Human extinction learning is accelerated by an angiotensin antagonist via ventromedial prefrontal cortex and its connections with basolateral amygdala

Feng Zhou[a], Yayuan Geng[a], Fei Xin[a], Jialin Li[a], Pan Feng[b,c], Congcong Liu[a], Weihua Zhao[a], Tingyong Feng[b,c], Adam J. Guastella[d,e], Keith M. Kendrick[a], Benjamin Becker[a]*

[a]Clinical Hospital of Chengdu Brain Science Institute, MOE Key Laboratory for Neuroinformation, University of Electronic Science and Technology of China, Chengdu, China

[b]Faculty of Psychology, Southwest University, Chongqing, China.

[c]Key Laboratory of Cognition and Personality, Ministry of Education, Southwest University, Chongqing, China.

[d]Autism Clinic for Translational Research, Brain and Mind Centre, Central Clinical School, Faculty of Medicine, University of Sydney, Camperdown, Australia.

[e]Youth Mental Health Unit, Brain and Mind Centre, Central Clinical School, Faculty of Medicine, University of Sydney, Camperdown, Australia.

***Correspondence**
Benjamin Becker
Center for Information in Medicine
University of Electronic Science and Technology of China
Chengdu 611731, China
Tel.: +86 2861 830 811
Mail: ben_becker@gmx.de

Abstract

Recent translational research suggests a role of the renin-angiotensin (RA) system in threat extinction and underlying neuroplasticity; however, whether and how pharmacological modulation of the RA system influences physiological and neural manifestations of threat during extinction learning in humans is unclear. Here we report that pre-extinction administration of losartan, an angiotensin II type 1 receptor antagonist, accelerated attenuation of physiological threat expression. During early extinction, losartan enhanced threat-signal specific ventromedial prefrontal cortex (vmPFC) activation and its coupling with the basolateral amygdala. Multivoxel pattern analysis revealed that losartan reduced whole brain, particularly vmPFC, threat expression and voxel-wise mediation analyses further confirmed that losartan-accelerated extinction crucially involved vmPFC processing. Overall the results provide initial evidence for a critical role of the RA system in extinction learning in humans and suggest that adjunct losartan administration may facilitate the efficacy of extinction-based therapies.

ClinicalTrials.gov, Identifier: NCT03396523

Introduction

Extinction learning refers to the attenuation of a previously learned defensive response when the threat predictive stimulus is repeatedly encountered in the absence of adverse consequences. Exposure-based interventions (partly) rely on extinction learning to reduce excessive fear responses in patients and are considered an efficient therapy for anxiety disorders[1]. However, extinction learning deficits are considered a core pathological mechanism across anxiety disorders and considerably impede the efficacy of exposure-based interventions[1]. The neural mechanisms of extinction are evolutionary extremely well-conserved and the translational identification of receptor targets to facilitate neural plasticity in pathways underlying extinction learning may thus help to augment the efficacy of exposure-based interventions[2].

Animal models and human neuroimaging research have demonstrated a crucial role of the infra-limbic cortex (IL) (homologous to the human ventromedial prefrontal cortex, vmPFC) and its interactions with the amygdala in extinction learning[3-9]. The vmPFC is critically engaged in the reduction of threat expression during extinction[1,6,10] and governs the amygdala inhibition of conditioned threat response[5,7]. Translational disorder models suggest that dysfunctions in amygdala-prefrontal plasticity (at least partly) contribute to the extinction-failure in anxiety disorders[11]. Converging evidence come from clinical research suggesting that anxiety disorders are characterized by deficient extinction, hypoactivation within the vmPFC and attenuated vmPFC-amygdala functional connectivity[4,5,9,10,12].

Initial preclinical research has identified the glutamatergic N-methyl-D-aspartate (NMDA) as a highly promising candidate system[13] to facilitate extinction. For instance, the NMDA receptor partial agonist D-cycloserine (DCS) has been repeatedly demonstrated to improve extinction learning in rodents[13], probably via augmenting prefrontal engagement[14]. However, translation into clinical utility has been limited by rapid attenuation of DCS-enhanced extinction with repeated administration[15], and clinical trials reporting small - or even detrimental - effects when DCS was administered in combination with exposure therapy[16].

Recent evidence suggests that the renin-angiotensin (RA) system, primarily known for its role as a blood pressure and renal absorption regulator, may represent a promising alternative target to facilitate extinction[2]. In addition to peripheral expression, angiotensin receptors are densely expressed and modulate neuroplasticity in limbic (e.g., amygdala) and prefrontal brain regions critically engaged in extinction[17-19]. Initial studies in rodents have successfully demonstrated the potential of pharmacological modulation of RA signaling to enhance and accelerate extinction using the selective competitive angiotensin II type 1 antagonist losartan (LT)[20,21]. Notably - in contrast to DCS - LT-induced enhancement was fully maintained with repeated administration in rodents[20]. Losartan is routinely prescribed to treat high blood pressure[22], and increasingly examined for potential neuroprotective effects after stroke and traumatic brain injury[23] and initial clinical observations suggest beneficial effects on memory formation and anxiety-symptomatology[2,22]. Losartan crosses the blood-brain barrier[23,24], and following oral administration peak plasma levels are

reached after 90 minutes[25,26]. Whereas effects on cardiovascular indices only become apparent after 3 hours, effects at central receptors following intravenous administration have been observed after 30 minutes[24]. In line with the pharmacodynamic profile, previous studies reported effects on neurocognitive performance[27] and fear-related neural activity[19] following 90 minutes after oral administration in healthy human participants. Taken together, accumulating evidence suggests that LT-induced modulation of the central RA system may represent a promising target to enhance extinction learning. Moreover, the excellent safety record of LT would facilitate rapid translation into clinical application.

