Brief Communications

Necessary, Yet Dissociable Contributions of the Insular and Ventromedial Prefrontal Cortices to Norm Adaptation: Computational and Lesion Evidence in Humans

Xiaosi Gu,[1,2]* Xingchao Wang,[3,4]* Andreas Hula,[1] Shiwei Wang,[3,4] Shuai Xu,[3,4] Terry M. Lohrenz,[2] Robert T. Knight,[5,6] Zhixian Gao,[3,4] Peter Dayan,[7] and P. Read Montague[1,2,8]

[1]Wellcome Trust Centre for Neuroimaging, University College London, London, United Kingdom WC1N 3BG, [2]Human Neuroimaging Laboratory, Virginia Tech Carilion Research Institute, Roanoke, Virginia 24016, [3]Department of Neurosurgery, Beijing Tiantan Hospital, Capital Medical University, and [4]China National Clinical Research Center for Neurological Diseases, Beijing, China 100050, [5]Helen Willis Neuroscience Institute and [6]Department of Psychology, University of California, Berkeley, California 94720, [7]Gatsby Computational Neuroscience Unit, University College London, London, United Kingdom WC1N 3AR, and [8]Department of Physics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061

Social norms and their enforcement are fundamental to human societies. The ability to detect deviations from norms and to adapt to norms in a changing environment is therefore important to individuals' normal social functioning. Previous neuroimaging studies have highlighted the involvement of the insular and ventromedial prefrontal (vmPFC) cortices in representing norms. However, the necessity and dissociability of their involvement remain unclear. Using model-based computational modeling and neuropsychological lesion approaches, we examined the contributions of the insula and vmPFC to norm adaptation in seven human patients with focal insula lesions and six patients with focal vmPFC lesions, in comparison with forty neurologically intact controls and six brain-damaged controls. There were three computational signals of interest as participants played a fairness game (ultimatum game): sensitivity to the fairness of offers, sensitivity to deviations from expected norms, and the speed at which people adapt to norms. Significant group differences were assessed using bootstrapping methods. Patients with insula lesions displayed abnormally low adaptation speed to norms, yet detected norm violations with greater sensitivity than controls. Patients with vmPFC lesions did not have such abnormalities, but displayed reduced sensitivity to fairness and were more likely to accept the most unfair offers. These findings provide compelling computational and lesion evidence supporting the necessary, yet dissociable roles of the insula and vmPFC in norm adaptation in humans: the insula is critical for learning to adapt when reality deviates from norm expectations, and that the vmPFC is important for valuation of fairness during social exchange.

Key words: **brain lesion; computational modeling; decision-making; insular cortex; social norms; ventromedial prefrontal cortex**

Introduction

Social norms and their enforcement are fundamental to societies. Violations of any prevailing norm must be detected so that they can be acted upon. This process requires an individual to be able to: (1) represent a shared norm about what is expected, (2) detect deviations from the norm, and (3) correct for norm deviations and adapt to norms (Montague and Lohrenz, 2007). It is believed that our nervous system has the "apparatus" and "mechanisms" to implement such complex social computations (Montague and Lohrenz, 2007; Rilling et al., 2008): for instance, the insular and ventromedial prefrontal cortices (vmPFC) have been shown to be activated during social decision-making in recent functional magnetic resonance imaging (fMRI) studies (Sanfey et al., 2003; Chang and Sanfey, 2013).

At least two issues remain unresolved. First, it is not clear what dissociable roles the insula and vmPFC serve in norm adaptation. The insula has been established as an interoceptive cortex, providing a bodily map for a wide range of mental processes (Craig, 2013; Critchley and Harrison, 2013; Gu et al., 2013a). Insular activity correlate swith predictions and prediction errors of value (Seymour et al., 2004) and risk (Preuschoff et al., 2008). We previously found that insular activation correlates with both deviations from norms and deviations from expected feelings, suggesting a role of the insula in linking feelings with decisions

Received July 15, 2014; revised Nov. 1, 2014; accepted Nov. 3, 2014.

Author contributions: X.G., Z.G., and P.R.M. designed research; X.G., X.W., S.W., S.X., and Z.G. performed research; X.G., A.H., R.T.K., and P.D. analyzed data; X.G., X.W., A.H., T.L., R.K., P.D., and P.R.M. wrote the paper.

This work was funded by a Principal Research Fellowship from The Wellcome Trust (P.R.M.), The Kane Family Foundation (P.R.M.), and NINDS Grant 2R37NS21135 (R.T.K.). X.W. and Z.G. are supported by National Natural Science Foundation of China (Grant No. 81328008). P.D. is funded by the Gatsby Charitable Foundation. We thank Jae Shin for help with preparing the task and IT support, Clay Clayworth for help with lesion reconstruction, Ming Hsu for helpful discussions, and Cathy Price for comments on the paper.

The authors declare no competing financial interests.

*X.G. and X.W. contributed equally to this work.

This article is freely available online through the *J Neurosci* Author Open Choice option.

Correspondence should be addressed to: Dr P. Read Montague, Virginia Tech Carilion Research Institute, 2 Riverside Circle, Roanoke, VA 24011. E-mail: read@vt.edu.

