (Siemens Healthineers, Erlangen, Germany).

(e) fMRI data analysis

fMRI data were analysed using Statistical Parametric Mapping (SPM12 (v 7771), Wellcome Trust Centre for Neuroimaging, London, UK) and MATLAB 2019b (MathWorks). In pre-processing, the first two functional images were discarded to ensure T1 saturation. Remaining functional images were corrected for slice-timing, realigned, distortion-corrected via gradient field-map (Realign & Unwarp of SPM12 using a voxel displacement map, [25,26]) and co-registered to the corresponding individual T1-weighted anatomical images. The anatomical images were segmented, with parameters used to normalize functional images with the Montreal Neurological Institute (MNI) brain template. Normalized functional data were smoothed with a three-dimensional isotropic Gaussian kernel 6 mm full width at half maximum. For statistical processing, the general linear model (GLM) was used at first-level analysis. Presentation of tweet, rating of belief score and rating of feeling score were modelled as events for each condition (2 levels of 2 factors) using the canonical haemodynamic response function (HRF). Additional regressors of no interest included the temporal derivative of each GLM predictor, high-pass filter (128 s) regressors and six rigid body transformation parameters from spatial realignment. Contrasts of parameter estimates (greater than fixation (not explicitly modelled)) of each event were used to compute contrast images at the first level, which were then entered into the second-level GLM model of a two-way repeated measures (2×2) flexible factorial ANOVA for the whole brain. We performed separate ANOVAs during the first presentation of a tweet, belief rating

task, and feeling rating task. Statistical threshold was set at $p < 0.001$ (uncorrected) at the voxel level, and with a cluster-level extent threshold of $p < 0.05$ (false discovery rate, FDR) applied for multiple corrections. Cluster parameter estimates included all voxels, extracted using the Marsbar toolbox [27].

3. Results

(a) Behavioural results

Rating scores and reaction times (RTs) collected during fMRI scans were standardized and analysed using R (v. 4.0.2), brms (v. 2.13.5) and STAN. All regression models sampled from four chains with 1000 trial burn-in and 6000 iterations each, using regularizing, weakly informative Gaussian priors ($\mu = 0$, $\sigma = 1$). All chains converged as indicated by both caterpillar plots and $\hat{R}$ estimates (all = 1.00). Intervals are 95% highest posterior density intervals (HDI) unless otherwise stated.

As belief and feeling ratings were negatively correlated, $r = -0.19$, HDI = $[-0.22, -0.16]$, we performed a 2×2 repeated measures Bayesian multilevel multivariate regression, with fixed effects of information source (Trusted, Distrusted) and valence (Negative, Positive) and random intercepts for participants and stimuli. Trusted sources increased belief ratings, $b = 0.22$, HDI = $[0.15, 0.29]$, but did not appreciably affect feeling ratings, $b = 0.01$, HDI = $[-0.05, 0.07]$. Positive valence/Belief-Incompatible information decreased belief ratings, $b = -0.79$, HDI = $[-0.97, -0.61]$, and increased positive feeling ratings, $b = 1.17$, HDI = $[1.05, 1.29]$. There was no meaningful

Table 1. fMRI ANOVA results during the belief rating task. R, right hemisphere.

brain regions	MNI coordinates			Z-score	cluster size (k)
	x	*y*	*z*		
main effect of information source					
R middle occipital gyrus	48	−76	−4	4.41	21
interaction: information source × emotional valence					
R cuneus and precuneus	21	−82	38	4.11	34

interaction for belief ratings, $b = -0.01$, HDI = [−0.10, 0.09], nor for feeling ratings, $b = 0.04$, HDI = [−0.05, 0.12].