Against this background the present pre-registered randomized placebo-controlled pharmacological between-subject functional magnetic resonance imaging (fMRI) experiment (ClinicalTrials.gov, Identifier: NCT03396523) aimed at determining the potential of LT (50mg, single dose p.o. administration) to facilitate extinction learning in humans and the underlying neural mechanisms. Based on previous translational research we expected that LT would (1) accelerate attenuation of the physiological threat responses, and that enhanced extinction would be neurally mediated by (2) increased activation in the vmPFC and attenuation of its threat expression in the context of (3) stronger functional interaction of the vmPFC with the amygdala.

To test our hypotheses 70 healthy males underwent a validated Pavlovian threat acquisition and extinction procedure with simultaneous fMRI and skin conductance response (SCR) acquisition. To reduce variance related to menstrual-cycle-associated

variation in extinction[28,29] only male participants were included in the present proof-of-concept experiment (for similar approach see ref. [30]). During the threat acquisition stage, participants were repeatedly presented with two different colored squares serving as CSs (conditioned stimulus). One of the CSs (CS^+) was pseudo-randomly paired with a mild electric shock (US), whereas the other CS (CS^-) was never paired with a US. Following the acquisition participants were administered either a single oral dose of LT or placebo (PLC) and underwent the extinction procedure during which the same CSs were presented in the absence of the US after table peak plasma concentrations of LT had been reached (90-min post treatment, detailed experimental protocols shown in **Supplementary Figure 1**).

Results

Participants and Potential Confounders

After initial quality assessment of the data the primary hypotheses were evaluated in n = 30 LT- and n = 29 PLC-treated subjects (exclusion according to recommendations from ref. [31], details provided in **Methods**). Although previous studies indicate no effects of LT-single dose administration on cardiovascular activity or affective state during the time window of the present extinction experiment[19,25,27] corresponding indices were monitored during the experimental period (0-120min after treatment). In line with the previous studies, no effects on cardiovascular and affective indices were observed, which together with the chance level guesses for treatment (LT vs PLC) argues against unspecific confounding effects of treatment (**Table 1**). During the pre-

treatment acquisition phase, both groups exhibited CS discrimination with enhanced SCR to the CS^+ relative to the CS^- (**Figure 1a**), confirming successful threat acquisition. On the neural level acquisition of threat was accompanied by stronger CS^+ versus CS^- responses in the threat acquisition networks (see e.g. refs. [31,32]) (**Supplementary Figure 2**). Importantly, two-sample t-tests did not reveal between-group activation differences during this stage arguing against confounding effects of pre-treatment differences in threat acquisition.

Losartan Treatment Reduces Physiological Threat Responses during Early Extinction Learning

A mixed-design analysis of variance (mixed ANOVA) model using physiological threat response (SCR CS^+ - CS^-) during extinction learning as dependent variable demonstrated a significant main effect of run ($F_{(1, 57)} = 17.063$, $P < 0.001$, partial $\eta^2 = 0.230$, η^2 indicates effect size in terms of eta squared) reflecting decreased physiological threat responses in run 2 compared to run 1 and a marginally significant main effect of phase ($F_{(1, 57)} = 3.387$, $P = 0.071$, partial $\eta^2 = 0.056$) reflecting decreased physiological threat responses during late extinction and demonstrating successful extinction learning. Importantly, a significant treatment × phase interaction effect ($F_{(1, 57)} = 5.017$, $P = 0.029$, partial $\eta^2 = 0.081$) was observed, with post hoc tests indicating that, relative to the PLC group, the LT group exhibited decreased physiological threat responses during early extinction learning across both runs ($t_{(57)} = -2.179$, $P = 0.034$, d = -0.567, d indicates effect size in terms of Cohen's d) (see insert

figure **in Figure 1c**), suggesting accelerated extinction learning. Exploratory run-specific analyses confirmed that LT-treatment enhanced early extinction learning on the level of the physiological threat responses during both, initial as well as repeated extinction learning (run 1 $t_{(57)}$ = -1.805, P = 0.038, d = -0.470; run 2 $t_{(57)}$ = -2.012, P = 0.024, d = -0.525; two-sample t-tests comparing the treatment groups, one-tailed, **Figure 1c**). No further significant main or interaction effects were observed on the differential threat responses ($CS^+ > CS^-$; all Ps > 0.4). An examination of the safety signal (CS^-) revealed no between-group differences (all Ps > 0.15, **Figure 1d**) confirming specific effects of LT on extinction learning and additionally arguing against unspecific effects of treatment on the SCR response.

Accelerated Extinction Learning Accompanied by Increased vmPFC Engagement

The neural basis of LT-facilitated extinction was examined employing concordant mixed ANOVA models to the concomitantly acquired fMRI data. Whole-brain voxel-wise analysis revealed a significant treatment × phase interaction effect on extinction-related neural activity ($CS^+ > CS^-$) in the vmPFC (peak MNI_{xyz} = [-3, 27, -12], $F_{(1,57)}$ = 23.582, $P_{clusterFWE}$ = 0.011, k = 187) (**Figure 2a**). To further disentangle the interaction effect, two different post hoc approaches were employed to establish high robustness and regional specificity of the findings. First, a leave-one-subject-out (LOSO) cross-validation (CV) procedure[33] was used to determine the robustness of the effects while ensuring that individual extracted regional signals were independent