DOI:10.1523/JNEUROSCI.2906-14.2015

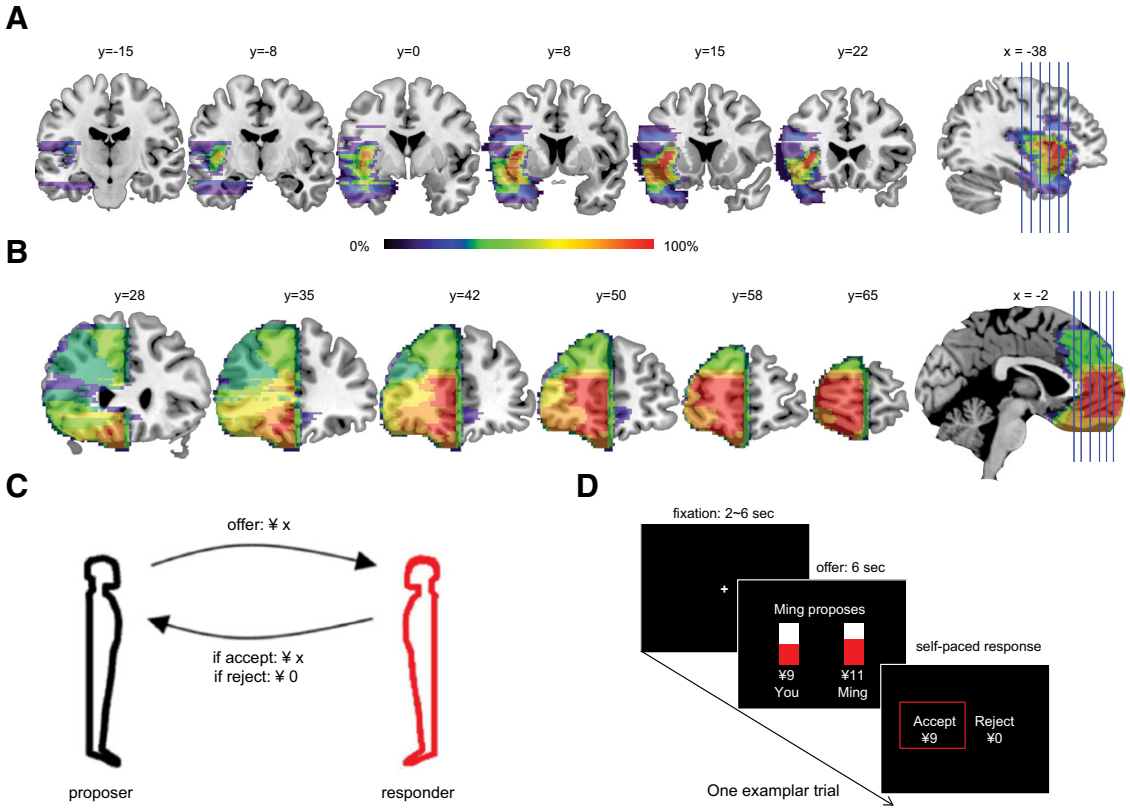

Figure 1. ***A***, Display of reconstruction of insula lesions ($n = 7$). ***B***, Display of reconstruction of vmPFC lesions ($n = 6$). Color bar represents the degree of lesion overlap among patients (0–100%). ***C***, A version of the Ultimatum task. Participants made "accept/reject" responses to an offer of s_i (of ¥20 Chinese Yuan). ***D***, Time course of one trial. Each participant played 45 rounds in total with a different virtual partner every round.

(Xiang et al., 2013). The vmPFC has been associated with valuation during value-based decision-making in both social (Behrens et al., 2008) and nonsocial (Rushworth et al., 2012) domains. It has also been shown to coactivate with the insula in norm processing (Xiang et al., 2013).

Second, there is little evidence supporting the necessity of the insula and vmPFC in norm adaptation. Existing fMRI evidence showing insula and vmPFC responses to norms are largely correlational. One recent study using transcranial direct brain stimulation (tDCS) suggests a causal role of right dorsolateral prefrontal cortex in norm compliance (Ruff et al., 2013). However, neither the insula nor vmPFC can be directly targeted with current brain stimulation methods such as tDCS due to their anatomical constraints. Thus, no causal evidence exists supporting the roles of the insula and vmPFC in norm adaptation.

Here, we address the hypothesis that the insula and vmPFC are necessary for dissociable processes during norm adaptation. To test this hypothesis, we used computational modeling of a fairness game in patients with either focal insula lesions or vmPFC lesions. We predicted that insula lesions would lead to failure in the ability to adapt to social norms, whereas vmPFC lesions would result in deficits in valuation in social situations.

Materials and Methods

Participants

We examined seven patients with focal lesions in the insula (4 females and 3 males) and six patients with vmPFC lesions (3 females and 3 males; see Fig. 1*A*,*B* for lesion display and Table 1 for all participants' characteristics), compared with 40 neurologically intact controls (NCs; 21 females and 19 males). Six patients with lesions other than the insula and vmPFC (i.e., temporal lobe lesions) were recruited as brain-damaged controls (BDCs; 2 females and 4 males). All lesions resulted from surgical removal of low-grade gliomas. All subjects had normal color vision, and reported no previous or current neurological or psychiatric conditions. All patients (insula, vmPFC, BDC) were recruited from the Patient's Registry of Beijing Tiantan Hospital, Beijing, China. Neurologically intact controls were recruited in local Beijing communities and were matched with patients for age, sex, and education. Informed consent was obtained for all research involving human participants. All procedures were approved by the Institutional Review Boards of Virginia Tech in Blacksburg, Virginia, and the Beijing Tiantan Hospital of Capital Medical University in Beijing.