Since RTs were also correlated ($r = 0.04$, HDI = [0.01, 0.07]), we repeated the above analysis. Predicting belief RTs revealed negligible effects of Trusted sources, $b = -0.04$, HDI = [−0.12, 0.04], Positive valence/Belief-Incompatible information, $b = -0.04$, HDI = [−0.15, 0.05], and no interaction, $b = 0.05$, HDI = [−0.07, 0.16]. For feeling RTs, the pattern was similar to source, $b = -0.06$, HDI = [−0.14, 0.02], Positive valence/Belief Incompatibility, $b = 0.10$, HDI = [−0.01, 0.19], and the interaction, $b = 0.11$, HDI = [−0.01, 0.20], though the latter two effects provide some suggestion that participants took slightly longer to give feeling ratings to Belief-Incompatible information, particularly when it came from a Trusted source (for descriptive statistics, see supplementary material, table S2).

(b) fMRI results

A two-way repeated measures ANOVA of whole brain during the belief rating task showed a significant main effect of the information source in the middle occipital gyrus (MOG) of the right hemisphere (table 1, MNI coordinates: $x = 48$, $y = -76$, $z = -4$; Z-score = 4.41), and a significant interaction in the cuneus and precuneus in the right hemisphere (figure 2*a*, MNI coordinates: $x = 21$, $y = -82$, $z = 38$; Z-score = 4.11), where the parameter estimates of Distrusted–Positive and Trusted–Negative were significantly greater than Distrusted–Negative and Trusted–Positive, respectively (figure 2*b*). There was a negative correlation, $r = -0.33$, HDI = [−0.66, −0.01], between parameter estimates in the cuneus/precuneus and belief ratings for Belief-Incompatible information from a Distrusted source (i.e. Distrusted source–Positive valence; figure 2*c*), but no equivalent result for compatible information from a Trusted source, $r = 0.12$, HDI = [−0.22, 0.48].

An identical analysis on the feeling rating task showed significant activations in the main effect of valence/belief compatibility across fronto-parietal–occipital lobes (figure 3, electronic supplementary material, table S3 and figure S1), which are largely clustered and include bilateral superior/middle/inferior occipital gyrus, bilateral postcentral gyrus, superior frontal gyrus, superior parietal gyrus, supplementary motor areas extending to middle cingulate gyrus, right putamen/insula, right Rolandic operculum extending to Heschl's gyrus, right precentral gyrus extending to superior fontal gyrus, angular gyrus and supramarginal gyrus. There was no significant activation for the main effect of information source nor interaction.

4. Discussion

Our participants rated their belief in, and feelings about, politically salient tweets. Behaviourally both (dis)trust in information source and emotional valence/belief compatibility affected belief ratings, but only valence/belief compatibility affected feeling ratings. RTs were slightly longer when rating feelings about Positive valence/Belief-Incompatible tweets. The fMRI analyses of the feeling rating task revealed only a main effect of valence/belief compatibility in a wide network with clustered regions related to multiple cognitive processes, including language processing in the frontal, temporal and parietal regions, motor movements in the precentral and postcentral gyrus, and visual processing in the occipital areas in both hemispheres. By contrast, when rating *belief* in our political tweets, there was a significant main effect of information source in right MOG and a significant interaction effect in right cuneus and precuneus. That interaction effect, driven by the negative correlation with belief ratings for Distrusted Positive-valenced stimuli, suggests additional processing occurred for information matching source expectations (i.e. Distrusted sources providing Belief-Incompatible information).