from the voxel-wise group level findings (for details see **Methods**). Results from the LOSO CV procedure demonstrated that compared to the PLC group, the LT group exhibited increased vmPFC activation during early ($t_{(57)}$ = 3.417, P = 0.001, d = 0.890), but not late ($t_{(57)}$ = -1.556, P = 0.125, d = -0.405), extinction learning (**Figure 3b**, **Figure 3a** displays an overlay of all leave-one-subject-out-ROIs, note that the vmPFC showed a robust treatment × phase interaction effect). Second, the regional specificity of effects was further confirmed by means of voxel-wise whole-brain post-hoc comparisons between the treatment groups demonstrating regional-specific increased vmPFC activity in the LT group during early extinction (peak MNI_{xyz} = [6, 48, -3], $t_{(57)}$ = 4.505, $P_{clusterFWE}$ = 0.013, two-tailed, k = 139), whereas no group differences were observed during late extinction (**Figure 2b**).

Moreover, extraction of the stimulus-specific neural signal from an anatomically-defined vmPFC ROI (details see **Methods**) revealed significant main effects of stimulus type and run, as well as a significant treatment × stimulus type interaction effect during early extinction (**Figure 6a,** details provided in **Supplementary Results**). Exploratory post-hoc two-sample t-tests demonstrated that LT enhanced vmPFC responses to the CS^+ during both the initial (run 1 $t_{(57)}$ = 2.141, P = 0.037, d = 0.557) and repeated extinction learning (run 2 $t_{(57)}$ = 2.126, P = 0.038, d = 0.554), but not to the CS^- (Ps > 0.29), further confirming LT-induced stimulus-specific effects on the threat signal (CS^+). In addition, in line with previous studies[34,35] exploratory post-hoc paired sample t-tests showed conditioned threat responses with a decrease in BOLD signal for the CS^+ relative to the CS^- in early extinction in PLC-treated

subjects in both run 1 ($t_{(28)}$ = -3.917, P < 0.001, d = -0.727) and run 2 ($t_{(28)}$ = -4.829, P < 0.001, d = -0.897), reflecting persistence of the acquired conditioning memory. In contrast LT-treated subjects exhibited comparable responses to the CS^+ and CS^- in both runs (Ps > 0.6), suggesting that following LT-treatment differential vmPFC responsivity was already normalized during early extinction (**Figure 4a**).

To further evaluate our hypothesis that LT would accelerate extinction learning the temporal pattern of LT-treatment effects was decoded employing an exploratory single trial analysis. In line with our hypothesis, analysis of single trial extinction-related vmPFC threat reactivity ($CS^+ > CS^-$) demonstrated that LT specifically increased activity in this region during the initial re-exposure to the previously conditioned threat stimuli (q < 0.05, FDR corrected) (**Figure 4b**).

Losartan Treament Reduces Neural Threat Pattern Expression to the Conditioned Stimulus during Early Extinction Learning

To test whether LT-treatment reduces threat expression during early extinction, a whole-brain multivariate threat-predictive pattern was decoded[36] (**Supplementary Figure 3a**). To this end a linear support vector machine (SVM) was employed to discriminate the threat (non-reinforced CS^+) and non-threat (CS^-) stimuli during acquisition using a LOSO CV procedure. Results showed a classification accuracy of 87.93% (P < 0.001) and sensitivity and specificity to threat were 0.862 (CI: 0.766-0.944) and 0.897 (CI: 0.811-0.967) respectively suggesting that the neural threat-predictive pattern could effectively distinguish neural threat versus non-threat

processing (before treatment, **Supplementary Figure 3b**). The neural threat-predictive pattern encompassed a distributed threat representation network including the vmPFC, insula, dorsal anterior cingulate cortex, thalamus and hippocampus (q < 0.05, FDR corrected, **Supplementary Figure 3a**) resembling findings from mass-univariate analyses of threat acquisition in the present as well as previous studies (overview see ref. [32]). Next the neural threat-predictive pattern was applied to the early extinction activation maps using a LOSO procedure (CS^+ > baseline and CS^- > baseline; dot-product of a vectorized activation image with the classifier weights, for details see **Methods**). In the entire sample the neural threat-predictive pattern (CS^+ - CS^-) was positively associated with physiological threat reactivity (differential SCR; $r_{(57)} = 0.571$, $P < 0.001$), confirming the functional relevance of the neural threat expression[36]. Comparing the groups further revealed that LT significantly decreased the magnitude of the differential threat expression (CS^+ - CS^-) relative to PLC ($t_{(57)} = -2.091$, $P = 0.041$, d = -0.544, two-sample t-test), suggesting attenuated neural expression of threat during early extinction.

Based on our a priori regional hypothesis and the key role of the vmPFC in the reduction of threat expression during extinction[1,6,10] an additional analysis specifically focused on the effects of LT on the vmPFC partial threat expression. Briefly, the unthresholded neural threat-predictive pattern was masked by the anatomical vmPFC mask and next applied to early extinction activation maps with the higher partial pattern expression chosen as the threat stimulus (i.e., a forced-choice test). Results revealed that only the PLC group (accuracy = 89.66%, $P < 0.001$; sensitive = 0.862,

CI: 0.774-1; specificity = 0.862, CI: 0.774-1, discrimination above chance) but not the LT group (accuracy = 56.67%, $P = 0.585$; sensitive = 0.567, CI: 0.382-0.742; specificity = 0.567, CI: 0.387-0.742) exhibited neural threat experience during early extinction. In line with the whole brain results, LT significantly attenuated early-extinction partial threat expression (CS^+ - CS^-) in the vmPFC ($t_{(57)} = -3.410$, $P = 0.001$, d = -0.888), indicating that LT-treatment successfully reduced vmPFC threat expression (**Supplementary Figure 3c**).