Patients did not differ from neurologically intact controls in age, sex, and education (all p values >0.05). Additionally, all patients were considered cognitively intact, measured by Mini-Mental Status Exam (MMSE; Folstein et al., 1975), a measurement of cognitive impairment

Table 1. Participant characteristics

	NC ($n = 40$)	BDC ($n = 6$)	Insula ($n = 7$)	vmPFC ($n = 6$)
Age (years)*	48 ± 12	42 ± 11	42 ± 8	54 ± 13
Sex	21 f/19 m	2 f/4 m	4 f/3 m	3 f/3 m
Handedness	39 R/1 L	6 R/0 L	7 R/0 L	6 R/1 L
Education (years)	11 ± 2	12 ± 3	13 ± 3	9 ± 3
MMSE	29 ± 2	28 ± 2	28 ± 2	26 ± 3
BDI	3 ± 4	3 ± 4	2 ± 1	4 ± 4
Chronicity (months)†	—	22 ± 13	18 ± 21	15 ± 2
Lesion laterality	—	3 L/3 R	4 L/3 R	3 L/2 R/1 B
Lesion size (ml)	—	42 ± 20	32 ± 13	112 ± 44

L, Left; R, right; B, bilateral; f, female; m, male. MMSE, Mini-Mental Status Exam (Folstein et al., 1975). BDI, Beck Depression Inventory (Knight, 1984). Data are represented as mean ± SD.

*The age at the testing date.

†The time length between surgery date and testing date.

although vmPFC patients' MMSE scores were slightly lower than the NC subjects ($p < 0.05$). All other patients' MMSE scores did not significantly differ from controls ($p > 0.05$). Patients did not show alteration in Beck Depression Inventory (BDI), a measurement of baseline mood (Knight, 1984) compared with NC participants ($p > 0.05$).

Lesion reconstruction

Lesion reconstruction was performed by a neurologist (R.T.K.), who was blind to the behavioral results. In brief, lesions evident on T1- and T2-weighted MRI were identified and transcribed onto corresponding sections of a template to create a volume of interest file. The template was derived from a digital MRI volume of a normal control (ch2.nii) created by Christopher Rorden (University of South Carolina, Columbia, SC) and provided for use with MRIcron (http://www.cabiatl.com/mricro/mricro/index.html). This was used to measure the location (MNI coordinates) and volume (in ml) of individual lesions and to create within group overlaps of multiple lesions using the MRIcron program.

Task and procedure

All participants played the role of the responder in the ultimatum game (Fig. 1*C*,*D*), with a different virtual proposer each round, for a total of 45 rounds. In each round, the subjects were first offered a split of ¥20 Chinese Yuan. The offer screen lasted 6 s. Next the subjects were presented with the choice options and were allowed as much time as required to respond. The offers were predetermined: 6 × ¥1, 6 × ¥2, 6 × ¥3, 6 × ¥4, 6 × ¥5, 3 × ¥6, 3 × ¥7, 3 × ¥8, 3 × ¥9, 3 × ¥10, presented in a randomized order. Participants received a fixed amount of payment plus a bonus of the outcome of one of the 45 rounds, drawn randomly.

Logistic regression

We first fitted the each subject's binary reject/accept response against offer size using a logistic regression as follows (mnrfit in MATLAB, MATLAB and Statistics Toolbox Release 2012b, MathWorks):

$$ln\left(\frac{\pi_{reject}}{\pi_{accept}}\right) = \beta_0 + \beta_1 \cdot \text{offer}.$$

π_{reject} denotes the probability of rejecting an offer, and π_{accept} represents the probability of accepting an offer. β_1 indicates the influence of each unit change in offer size on the change of the probability of reject (relative to the probability of accept), and hence, represents sensitivity to fairness.

Norm adaptation models

The logistic regression defines subjects' behavior as a function of offer size, but does not inform us about their underlying learning structure. Therefore, we conducted model-based analysis based on the observation that subjects playing the ultimatum game have an internal representation of norms and have the ability to adapt to changing environments (Montague and Lohrenz, 2007; Xiang et al., 2013). Specifically, we assume that the subject's behavior can be modeled by individual aversion to unequal splits. The responder's utility on iteration or round i of the exchange can be represented using the Fehr–Schmidt (FS) inequality aversion utility as follows (Fehr and Schmidt, 1999):

$$V_i(s_i) = s_i - \alpha \max\{f_i - s_i, 0\}.$$

Here, α represents sensitivity to norm prediction error ("envy", $\alpha \in [0,1]$), that is, the subject's unwillingness to accept an offer s_i below the reference value f_i ("internal norm"). Sensitivity to advantageous norm prediction errors ("guilt") was not modeled because the experiment setup only included disadvantageous offers (≤ 10).

Given $V_i(s_i)$, we model the probability of accepting an offer as follows:

$$p_i(s_i) = \frac{e^{\gamma V_i(s_i)}}{1 + e^{\gamma V_i(s_i)}}.$$

Here γ is the inverse temperature parameter of the softmax. The lower is γ, the more diffuse and variable are the choices ($\gamma \in [0,1]$).

We considered two classes of learning models governing the evolution of the internal norm f_i: a Rescorla–Wagner (RW) model based on temporal difference learning (Rescorla and Wagner, 1972; Sutton and Barto, 1998) and a Bayesian observer model (Xiang et al., 2013). Both models assume the internal norm f_i evolves as a function of observed offers, but differ in the updating rule. We also considered two ways of setting the initial norm f_0 for both models: in the fixed initial norm condition, $f_0 = 10$; in the variable norm condition, f_0 was fitted individually to each subjects' data ($f_0 \in [0,20]$). A non-learning FS model with fixed norm $f = 10$ and two parameters ($\gamma \in [0,1]$ and $\alpha \in [0,1]$) was also considered as a baseline model. All models were fitted at the individual subject's level. We estimated parameters multiple times per subject with varying starting points to avoid the possibility of being stuck in local minima. Because these two classes of models have very distinct assumptions of the underlying cognitive process in the subject, and that it was unknown to us which mechanism was more plausible, we performed model comparison using a modified Bayesian information criterion score (BIC) for studies with small numbers of observations (Haughton, 1988), which is an approximation of the model evidence: the model with the lowest BIC score has the highest model evidence. BIC scores were normalized based on sample size.