While these results are limited, and somewhat unexpected given our initial hypotheses centred on PFC activity, they seem most compatible with the I-bH [16] and SR accounts [10,28,29] of misinformation acceptance. Emotion and attention are known to heighten sensitivity to visual cues [30], and recent neurocomputational work suggests that emotion-related activity in visual cortex, such as we found in the belief rating task, may represent an interface of sensory representations of the environment and previous knowledge [31]. Encoding of high-arousal negative information is associated with activity in the MOG, an area that may cooperate with closely related areas (e.g. posterior fusiform and inferior occipital gyrus) in a network specializing in identifying emotionally important visual clues [32]; for example, both MOG and cuneus activity are involved in inferring threat from non-facial cues/body language [33]. Given the identity-relevant nature of the information presented in our study, the activation we observe in MOG and cuneus is also consistent with middle occipital and cuneus involvement in both judgements and attitudes about others and the self [34,35]. Similarly, in close parallel to our findings, the cuneus is involved in information processing when invalid cues appear (i.e. when information is detected to be misleading/wrong) and even more so when they are the result of human intention (versus preprogrammed stimuli; [36]), as well as when people disbelieve unpalatable political information relative to non-political information [23]. People often deliberately scrutinize or attack incompatible and/or distrusted information more than they carefully consider appealing belief-compatible information [9], which is the pattern of