vmPFC Activity Drives Losartan Induced Extinction Enhancement

A key question is how LT-treatment leads to accelerated extinction learning as reflected by reduced physiological threat responses in early extinction. Based on previous studies indicating crucial contributions of the vmPFC to extinction and the facilitation of extinction[5,10,37-39] we hypothesized that treatment-induced effects on vmPFC activation would critically mediate accelerated extinction. To test our hypothesis we employed a voxel-wise bootstrapping-based mediation analysis within the anatomically defined vmPFC (details see **Methods**). Effects of LT-treatment on vmPFC activity (path a), vmPFC activity on physiological threat responses (path b) and the mediation effect (a × b) were determined in a single model. LT-treatment significantly contributed to both the reduced physiological threat responses (path c; $t_{(57)} = -2.179$, $P = 0.034$) and enhanced vmPFC activation (path a; peak MNI_{xyz} = [6, 36, -21], Z = 4.616, $q < 0.05$, FDR corrected within pre-defined vmPFC mask, k = 544, **Figure 5A in yellow**). Furthermore, a significant negative correlation between

vmPFC activity and the physiological threat responses (path b; controlled for treatment, peak MNI_{xyz} = [-6, 48, -15], Z = -4.882, $q < 0.05$, FDR corrected within pre-defined vmPFC mask, k = 221, **Figure 5A in green**) as well as a significant negative mediation effect in the vmPFC (peak MNI_{xyz} = [-3, 45, -18], Z = -3.795, $q <$ 0.05, FDR corrected within pre-defined vmPFC mask, k = 139, **Figure 5A in blue**) were observed. Importantly, a vmPFC cluster exhibited conjunction effects of both LT-treatment (path a) and fear expression (path b) as well as a mediation effect (peak MNI_{xyz} = [-3, 45, -15], Z = -3.692, $q < 0.05$, FDR corrected within pre-defined vmPFC mask, k = 138), suggesting that activation in this region significantly mediated LT-induced attenuation of physiological threat responses. An independent vmPFC-focused (ROI) mediation analyses confirmed that activation in this region significantly mediated the impact of treatment on physiological threat responses (**Figure 5B**; a × b effect, bootstrapped *P* value = 0.016; each path is shown in **Figure 5C**). Together these results indicate an indirect pathway, with the vmPFC mediating the effects of LT on extinction acceleration.

Losartan Treatment Enhanced vmPFC-Amygdala Coupling

Based on the critical contribution of vmPFC-amygdala circuitry to extinction learning[7-9] we hypothesized that LT-induced extinction enhancement would be accompanied by stronger functional interaction between these regions. To test our hypothesis extinction-related ($CS^+ > CS^-$) vmPFC-amygdala functional coupling was examined using a context-dependent physiological interaction analyses (gPPI)[40]. This

analysis revealed LT-induced enhanced functional coupling between the vmPFC and the right amygdala (peak MNI_{xyz} = [24, -3, -24], $t_{(57)} = 3.557$, $q < 0.05$ (one-tailed), FDR corrected within pre-defined amygdala mask, k = 13, **Figure 6a**) during early extinction. Cytoarchitectonic probabilistic mapping[41] further revealed that LT specifically enhanced vmPFC coupling with the basolateral amygdala (BLA) subregion. Subsequent examination of stimulus-specific connectivity estimates from the vmPFC-bilateral BLA pathway confirmed specific effects of LT on threat signal (CS^+) processing ($t_{(57)} = 2.147$, $P = 0.036$, d = 0.559) (**Figure 6b**).

Discussion

Using a translational pharmacological imaging approach, the present study demonstrated for the first time that transient pharmacological modulation of the RA system via LT enhanced extinction learning as reflected by accelerated attenuation of SCR to the previously conditioned threat stimulus. During early extinction, the treatment effect was mediated by enhanced threat-signal specific vmPFC activity and stronger functional coupling of the vmPFC with the BLA. These effects were further paralleled by a threat-predictive pattern classification approach showing that LT-treatment accelerated attenuation of the expression of threat pattern which was established during threat acquisition. Overall, the present findings provide first evidence for an important contribution of the RA system to fear extinction in humans and the potential of LT to accelerate extinction via effects on the vmPFC and its inhibitory connections with the BLA.

In humans, successful extinction is accompanied by decreased physiological threat reactivity and concomitantly increased vmPFC activation in response to the threat signal[34,35,42]. In the present study LT-treatment reduced the physiological threat responses and selectively enhanced vmPFC activation in response to the threat signal (CS^+) during early extinction, indicating its potential to accelerate extinction learning in humans. Moreover, the acceleration effects were found not only during the initial extinction learning (i.e., run 1), but also in the following "new" learning process (for extinction run 2, participants were told that the two extinction runs were independent, and thus they could not predict the absence of the US during the extinction run 2). The findings resemble previously observed LT-enhanced extinction learning in rodents[20,21] and further confirm the important contribution of the vmPFC to successful extinction. Exploring the stimulus-specific effects of LT-treatment on vmPFC activation revealed reactivity to the safety signal (CS^-) remained unaffected and decoding the temporal pattern of LT-effects further suggested that LT specifically attenuated vmPFC reactivity during early re-exposure towards the previously conditioned threat signal. Non-pharmacological stimulation of the vmPFC homologous infra-limbic cortex accelerates extinction learning in rodents[37-39] while inactivation or lesion of this region critically impede threat reduction during extinction (for reviews see refs. [5,10]). In accordance with the present findings, previous studies demonstrated that non-pharmacological stimulation of the vmPFC can enhance early extinction learning in humans[43], although unspecific effects on CS^- reactivity have also been reported[44].