Rescorla–Wagner norm adaptation models. These models use the RW rule (Rescorla and Wagner, 1972; Sutton and Barto, 1998) to update an individual's internal norm:

$$f_i = f_{i-1} + \varepsilon(s_i - f_{i-1}).$$

Here the norm adaptation rate $\varepsilon \in [0,1]$, determines the extent to which the internal norm is influenced by the immediately preceding offer. A high rate ε means high impact of offers on internal norms, while a low rate ε indicates unwillingness to adapt. The use of ε inside the inequality term could potentially create problems for our estimation, given high ε or low α. Therefore, we compared the α values from learning models with those from the FS model. We find the values to be close for both models; hence, the adaption rate ε significantly improved likelihoods but only slightly altering the values of α. We expected that subjects with low α and low ε should show low rejection, little influence by the immediately preceding offer, and consistent rejection behavior; subjects with low α and high ε should have lowest rejection overall, but proportionally more rejections for low offers following high offers, and fluctuating rejection behavior; those with high α and low ε should have high rejection, little influence by immediate preceding offer, and consistent rejection behavior; people with high α and high ε should show low rejection for high offers following low offers, high rejection for low offers following high offers, and overall fluctuating rejection behavior.

Bayesian norm adaption models. Here we assume that the subjects model the offers from a normal distribution with uncertain mean and variance. The subjects start with a prior on the mean μ and variance σ^2, and update it as offers are made. More specifically, the subject assumes that conditional on μ and σ^2, offers X are given by $X \mid \mu,\sigma^2 \sim N(\mu,\sigma^2)$. If we assume a prior $p(\mu,\sigma^2)$ on μ and σ^2 after observing the first offer s_i we have the posterior as follows:

$$p(\mu,\sigma^2|s_i) = p(s_i|\mu,\sigma^2)\,p(\mu,\sigma^2).$$

Concretely we assume a conjugate prior:

$$p(\mu,\sigma^2) = P_1(\mu|\sigma^2)P_2(\sigma^2), \text{ with}$$

$$p_1(\mu|\sigma^2) = NormalDensity(\mu|\mu_0, \sigma^2/k_0),$$

$$p_2 = Inv - \chi^2(\sigma^2/\nu_0).$$

After observing s_i, the posteriors are of the same form but with parameters given by the recursive formulae:

$$k_{i+1} = k_i + 1,\ \nu_{i+1} = \nu_i + 1$$

$$\mu_{i+1} = \mu_i + \frac{1}{k_{i+1}}(s_i - \mu_i)$$

$$\nu_{i+1}\sigma^2_{i+1} = \nu_i\sigma^2_i + \frac{k_i}{k_{i+1}}(s_i - \mu_i)^2.$$