activation we find in (pre)cuneus (figure 2*b*). The lack of significant PFC activation for belief-compatible information (from any source) seems consistent with this, and inconsistent with a role for careful reasoning in our task, though we cannot draw strong inferences from the lack of an effect in this context. It may well be that our participants, selected for strong Brexit views, were immediately 'on guard' when our stimuli were Brexit-related. Such identity-based motivated scrutiny is likely somewhat automatic, with respect to core identity beliefs.

Indeed, this type of motivated rejection effect is also observed when participants refuse to update beliefs about
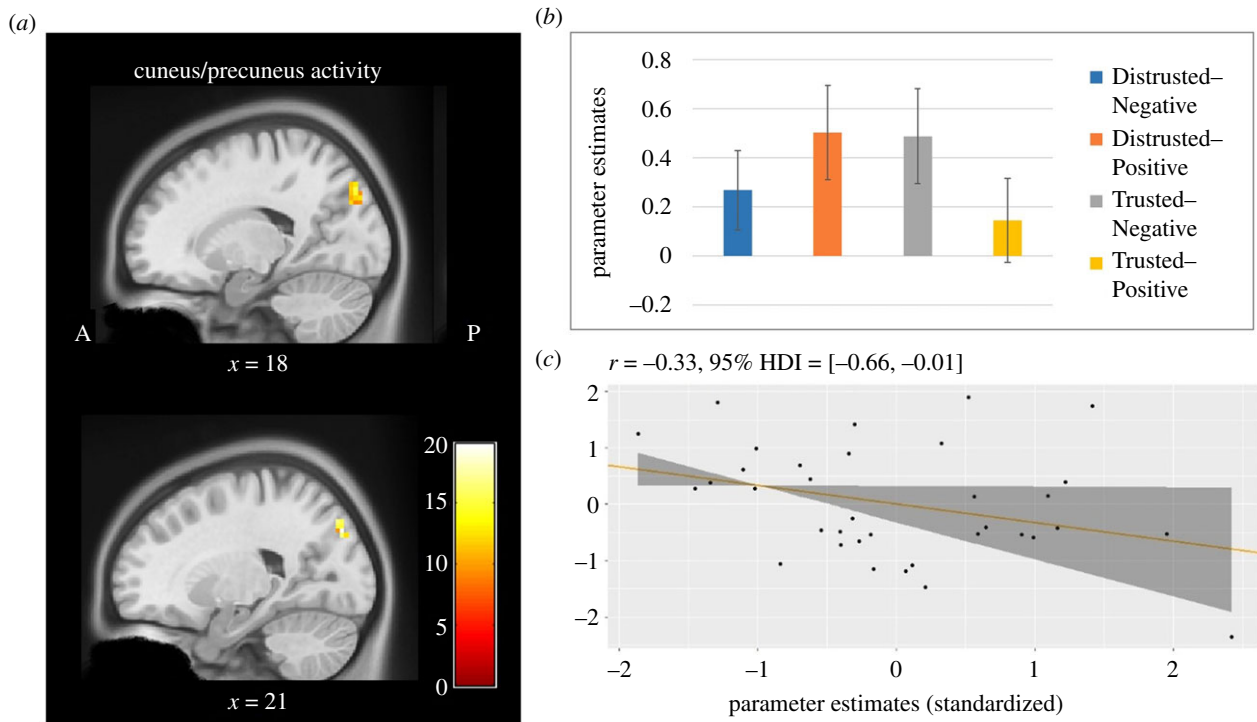
Figure 2. Significant interaction between emotional valence and information source during the belief rating task (*a*) Cuneus and precuneus activity in the right hemisphere. (*b*) Parameter estimates of the cuneus/precuneus activity (error bars denote ±s.e.m.). (*c*) Correlation (and 95% HDI) between parameter estimates and belief ratings (Distrusted–Positive).
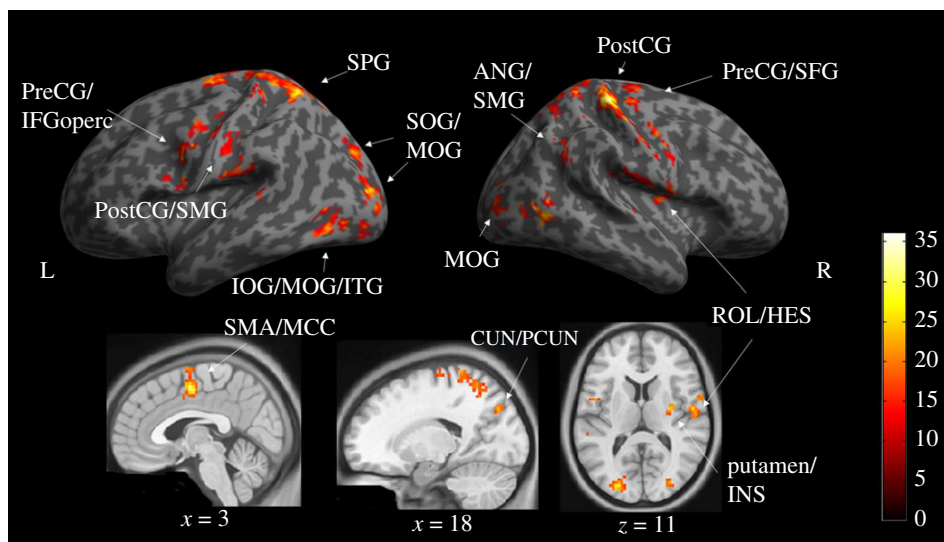
Figure 3. The main effect of emotional valence on neural activation during the feeling rating task (see also electronic supplementary material, table S3 and figure S1). PreCG, precentral gyrus; IFGoperc, inferior frontal gyrus, opercular part; PostCG, postcentral gyrus; SMG, supramarginal gyrus; SMA, supplementary motor area; MCC, middle cingulate gyrus; ROL, rolandic operculum; HES, heschl's gyrus; SPG, superior parietal gyrus; SOG, superior occipital gyrus; IOG, inferior occipital gyrus; ITG, inferior temporal gyrus; MTG, middle temporal gyrus; MOG, middle occipital gyrus; ANG, angular gyrus; INS, insula; CUN, cuneus; PCUN, precuneus.