Converging evidence from different lines of research suggests that the vmPFC, or the homologous IL in rodents, critically contributes to the reduction of threat expression during extinction learning[1,5,6,10] and regulates amygdala output to inhibit the conditioned threat response[5,7]. In line with the proposed contribution of the vmPFC to extinction learning, LT-attenuated physiological threat responses during early extinction were accompanied by an attenuated neural threat expression, particularly in the anatomically defined vmPFC. In line with previous studies reporting critical contributions of the vmPFC to extinction enhancement[37-39], and direct associations between activity in this region and physiological threat reactivity during extinction[45], an additional mediation analysis revealed that higher vmPFC activation was associated with stronger suppression of the physiological threat response. Notably, LT-facilitated suppression of the physiological threat response crucially involved enhanced vmPFC activation (for convergent mediation effects of vmPFC threat expression see **Supplementary Methods and Results**) further emphasizing the key role of this region in extinction enhancement.

On the network level LT-accelerated threat reduction during early extinction was parallel by stronger functional communication between the vmPFC and the amygdala, specifically the basolateral subregion. Previous lesion studies in humans demonstrated a critical role of the BLA in threat processing[46] and of the vmPFC in inhibiting amygdala threat responses by exerting top-down control over this region[47]. Animal models further confirmed the importance of pathway-specific neuroplastic changes in the vmPFC-amygdala circuitry during extinction memory formation[7-9] and suggest

that vmPFC inputs to the amygdala instruct threat memory formation and/or gate the expression of conditioned threat[7] during early extinction[8]. The present findings of CS^+-specific increased vmPFC-BLA functional connectivity following LT-treatment may therefore reflect an important modulatory role of angiotensin signaling on vmPFC regulation of the amygdala. In the context of previous animal models demonstrating that stimulation of vmPFC inputs to the amygdala promotes the formation of extinction memories[7], enhanced transmission in this pathway may reflect a core mechanism underlying angiotensin regulation of extinction learning. Angiotensin receptors are densely expressed and considered to modulate learning-related neuroplasticity in limbic and prefrontal regions critically engaged in extinction[17-19]. LT is a selective competitive antagonist of the angiotensin II type 1 receptor but also increases availability of angiotensin II-converted angiotensin IV-an agonist at the AT4 receptor subtype. The AT4 system is thought to play a role in neuroplasticity and learning and memory[17,18,27,48], a mechanism which may (partly) contribute to LT-induced extinction enhancement.

In line with previous animal models demonstrating the potential of LT to enhance extinction in rodents[20,21], the present study successfully demonstrated the potential of a single, low-dose administration of LT to facilitate extinction learning in humans thereby providing initial evidence that LT may potentially augment the efficacy of exposure-based interventions. On the neural level the effects of LT were mediated by circuits consistently involved in anxiety disorders with exaggerated threat reactivity and deficient extinction being associated with decreased vmPFC activation and

dysfunction in the vmPFC-BLA circuit[4,5,9,10]. Importantly, dysregulations in this circuitry normalize during the course of successful treatment[49] suggesting that they represent treatment-responsive - rather than stable - markers and consequently promising targets for innovative therapeutic interventions. Given a previous human neuroimaging study reporting that LT can improve threat discrimination in high anxious individuals[19], together with LT's excellent safety record in clinical applications[22], the observed extinction enhancing potential and selective effects on vmPFC-BLA threat signaling make it an attractive candidate for augmenting the effects of exposure therapy.

However, despite these initial promising results subsequent studies need to (1) examine effects of LT on subsequent extinction consolidation and recall in humans (as previously demonstrated in rodents[20]), (2) determine the generalization of the effects to female subjects and, finally, (3) evaluate its potential to enhance exposure-based interventions in clinical trials.

Overall, the present results indicate an important regulatory role of the RA system in fear extinction learning in humans that are mediated by modulatory effects on vmPFC threat processing and its interaction with the amygdala. From a clinical perspective adjunct LT-treatment may represent an innovative strategy to enhance the efficacy of exposure-based interventions.

Acknowledgements

We would like to thank Marianne Cumella Reddan for assistance with the threat expression analysis. This work was supported by grants from National Natural Science Foundation of China (NSFC) [91632117; 31530032;]; Fundamental Research Funds for the Central Universities [ZYGX2015Z002]; Science, Innovation and Technology Department of the Sichuan Province [2018JY0001].

Author contributions

F.Z. and B.B. designed the study, analyzed the data and wrote the manuscript. F.Z., Y.G., F.X., J.L., P.F., C.L, and W.Z. conducted the experiment. P.F., T.F., A.G. and K.K. revised the manuscript draft.

Competing interests

The authors declare no competing interests.

References

1 Quirk, G. J. & Mueller, D. Neural mechanisms of extinction learning and retrieval. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* **33**, 56 (2008).

2 Morrison, F. G. & Ressler, K. J. From the neurobiology of extinction to improved clinical treatments. *Depression and anxiety* **31**, 279-290 (2014).

3 Milad, M. R. & Quirk, G. J. Neurons in medial prefrontal cortex signal memory for fear extinction. *Nature* **420**, 70 (2002).

4 Milad, M. R. & Quirk, G. J. Fear extinction as a model for translational neuroscience: ten years of progress. *Annual review of psychology* **63**, 129-151 (2012).