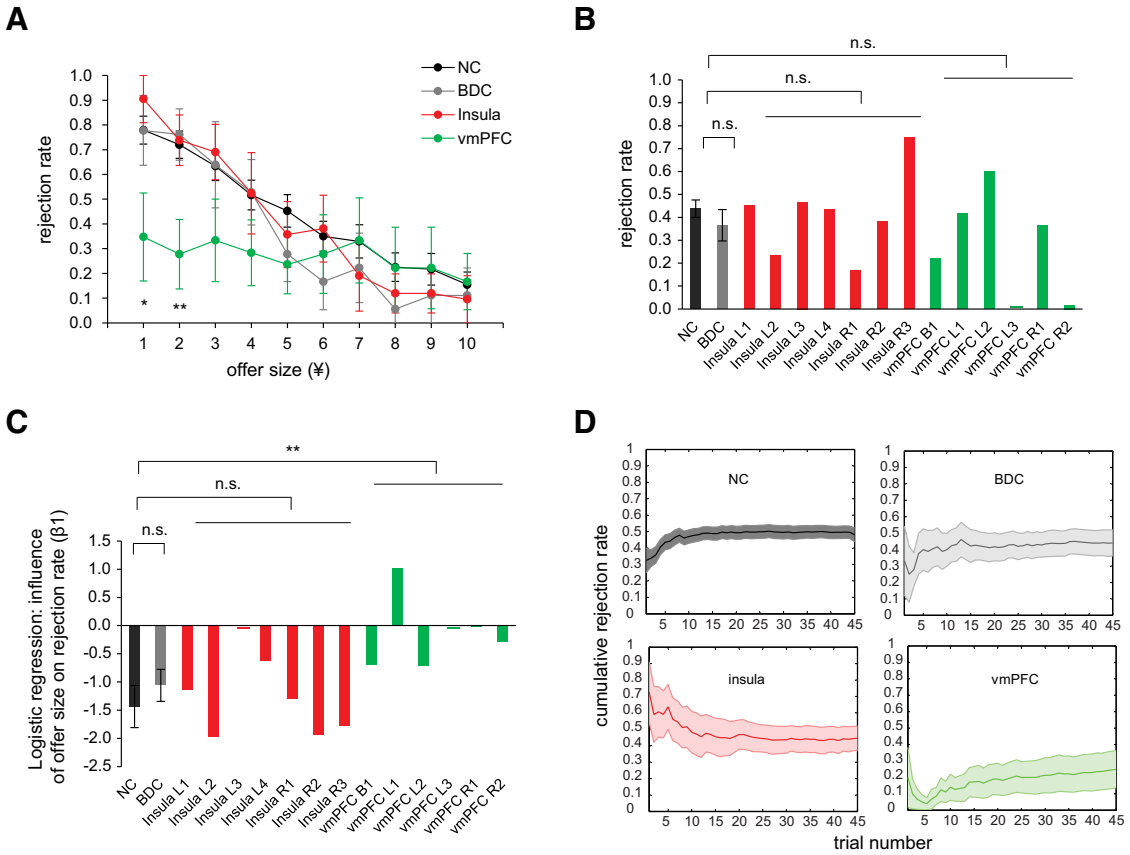

Figure 2. ***A***, Rejection rate at each offer size. Patients with vmPFC lesions rejected less in the most unfair conditions (¥1 and ¥2). ***B***, There were no significant group differences in overall rejection rate. ***C***, The influence of offer size on rejection calculated by a logistic regression model was significantly reduced in patients with vmPFC lesions. ***D***, Trajectories of cumulative rejection rate; *$p < 0.05$; **$p < 0.01$; n.s., not significant. L, Left; R, right; B, bilateral. Data are represented as mean ± SEM.

On trial $i + 1$ the prevailing norm is

$$E_i[X] = \int xp(x|\mu,\sigma^2)\, p_1(\mu|\sigma^2,\, \mu_i,k_i)\, p_2(\sigma^2|v_i) = \mu_i,$$

so that the utility of the offer at trial $i + 1$ is then given by the following:

$$V_i(s_i) = s_i - \alpha \max\{\mu_i - s_i, 0\}.$$

and the probability of accepting the offer is as follows:

$$\Pr(accept) = \frac{1}{1 + \exp(-\gamma V(s_i))}.$$

Statistical method

Because the current dataset does not meet the assumptions of parametric tests and all comparisons were based on a priori hypotheses in small samples, we used a bootstrapping method (Gu et al., 2012) to assess the probability of observing a difference between two groups. We consider the difference between two groups significant if the probability of obtaining the actual t statistic is <5% along the permutated (10,000 iterations) distribution of t statistics $p < 0.05$ (one-tailed).

Results

Rejection rate and sensitivity to fairness

Figure 2*A* illustrates the rejection rate by offer size for each group. Patients with vmPFC lesions rejected much less than the NC group for the most unfair offer sizes of ¥1 (bootstrapping $p < 0.05$) and ¥2 ($p < 0.01$). There was no other group difference at other offer sizes ($p > 0.1$). Although patients with vmPFC lesions showed marginally lower overall rejection rate than the NC group ($p = 0.09$), there was no significant group difference in overall raw rejection rate for any patient group relative to NC subjects (Fig. 2*B*; $p > 0.1$).

Figure 2*C* shows the β coefficient β_1 from the logistic regression, representing the influence of offer size on rejection rate. Patients with vmPFC lesions had significantly lower β coefficient than NC subjects ($p < 0.01$), suggesting that the influence of offer size on rejection rate was reduced in these patients. Neither patients with insula lesions, nor the BDC patients showed such a pattern ($p > 0.1$). These results suggest that patients with vmPFC lesions, but not those with insula lesions, showed reduced sensitivity to offer size.

To better understand how subjects' behavior evolved over time, we further examined subjects' cumulative response. Figure 2*D* demonstrates the averaged cumulative rejection rate: patients with insula lesions showed a relatively flattened cumulative rejection curve; in contrast, vmPFC patients had a continued increase of cumulative rejection rate.

Together, these results demonstrate that patients with vmPFC lesions, but not those with insula lesions, showed reduced sensitivity to offer size, especially when the offers were highly unfair. These findings provide a baseline comparison for the computational models presented below and suggest a critical role of the vmPFC in valuation in the realm of social norms.

Rescorla–Wagner norm adaptation model parameters

We further demonstrate the patients' deficits based on parameters from the RW model with variable initial norms (Fig. 3*A–E*),