the self or future finances when information is negative/undesirable [37]. Many also reject overwhelming scientific consensus on a variety of topics by positing complex, though absurd, conspiracy theories—essentially attacking the source of unpalatable information (e.g. [38,39]), but typically only when that information presents a challenge to their pre-existing worldview [10]. Such reactions, though relying on some amount of deliberate cognition, are typically not products of systematic reasoning, but rather affective reactivity [40] that is largely effortless [41], though it can sometimes be the result of rational calculation to discredit information or sources that oppose or hinder one's goals [42,43]. Our results

suggest that participants integrated information regarding source trustworthiness into their belief ratings, and when confronted with belief-incompatible information from a distrusted source, believed it less, possibly as a function of differentially critical processing.

5. Conclusion

Under motivated cognition accounts of misinformation acceptance (SR and I-bH), people can actively incorporate belief-compatible information and critically scrutinize

incompatible information [9]. In the latter case, people do not necessarily systematically reason about strengths and weakness of offending information, but rather critique to reject it and protect the self (cf. [44]). Remain and Leave identities, prevalent in the aftermath of the UK's EU referendum, are personally important and cut across traditional party lines, generating affective polarization as intense as that of partisanship across a range of measures [45]. Our results seem to corroborate the view of motivated rejection of belief-incompatible information. Belief incompatibility itself and beliefs about trustworthiness of sources may be combined to motivate rejection of politically offensive information when that information is a threat to a deeply held/self-defining ideological belief.

References

1. Allcott H, Gentzkow M. 2017 Social media and fake news in the 2016 election. *J. Econ. Persp.* **31**, 211–236. (doi:10.1257/jep.31.2.211)

2. Gaughan AJ. 2017 Illiberal democracy: the toxic mix of fake news, hyperpolarization, and partisan election administration. *Duke J. Constit. Law Publ. Policy* **12**, 57–139. See https://scholarship.law.duke.edu/djclpp/vol12/iss3/3.

3. Hansen I, Lim DJ. 2019 Doxing democracy: influencing elections via cyber voter interference. *Contemp. Politics* **25**, 150–171. (doi:10.1080/13569775.2018.1493629)

4. Colley T, Granelli F, Althuis J. 2020 Disinformation's societal impact: Britain, Covid, and beyond. *Defence Strat. Commun.* **8**, 89–140. (doi:10.30966/2018.RIGA.8)

5. Tucker JA, Guess A, Barbera P, Vaccari C, Siegel A, Sanovich S, Stukal D, Nyhan B. 2018 Social media, political polarization, and political disinformation: a review of the scientific literature. *SSRN*, 21 Mar 2018. (doi:10.2139/ssrn.3144139)

6. Kahan DM, Peters E, Wittlin M, Slovic P, Ouellette P, Braman D, Mandel G. 2012 The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nat. Clim. Change* **2**, 732–735. (doi:10.1038/nclimate1547)

7. Knobloch-Westerwick S, Mothes C, Polavin N. 2020 Confirmation bias, ingroup bias, and negativity bias in selective exposure to political information. *Commun. Res.* **47**, 104–124. (doi:10.1177/0093650217 719596)

8. Peterson MB, Skov M, Serritzlew S, Ramsoy T. 2013 Motivated reasoning and political parties: evidence for increased processing in the face of party cues. *Polit. Behav.* **35**, 831–854. (doi:10.1007/s11109-012-9213-1)

9. Ditto PH, Lopez DF. 1992 Motivated skepticism: use of differential decision criteria for preferred and non-preferred conclusions. *J. Pers. Social Psychol.* **63**, 568–584. (doi:10.1037/0022-3514.63.4.568)

10. Lewandowsky S, Oberauer K. 2016 Motivated rejection of science. *Curr. Dir. Psychol. Sci.* **25**, 217–222. (doi:10.1177/0963721416654436)

11. Pennycook G, Cheyne JA, Barr N, Koehler DJ, Fugelsang JA. 2015 On the reception and detection of pseudo-profound bullshit. *Judg. Decision Making* **10**, 549–563. (doi:10.1037/t01307-000)

12. Pennycook G, Rand DG. 2019 Lazy, not biased: susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition* **188**, 39–50. (doi:10.1016/J.cognition.2018.06.011)

13. Baron J, Jost JT. 2019 False equivalence: are liberals and conservatives in the United States equally biased? *Perspect. Psychol. Sci.* **14**, 292–303. (doi:10.1177/1745691618788876)

14. Jost JT. 2017 Ideological asymmetries and the essence of political psychology. *Polit. Psychol.* **38**, 167–208. (doi:10.1111/pops.12407)

15. Ditto PH, Liu BS, Clark CJ, Wojcik SP, Chen EE, Grady RH, Celniker JB, Zinger JF. 2019. At least bias is bipartisan: a meta-analytic comparison of partisan bias in liberals and conservatives. *Persp. Psychol. Sci.* **14**, 273–291. (doi:10.1177/1745691617746796)