5 Giustino, T. F. & Maren, S. The role of the medial prefrontal cortex in the conditioning and extinction of fear. *Frontiers in behavioral neuroscience* **9**, 298 (2015).

6 Schiller, D. & Delgado, M. R. Overlapping neural systems mediating extinction, reversal and regulation of fear. *Trends in cognitive sciences* **14**, 268-276 (2010).

7 Bukalo, O. *et al.* Prefrontal inputs to the amygdala instruct fear extinction memory formation. *Science advances* **1**, e1500251 (2015).

8 Delgado, M. R. *et al.* (Nature Publishing Group, 2016).

9 Dejean, C. *et al.* Neuronal circuits for fear expression and recovery: recent advances and potential therapeutic strategies. *Biological psychiatry* **78**, 298-306 (2015).

10 Sotres-Bayon, F., Cain, C. K. & LeDoux, J. E. Brain mechanisms of fear extinction: historical perspectives on the contribution of prefrontal cortex. *Biological psychiatry* **60**, 329-336 (2006).

11 Fragale, J. E. *et al.* Dysfunction in amygdala–prefrontal plasticity and extinction-resistant avoidance: A model for anxiety disorder vulnerability. *Experimental neurology* **275**, 59-68 (2016).

12 Sylvester, C. *et al.* Functional network dysfunction in anxiety and anxiety disorders. *Trends in neurosciences* **35**, 527-535 (2012).

13 Walker, D. L., Ressler, K. J., Lu, K.-T. & Davis, M. Facilitation of conditioned fear extinction by systemic administration or intra-amygdala infusions of D-cycloserine as assessed with fear-potentiated startle in rats. *Journal of Neuroscience* **22**, 2343-2351 (2002).

14 Baker, K. D., McNally, G. P. & Richardson, R. d-Cycloserine facilitates fear extinction in adolescent rats and differentially affects medial and lateral prefrontal cortex activation. *Progress in Neuro-Psychopharmacology and Biological Psychiatry* (2018).

15 Norberg, M. M., Krystal, J. H. & Tolin, D. F. A meta-analysis of D-cycloserine and the facilitation of fear extinction and exposure therapy. *Biological psychiatry* **63**, 1118-1126 (2008).

16 Reinecke, A. & Harmer, C. J. A cognitive-neuropsychological account of treatment action in anxiety: can we augment clinical efficacy? *Psycopathology Review* **3**, pr. 035113 (2016).

17 Chai, S. Y. *et al.* Distribution of angiotensin IV binding sites (AT4 receptor) in the human forebrain, midbrain and pons as visualised by in vitro receptor autoradiography. *Journal of chemical neuroanatomy* **20**, 339-348 (2000).

18 Wright, J. W. & Harding, J. W. The brain renin–angiotensin system: a diversity of functions and implications for CNS diseases. *Pflügers Archiv-European Journal of Physiology* **465**, 133-151 (2013).

19 Reinecke, A. *et al.* Angiotensin regulation of amygdala response to threat in high-trait anxious individuals. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* (2018).

20 Marvar, P. J. *et al.* Angiotensin type 1 receptor inhibition enhances the extinction of fear memory. *Biological psychiatry* **75**, 864-872 (2014).

21 do Nascimento Lazaroni, T. L., Bastos, C. P., Moraes, M. F. D., Santos, R. S. & Pereira, G. S. Angiotensin-(1-7)/Mas axis modulates fear memory and extinction in mice. *Neurobiology of learning and memory* **127**, 27-33 (2016).

22 Khoury, N. M. *et al.* The renin-angiotensin pathway in posttraumatic stress disorder: angiotensin-converting enzyme inhibitors and angiotensin receptor blockers are associated with fewer traumatic stress symptoms. *The Journal of clinical psychiatry* **73**, 849-855 (2012).

23 Thöne-Reineke, C., Steckelings, U. M. & Unger, T. Angiotensin receptor blockers and cerebral protection in stroke. *Journal of Hypertension* **24**, S115-S121 (2006).

24 Culman, J., von Heyer, C., Piepenburg, B., Rascher, W. & Unger, T. Effects of systemic treatment with irbesartan and losartan on central responses to angiotensin II in conscious, normotensive rats. *European journal of pharmacology* **367**, 255-265 (1999).

25 Ohtawa, M., Takayama, F., Saitoh, K., Yoshinaga, T. & Nakashima, M. Pharmacokinetics and biochemical efficacy after single and multiple oral administration of losartan, an orally active nonpeptide angiotensin II receptor antagonist, in humans. *British journal of clinical pharmacology* **35**, 290-297 (1993).

26 Lo, M. W. *et al.* Pharmacokinetics of losartan, an angiotensin II receptor antagonist, and its active metabolite EXP3174 in humans. *Clinical Pharmacology & Therapeutics* **58**, 641-649 (1995).

27 Mechaeil, R., Gard, P., Jackson, A. & Rusted, J. Cognitive enhancement following acute losartan in normotensive young adults. *Psychopharmacology* **217**, 51-60 (2011).

28 Garcia, N., Walker, R. & Zoellner, L. Estrogen, progesterone, and the menstrual cycle: A systematic review of fear learning, intrusive memories, and PTSD. *Clinical psychology review* (2018).

29 Hwang, M. J. *et al.* Contribution of estradiol levels and hormonal contraceptives to sex differences within the fear network during fear conditioning and extinction. *BMC psychiatry* **15**, 295 (2015).

30 Eckstein, M. *et al.* Oxytocin facilitates the extinction of conditioned fear in humans. *Biological psychiatry* **78**, 194-202 (2015).