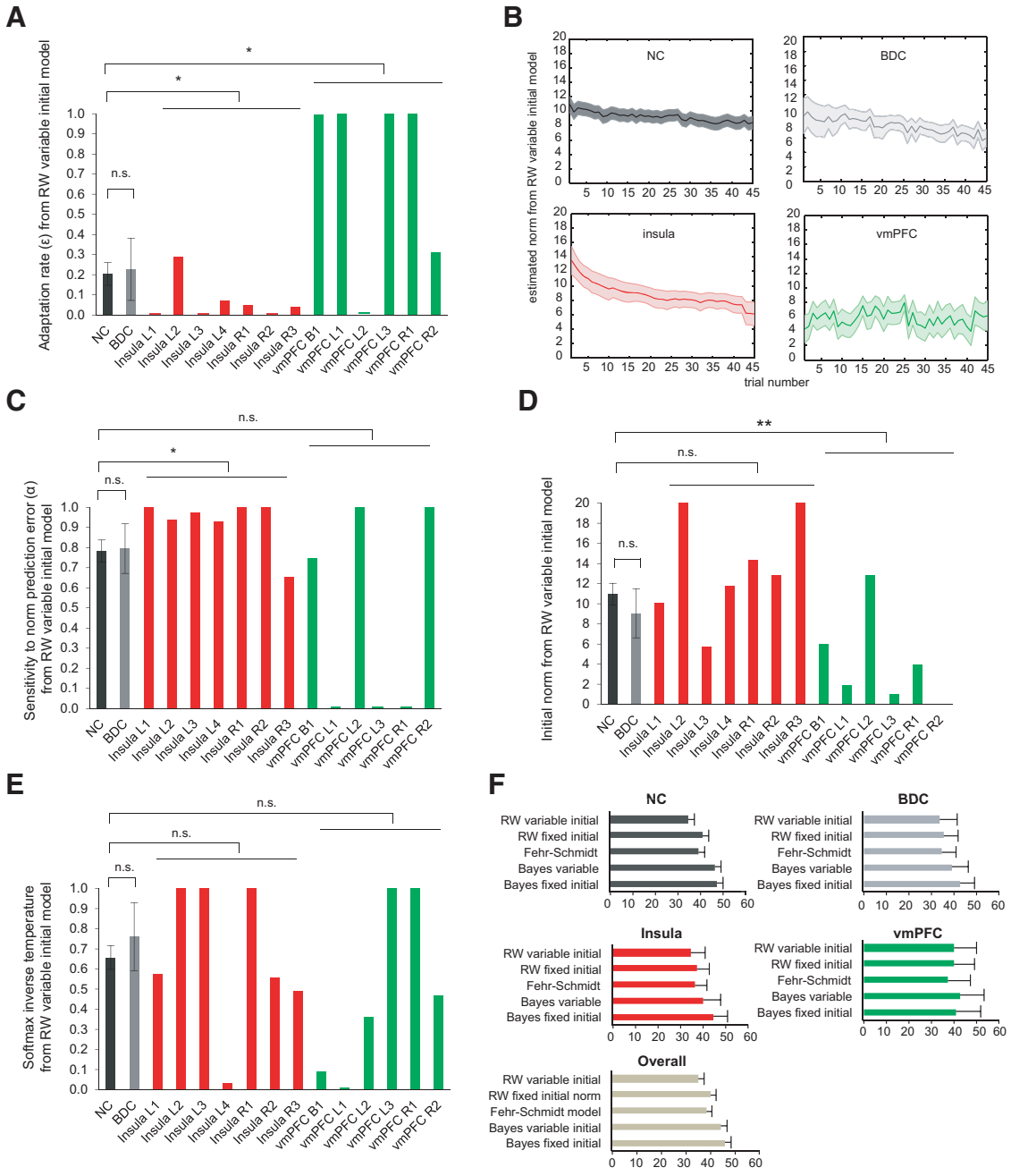

Figure 3. ***A***, Patients with insula lesions showed slower adaptation rate ε of the RW model with variable initial norm, while patients with vmPFC lesions had higher ε. ***B***, Estimated trajectory of internal norms. ***C***, Patients with insula lesions, but not vmPFC or BDC patients, showed increased sensitivity to norm prediction error α of the RW model. ***D***, Patients with vmPFC lesions, but not insula or BDC patients, showed decreased initial norm of the RW model. ***E***, No significant group difference was detected for the softmax inverse temperature. ***F***, Normalized Bayesian information criterion scores of RW, FS, and Bayes models: the model with the lowest BIC score has the maximal model evidence; *$p < 0.05$; **$p < 0.01$; n.s, not significant. L, Left; R, right; B, bilateral. Data are represented as mean ± SEM.

which was the superior model among all models based on lowest overall BIC (Fig. 3*F*). This model was also the winning model for NC, BDC, and insula groups individually, but not for vmPFC patients, possibly due to their valuation deficits in the first place. Therefore, the parameters of the vmPFC group should be interpreted with caution.

Figure 3*A* shows the results on the norm adaptation rate ε parameter: patients with insula lesions showed a lower adaptation rate ε than neurologically intact controls, suggesting they adjusted internal norms more slowly (Fig. 3*A*; $p < 0.05$); in contrast, patients with vmPFC lesions showed a higher adaptation rate ε and adjusted internal norms more quickly than NC subjects (Fig. 3*A*; $p < 0.05$). BDC patients did not show significant alternation in adaptation rate ε (Fig. 3*A*; $p > 0.1$), confirming that the results observed in insula and vmPFC patients were not due to general loss of brain tissues or surgical influences. Figure 3*B* illustrates the estimated adaptation trajectories: patients with insula lesions showed flattened updating curve in their internal norm, corresponding to their low adaptation rate. By contrast, patients with vmPFC lesions showed faster fluctuations in their internal norms, consistent with their high adaptation rate.

We also found higher α in patients with insula lesions than NC subjects (Fig. 3*C*; $p < 0.05$), indicating enhanced sensitivity to norm prediction errors in these patients. Patients with vmPFC lesions showed a trend of lower α than controls (Fig. 3*C*; $p = 0.1$). The BDC group did not show any significant abnormality in α (Fig. 3*C*; $p > 0.1$). Additionally, patients with vmPFC lesions (Fig. 3*D*; $p < 0.01$), but not insula or BDC patients ($p > 0.1$), showed lower initial norm. There was no significant difference in the softmax inverse temperature parameter γ between any patient group and NC subjects (Fig. 3*E*; $p > 0.1$).

Together, these model-based analyses suggest abnormal adaptation rate and sensitivity to norm prediction errors associated with insula lesions, highlighting a critical role of the insular cortex in norm learning.

Discussion

Our main findings are twofold. First, patients with insula lesions displayed abnormally low adaptation speed to norms in a changing environment; conversely, the behavior of these patients was nevertheless more sensitive to norm violations. Second, patients with vmPFC lesions showed diminished sensitivity to fairness and were more willing to accept unfair offers. These findings provide compelling human lesion evidence suggesting that the insular cortex is necessary for learning to adapt when reality deviates from norm expectations, and that the vmPFC is critical for valuation of fairness during social exchange.

Our results provide the first causal evidence supporting a critical role of the insula in norm learning. Several fMRI studies show insular activation signals fairness and reciprocity during social interaction (Sanfey et al., 2003; Xiang et al., 2013). Insula lesions have been linked to difficulty in adjusting choices with the odds of winning in risky situations (Clark et al., 2008) and learning about negative values (Palminteri et al., 2012) in gambling tasks. Consistent with these previous findings, our findings further suggest that patients with insula lesions also failed to adapt to norms with preserved sensitivity to fairness, highlighting a necessary role of insula in learning in social environments. Future studies are needed to assess whether insula lesion-related social and nonsocial learning deficits directly correlate with each other.