16. Van Bavel JJ, Pereira A. 2018 The partisan brain: an identity-based model of political belief. *Trends Cogn. Sci.* **22**, 213–224. (doi:10.1016/j.tics.2018.01.004)

17. Westen D, Blagov PS, Harenski K, Kilts C, Hamann S. 2006 Neural bases of motivated reasoning: an fMRI study of emotional constraints on partisan political judgment in the 2004 U.S. Presidential election. *J. Cogn. Neurosci.* **18**, 1947–1958. (doi:10.1162/jocn.2006.18.11.1947)

18. Cunningham WA, Raye CL, Johnson MK. 2004 Implicit and explicit evaluation: fMRI correlates of valence, emotional intensity, and control in the processing of attitudes. *J. Cogn. Neurosci.* **16**, 1717–1729. (doi:10.1162/0898929042947919)

19. Goel V, Dolan RJ. 2003 Explaining modulation of reasoning by belief. *Cognition* **87**, B11–B22. (doi:10.1016/S0010-0277(02)00185-3)

20. Harris S, Shet SA, Cohen MS. 2008 Functional neuroimaging of belief, disbelief, and uncertainty. *Ann. Neurol.* **63**, 141–147. (doi:10.1002/ana.21301)

21. Cohen G. 2003 Party over policy: the dominating impact of group influence on political beliefs. *J. Pers. Social Psychol.* **85**, 808–822. (doi:10.1037/0022-3514.85.5.808)

22. Cram L, Moore A, Olivieri V, Suessenbach F. 2018 Fair is fair, or is it? Territorial identity triggers influence ultimatum game behavior. *Polit. Psychol.* **39**, 1233–1250. (doi:10.1111/pops.12543)

23. Kaplan JT, Gimbel SI, Harris S. 2016 Neural correlates of maintaining one's political beliefs in the face of counterevidence. *Scient. Rep.* **6**, 39589. (doi:10.1038/srep39589)

24. Peirce JW, Gray JR, Simpson S, MacAskill MR, Höchenberger R, Sogo H, Kastman E, Lindeløv J. 2019 PsychoPy2: experiments in behavior made easy. *Behav. Res. Methods* **51**, 195–203. (doi:10.3758/s13428-018-01193-y)

25. Jezzard P, Balaban RS. 1995 Correction for geometric distortion in echo planar images from B0 field variations. *Magn. Reson. Med.* **34**, 65–73. (doi:10.1002/mrm.1910340111)

26. Hutton C, Bork A, Josephs O, Deichmann R, Ashburner J, Turner R. 2002 Image distortion correction in fMRI: a quantitative evaluation. *Neuroimage* **16**, 217–240. (doi:10.1006/nimg.2001.1054)

27. Brett M, Anton JL, Valabregue R, Poline JB. 2002 Region of interest analysis using an SPM toolbox. In *8th Int. Conf. Functional Mapping of the Human Brain, Sendai, Japan, 2–6 June 2002. NeuroImage* **16**, abstr. 497. See http://matthew.dynevor.org/research/abstracts/marsbar/marsbar_abstract.pdf.

28. Kahan D. 2013 Ideology, motivated reasoning, and cognitive reflection. *Judg. Decision Making* **8**, 407–424. (doi:10.2139/ssrn.2182588).

6

royalsocietypublishing.org/journal/rstb   *Phil. Trans. R. Soc. B* **376**: 20200140

29. Kahan D. 2017 Misconceptions, misinformation, and the logic of identity-protective cognition. *SSRN*, 24 May 2017. (doi:10.2139/ssrn.2973067)

30. Lane RD, Chua PM-L, Dolana RJ. 1999 Common effects of emotional valence, arousal and attention on neural activation during visual processing of pictures *Neuropsychologia* **37**, 989–997. (doi:10.1016/S0028-3932(99)00017-2)

31. Kragel PA, Reddan MC, LaBar KS, Wager TD. 2019 Emotion schemas are embedded in the human visual system. *Sci. Adv.* **5**, eaaw4358. (doi:10.1126/sciadv.aaw4358)