31 Boeke, E. A., Moscarello, J. M., LeDoux, J. E., Phelps, E. A. & Hartley, C. A. Active avoidance: Neural mechanisms and attenuation of Pavlovian conditioned responding. *Journal of Neuroscience* **37**, 4808-4818 (2017).

32 Fullana, M. *et al.* Neural signatures of human fear conditioning: an updated and extended meta-analysis of fMRI studies. *Molecular Psychiatry* **21**, 500 (2016).

33 Esterman, M., Tamber-Rosenau, B. J., Chiu, Y.-C. & Yantis, S. Avoiding non-independence in fMRI data analysis: leave one subject out. *Neuroimage* **50**, 572-576 (2010).

34 Schiller, D., Kanen, J. W., LeDoux, J. E., Monfils, M.-H. & Phelps, E. A. Extinction during reconsolidation of threat memory diminishes prefrontal cortex involvement. *Proceedings of the National Academy of Sciences* **110**, 20040-20045 (2013).

35 Phelps, E. A., Delgado, M. R., Nearing, K. I. & LeDoux, J. E. Extinction learning in humans: role of the amygdala and vmPFC. *Neuron* **43**, 897-905 (2004).

36 Reddan, M. C., Wager, T. D. & Schiller, D. Attenuating Neural Threat Expression with Imagination. *Neuron* **100**, 994-1005. e1004 (2018).

37 Vidal-Gonzalez, I., Vidal-Gonzalez, B., Rauch, S. L. & Quirk, G. J. Microstimulation reveals opposing influences of prelimbic and infralimbic cortex on the expression of conditioned fear. *Learning & memory* **13**, 728-733 (2006).

38 Orsini, C. A. & Maren, S. Neural and cellular mechanisms of fear and extinction memory formation. *Neuroscience & Biobehavioral Reviews* **36**, 1773-1802 (2012).

39 Do-Monte, F. H., Manzano-Nieves, G., Quiñones-Laracuente, K., Ramos-Medina, L. & Quirk, G. J. Revisiting the role of infralimbic cortex in fear extinction with optogenetics. *Journal of Neuroscience* **35**, 3607-3615 (2015).

40 McLaren, D. G., Ries, M. L., Xu, G. & Johnson, S. C. A generalized form of context-dependent psychophysiological interactions (gPPI): a comparison to standard approaches. *Neuroimage* **61**, 1277-1286 (2012).

41 Eickhoff, S. B. *et al.* A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* **25**, 1325-1335 (2005).

42 Milad, M. R. *et al.* Recall of fear extinction in humans activates the ventromedial prefrontal cortex and hippocampus in concert. *Biological psychiatry* **62**, 446-454 (2007).

43 Guhn, A. *et al.* Medial prefrontal cortex stimulation modulates the processing of conditioned fear. *Frontiers in behavioral neuroscience* **8**, 44 (2014).

44 Dittert, N., Huettner, S., Polak, T. & Herrmann, M. J. H. Augmentation of fear extinction by transcranial direct current stimulation (tDCS). *Frontiers in Behavioral Neuroscience* **12**, 76 (2018).

45 Gottfried, J. A. & Dolan, R. J. Human orbitofrontal cortex mediates extinction learning while accessing conditioned representations of value. *Nature neuroscience* **7**, 1144 (2004).

46 Becker, B. *et al.* Fear processing and social networking in the absence of a functional amygdala. *Biological psychiatry* **72**, 70-77 (2012).

47 Motzkin, J. C., Philippi, C. L., Wolf, R. C., Baskaya, M. K. & Koenigs, M. Ventromedial prefrontal cortex is critical for the regulation of amygdala activity in humans. *Biological psychiatry* **77**, 276-284 (2015).

48 Braszko, J. J., Walesiuk, A. & Wielgat, P. Cognitive effects attributed to angiotensin II may result from its conversion to angiotensin IV. *Journal of the Renin-Angiotensin-Aldosterone System* **7**, 168-174 (2006).

49 Milad, M. R., Rosenbaum, B. L. & Simon, N. M. Neuroscience of fear extinction: implications for assessment and treatment of fear-based and anxiety related disorders. *Behaviour research and therapy* **62**, 17-23 (2014).

Methods

Participants

Seventy healthy male university students (age between 18-30 years) were enrolled in the present study. Exclusion criteria included color blindness; systolic/diastolic blood pressure > 130/90 mmHg or < 90/60 mmHg; Body Mass Index (BMI) > 30 kg m^{-2} or < 18 kg m^{-2}; current or regular substance or medication use; current or history of medical or psychiatric disorders; any endocrinological abnormalities or contraindications for LT administration and MRI. In line with previous studies targeting extinction processes[31] n = 8 participants (LT group, n = 3) who failed to acquire a conditioned threat response (average CS^+ SCR < CS^- SCR during acquisition) were excluded. Three additional participants (LT group, n = 2) were excluded due to incomplete SCR data (technical issues). During acquisition of the primary neural outcome assessment (extinction) no subjects showed excessive head motion (> 3 mm translation or 3° rotation) leading to a final sample of n = 30 LT- and n = 29 PLC-treated subjects for the evaluation of the primary hypotheses of the present study. One participant (LT) was excluded from the fMRI analyses of the acquisition phase due to excessive head motion during acquisition.

Study procedures had full approval from the local ethics committee at the University of Electronic Science and Technology of China and adhered to the latest revision of the Declaration of Helsinki. The study and procedures were pre-registered (ClinicalTrials.gov, Identifier: NCT03396523).