Physiological signals and feelings arising from the body are an essential part of cognition and decision-making (Damasio, 1996; Gu et al., 2013a; Gu and FitzGerald, 2014). This notion can be supported by the highly consistent findings in recent studies demonstrating the involvement of the insula in a wide range of cognitive and decision-making tasks (Preuschoff et al., 2008; Gu et al., 2012, 2013b, 2014; Palminteri et al., 2012; Xiang et al., 2013; Kirk et al., 2014), although it has been traditionally considered as an interoceptive cortex (Craig, 2013). Thus, it is not surprising that insular lesions could lead to deficits in the processing of physiological signals from the body and the subjective experience of these bodily signals ("feelings") in a decision-making task, and contribute to learning deficits. We speculate that insula patients in our study could not accurately represent bodily feelings associated with deviations from norm expectations and therefore, chose to stay with their original representation of the norms and not to adapt. In contrast, these insula patients all had intact vmPFC and were able to compute value signals and respond to offer size normally.

One limitation of our study is that we did not directly measure subjective feelings or related physiological responses. Future studies are needed to better understand whether the learning deficits observed in insula patients arise purely from a lack of physiological/feeling signals that are normally provided by the insula, or rather, also higher-level abnormalities in associating bodily responses with value and prediction error signals. One study suggests that insula lesions disrupted craving in chronic smokers, but spared pleasurable bodily feelings associated with eating and taste (Naqvi et al., 2007); that is, insula is critical for bodily feelings acquired through learning (e.g., smoking), but might not be crucial for bodily feelings that are inheritably pleasurable (e.g., eating). These results are consistent with the findings on learning deficits observed in our patients with insula lesions, and are in favor of the possibility that insula is critical for linking bodily responses with decision signals, rather than representing interoceptive signals *per se*. Together, the difficulty in learning to adapt when reality deviates from norm expectations, as observed in patients with insula lesions, highlights a critical role of interoceptive representation in the insular cortex during social learning.

Our results also support a critical role of the vmPFC in valuation, consistent with findings from fMRI studies on the vmPFC (Behrens et al., 2008; Rushworth et al., 2012) and lesion evidence supporting diminished sense of "guilt" in patients with vmPFC lesions (Krajbich et al., 2009). One previous study showed increased rejection rates in patients with vmPFC lesions (Koenigs and Tranel, 2007). We speculate that this might be due to the following reasons. First, Koenigs and Tranel's (2007) vmPFC patients had acquired sociopathy, which was not the case for our patients. Thus, the lower rejection rate could also be attributed to sociopathy, rather than vmPFC lesions *per se*. Second, offers were presented in a fixed order in the Koenigs and Tranel (2007) study ($5, $4, $3, $2, $1) but randomized in our design. Presenting high offers first has been shown to increase subjects' norm expectations and decrease acceptance rate at subsequent low offers (Xiang et al., 2013). Therefore, it is possible that presenting high offers first had a greater impact on vmPFC patients' norm expectations, and that such learning dynamics were not captured by the analysis in the Koenigs and Tranel (2007) study. Thirdly, the proposer's face was displayed in the Koenigs and Tranel (2007) study. Considering the well known role of facial features in social interaction (Winston et al., 2002), viewing a partner's face could have impacted vmPFC patients' choices in that case.

Other lesion studies on vmPFC and decision-making have largely used nonsocial paradigms. Using a simple pairwise choice paradigm, Fellows and Farah reported greater inconsistency in preference in patients with vmPFC lesions (Fellows and Farah, 2007). Other studies have found impaired reversal learning associated with vmPFC lesions (Fellows and Farah, 2003; Hornak et al., 2004). Our findings complement and extend these previous studies on nonsocial scenarios, and point to a critical role played by the vmPFC in valuation in social situations by suggesting that decreased sensitivity to fairness and subsequently volatile representations of the norm can be linked to vmPFC impairment.

In summary, we provide compelling human lesion evidence supporting necessary, yet dissociable roles of the insula and vmPFC in norm adaptation. Our computational approach also invites future investigations using combined lesion and modeling approaches.

Notes

Supplemental material for this article is available at https://sites.google.com/site/xgufmri/download. Supplemental Figure 1: Simulations of internal norms and rejection behavior captured by low/high adaptation rate ($\varepsilon = 0.08$ or 0.8) and low/high sensitivity to norm prediction error ($\alpha = 0.15$ or 0.9). Supplemental Figure 2: Ideal Bayesian Observer Model: internal norm trajectory. Supplemental Table 1: Parameter estimates of all five models

tested. Supplemental analysis: Random effects model based on Rescorla–Wagner learning with variable initial norm. This material has not been peer reviewed.