32. Mickley Steinmetz KR, Addis DR, Kensinger EA. 2010 The effect of arousal on the emotional memory network depends on valence. *Neuroimage* **53**, 318–324. (doi:10.1016/j.neuroimage.2010.06.015)

33. Kret ME, Pichon S, Grèzes J, de Gelder B. 2011 Similarities and differences in perceiving threat from dynamic faces and bodies. An fMRI study. *Neuroimage* **54**, 1755–1762. (doi:10.1016/j.neuroimage.2010.08.012)

34. Bruneau EG, Saxe R. 2010 Attitudes towards the outgroup are predicted by activity in the precuneus in Arabs and Israelis. *Neuroimage* **52**, 1704–1711. (doi:10.1016/j.neuroimage.2010.05.057)

35. Ochsner KN, Knierim K, Ludlow DH, Hanelin J, Ramachandran T, Glover G, Mackey SC. 2004 Reflecting upon feelings: an fMRI study of neural systems supporting the attribution of emotion to self and other. *J. Cogn. Neurosci.* **16**, 1746–1772. (doi:10.1162/0898929042947829)

36. Ozdem C, Wiese E, Wykowska A, Muller H, Brass M, van Overwalle F. 2017 Believing androids – fMRI activation in the right temporo-parietal junction is modulated by ascribing intentions to non-human agents. *Social Neurosci.* **12**, 582–593. doi:10.1080/17470919.2016.1207702

37. Coutts A. 2019 Good news and bad news are still news: experimental evidence on belief updating. *Exp. Econ.* **22**, 369–395. (doi:10.1007/s10683-018-9572-5)

38. Diethelm P, McKee M. 2009 Denialism: what is it and how should scientists respond? *Eur. J. Publ. Health* **19**, 2–4. (doi:10.1093/eurpub/ckn139)

39. Lewandowsky S, Cook J, Oberauer K, Brophy S, Lloyd EA, Marriott M. 2015 Recurrent fury: conspiratorial discourse in the blogosphere triggered by research on the role of conspiracist ideation in climate denial. *J. Social Polit. Sci.* **3**, 142–178. (doi:10.5964/jspp.v3i1.443)

40. Van Prooijen JW, Douglas KM. 2018 Belief in conspiracy theories: basic principles of an emerging research domain. *Eur. J. Social Psychol.* **48**, 897–908. (doi:10.1002/ejsp.2530)

41. Hughes BL, Zaki J. 2015 The neuroscience of motivated cognition. *Trends Cogn. Sci.* **19**, 62–64. doi:10.1016/j.tics.2014.12.006

42. Cook J, Lewandowsky S. 2016 Rational irrationality: modeling climate change belief polarization using Bayesian networks. *Top. Cogn. Sci.* **8**, 160–179. (doi:10.1111/tops.12186)

43. Dunlap RE, Jacques PJ. 2013 Climate change denial books and conservative think tanks: exploring the connection. *Am. Behav. Sci.* **57**, 1–33. (doi:10.1177/0002764213477096)

44. Amit A, Mentser S, Arieli S, Porzycki N. In press. Distinguishing deliberate from systematic thinking. *J. Pers. Social Psychol.* (doi:10.1037/pspp0000284)

45. Hobolt S, Leeper T, Tilley J. In press. Divided by the vote: affective polarization in the wake of the Brexit referendum. *Br. J. Polit. Sci.* (doi:10.1017/S0007123420000125)

46. Engelmann JB, Meyer F, Ruff CC, Fehr E. 2019 The neural circuitry of affect-induced distortions of trust. *Sci. Adv.* **5**, eaau3413. (doi:10.1126/sciadv.aau3413)

47. Rolls ET, Huang CC, Lin, C-P, Feng J, Joliot M. 2020 Automated anatomical labelling atlas 3, *Neuroimage* **206**, 116189. (doi:10.1016/j.neuroimage.2019.116189)