References

Behrens TE, Hunt LT, Woolrich MW, Rushworth MF (2008) Associative learning of social value. Nature 456:245–249. CrossRef Medline

Chang LJ, Sanfey AG (2013) Great expectations: neural computations underlying the use of social norms in decision-making. Soc Cogn Affect Neurosci 8:277–284. CrossRef Medline

Clark L, Bechara A, Damasio H, Aitken MR, Sahakian BJ, Robbins TW (2008) Differential effects of insular and ventromedial prefrontal cortex lesions on risky decision-making. Brain 131:1311–1322. CrossRef Medline

Craig AD (2013) An interoceptive neuroanatomical perspective on feelings, energy, and effort. Behav Brain Sci 36:685–686. CrossRef Medline

Critchley HD, Harrison NA (2013) Visceral influences on brain and behavior. Neuron 77:624–638. CrossRef Medline

Damasio AR (1996) The somatic marker hypothesis and the possible functions of the prefrontal cortex. Philos Trans R Soc Lond B Biol Sci 351: 1413–1420. CrossRef Medline

Fehr E, Schmidt KM (1999) A theory of fairness, competition, and cooperation. Q J Economics 114:817–868. CrossRef

Fellows LK, Farah MJ (2003) Ventromedial frontal cortex mediates affective shifting in humans: evidence from a reversal learning paradigm. Brain 126:1830–1837. CrossRef Medline

Fellows LK, Farah MJ (2007) The role of ventromedial prefrontal cortex in decision making: judgment under uncertainty or judgment per se? Cereb Cortex 17:2669–2674. CrossRef Medline

Folstein MF, Folstein SE, McHugh PR (1975) "Mini-mental state." A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 12:189–198. CrossRef Medline

Gu X, FitzGerald TH (2014) Interoceptive inference: homeostasis and decision-making. Trends Cogn Sci 18:269–270. CrossRef Medline

Gu X, Gao Z, Wang X, Liu X, Knight RT, Hof PR, Fan J (2012) Anterior insular cortex is necessary for empathetic pain perception. Brain 135: 2726–2735. CrossRef Medline

Gu X, Hof PR, Friston KJ, Fan J (2013a) Anterior insular cortex and emotional awareness. J Comp Neurol 521:3371–3388. CrossRef Medline

Gu X, Liu X, Van Dam NT, Hof PR, Fan J (2013b) Cognition-emotion integration in the anterior insular cortex. Cereb Cortex 23:20–27. CrossRef Medline

Gu X, Kirk U, Lohrenz TM, Montague PR (2014) Cognitive strategies regulate fictive, but not reward prediction error signals in a sequential investment task. Hum Brain Mapp 35:3738–3749. CrossRef Medline

Haughton DMA (1988) On the choice of a model to fit data from an exponential family. Ann Statist 16: 342–355. CrossRef

Hornak J, O'Doherty J, Bramham J, Rolls ET, Morris RG, Bullock PR, Polkey CE (2004) Reward-related reversal learning after surgical excisions in orbito-frontal or dorsolateral prefrontal cortex in humans. J Cogn Neurosci 16:463–478. CrossRef Medline

Kirk U, Gu X, Harvey AH, Fonagy P, Montague PR (2014) Mindfulness training modulates value signals in ventromedial prefrontal cortex through input from insular cortex. Neuroimage 100:254–262. CrossRef Medline

Knight RG (1984) Some general population norms for the short form beck depression inventory. J Clin Psychol 40:751–753. CrossRef Medline

Koenigs M, Tranel D (2007) Irrational economic decision-making after ventromedial prefrontal damage: evidence from the ultimatum game. J Neurosci 27:951–956. CrossRef Medline

Krajbich I, Adolphs R, Tranel D, Denburg NL, Camerer CF (2009) Economic games quantify diminished sense of guilt in patients with damage to the prefrontal cortex. J Neurosci 29:2188–2192. CrossRef Medline

Montague PR, Lohrenz T (2007) To detect and correct: norm violations and their enforcement. Neuron 56:14–18. CrossRef Medline

Naqvi NH, Rudrauf D, Damasio H, Bechara A (2007) Damage to the insula disrupts addiction to cigarette smoking. Science 315:531–534. CrossRef Medline

Palminteri S, Justo D, Jauffret C, Pavlicek B, Dauta A, Delmaire C, Czernecki V, Karachi C, Capelle L, Durr A, Pessiglione M (2012) Critical roles for anterior insula and dorsal striatum in punishment-based avoidance learning. Neuron 76:998–1009. CrossRef Medline

Preuschoff K, Quartz SR, Bossaerts P (2008) Human insula activation reflects risk prediction errors as well as risk. J Neurosci 28:2745–2752. CrossRef Medline

Rescorla RA, Wagner AR (1972) A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: Classical conditioning II: current research and theory (Black AH, Prokasy WE eds), pp 64–99. New York: Appleton-Century-Crofts.

Rilling JK, King-Casas B, Sanfey AG (2008) The neurobiology of social decision-making. Curr Opin Neurobiol 18:159–165. CrossRef Medline

Ruff CC, Ugazio G, Fehr E (2013) Changing social norm compliance with noninvasive brain stimulation. Science 342:482–484. CrossRef Medline

Rushworth MF, Kolling N, Sallet J, Mars RB (2012) Valuation and decision-making in frontal cortex: one or many serial or parallel systems? Curr Opin Neurobiol 22:946–955. CrossRef Medline

Sanfey AG, Rilling JK, Aronson JA, Nystrom LE, Cohen JD (2003) The neural basis of economic decision-making in the ultimatum game. Science 300:1755–1758. CrossRef Medline

Seymour B, O'Doherty JP, Dayan P, Koltzenburg M, Jones AK, Dolan RJ, Friston KJ, Frackowiak RS (2004) Temporal difference models describe higher-order learning in humans. Nature 429:664–667. CrossRef Medline

Sutton RS, Barto AG (1998) Reinforcement learning: an introduction. Cambridge, MA: MIT Press.

Winston JS, Strange BA, O'Doherty J, Dolan RJ (2002) Automatic and intentional brain responses during evaluation of trustworthiness of faces. Nat Neurosci 5:277–283. CrossRef Medline

Xiang T, Lohrenz T, Montague PR (2013) Computational substrates of norms and their violations during social exchange. J Neurosci 33:1099–1108a. CrossRef Medline