Experimental Design

The present study employed a randomized, placebo-controlled, double-blind, between-subject pharmaco-fMRI design. The experiment consisted of three sequential stages: (1) acquisition, (2) treatment administration and, (3) extinction. During acquisition and extinction fMRI and electrodermal activity (skin conductance responses, SCR) were simultaneously acquired. Approximately 20-min after acquisition, participants were administered either a single 50mg (p.o.) dose of the selective, competitive angiotensin II type 1 receptor (AT_1) antagonist Losartan (LT, Cozaar; Merck, USA) or placebo (PLC, sugar), packed in identical capsules. LT is routinely prescribed for the management of hypertension[50] and penetrates the blood-brain barrier[51]. In line with the pharmacodynamic profile of LT[52] and previous studies examining the cognitive enhancing properties of LT in healthy subjects[27] the experimental paradigm (extinction) started 90 minutes after administration. To enhance threat memory acquisition as well as to increase the statistical power to determine treatment effects, both learning phases included two subsequent runs of the task (for a similar approach see ref. [53]). Subjects were instructed before each run that "the experimental runs are independent and they may or may not receive the electric shock" and thus participants were not able to predcit the absence of the US at the beginning of extinction run 2.

Although previous studies reported a lack of effects of single-dose LT administration on cardiovascular activity before 3h after treatment[25], blood pressure and heart rate were assessed before drug administration, as well as before and after

the extinction paradigm to further control for potential confounding effects of LT on cardiovascular activity. To control for potential effects of LT on cardiovascular activity blood pressure and heart rate were assessed before drug administration, as well as before and after the extinction paradigm. To further control for unspecific emotional effects of LT, participants completed the Spielberger State–Trait Anxiety Inventory (STAI) and the Positive and Negative Affective Scale (PANAS) before drug administration and after the extinction paradigm[54,55] (**Supplementary Figure 1**).

Functional MRI Acqusition and Extinction Paradigms

During acquisition participants underwent an adapted version of a validated Pavlovian discrimination threat-conditioning procedure with partial reinforcement[53]. Briefly, one colored square (CS^+, 4s) coincided with a mild electric shock (US, 2ms) to the right wrist with 43% contingency, whereas the other differentially colored square (CS^-, 4s) was never paired with the US. Acquisition included two runs and each run contained eight non-reinforced presentations of the CS^+ and the CS^-, intermixed with an additional six presentations of the CS^+ paired with the shock (CS^+U). In line with Schiller, et al. [56] first trial during acquisition was a reinforced one and we included one occurrence of consecutive reinforced CS^+U trials. Stimuli were presented in a pseudorandom order with a 9-12s interstimulus interval (ISI) (fixation-cross) that served as low level baseline. Before the experiment participants were instructed to pay attention to the stimuli and find out the relationship between the stimuli and the

shocks. In post-acquisition interviews, all of the subjects reported correct relationships.

The extinction procedure consisted 2 runs with each run encompassing10 CS^+ trials without the US and 10 trails of the CS^-. The colors of stimuli, durations of trials and the range of ISI were identical to the acquisition phase. No trial-type was repeated more than two times in a row during either acquisition or extinction. The stimulus presentation and timing were optimized employing the make_random_timing.py script implemented in AFNI (Analysis of Functional NeuroImages, https://afni.nimh.nih.gov/) and were selected from 10,000 potential designs to maximize design efficiency. Two color sets were used and balanced across subjects to control for effects of the designated colors (color set A: red = CS^+, blue = CS^-; color set B: Green = CS^+, pink = CS^-).

Unconditioned Stimulus and Physiologic Threat Response

The US consisted of mild electrical shocks of 2ms duration that were individually adapted to be “highly uncomfortable, but not painful”. During both the threat conditioning and extinction procedure, the SCRs were assessed from carbon snap electrodes attached to the first and second fingers of the left hand between the first and second phalanges were sampled simultaneously with functional MRI scans. Detailed information on the setup is provided in **Supplementary Methods**.

Skin Conductance Response Analysis

The level of skin conductance response was assessed for each trial as the maximum SCR signal during a time window beginning 1s and ending 5s after stimulus onset minus the mean conductance in the 2s immediately before the onset of the stimulus and was further normalized per participant by dividing all responses to their US response. The physiological threat responses were then defined as baseline-corrected CS^+ by subtracting the mean responses to the CS^- stimuli (see **Supplementary Methods** for more details). To initially test whether the threat responses were equally acquired to the conditioned threat between LT and PLC groups, we performed two-sample t-tests on the mean physiological threat responses during acquisition. Note that we only included the non-reinforced trials, meaning that the CS^+U trials followed by shocks were excluded from SCR analyses. To initially examine whether LT had unspecific effects on the SCR signal two-sample t-tests on the mean SCR for CS^- trials during extinction were conducted. To test whether LT had an effect on extinction learning, a 2 × 2 × 2 mixed-design analysis of variance (mixed ANOVA) model with the within-subject factors "phase" (early, late) and "run" (run 1, run 2), the between-subject variable "treatment" (LT, PLC) and the mean differential SCR during extinction as dependent variable was performed. All statistical analyses were performed in SPSS 22.0 (SPSS Inc) with Greenhouse-Geisser correction for non-sphericity if indicated. Significant main or interaction effects were further disentangled using appropriate post hoc t-tests focused on differences between the treatment groups (Bonferroni corrected). Tests employed two-tailed *P*-values, with $P < 0.05$ considered significant.

MRI Data Acquisition and Analysis

All MRI data were acquired using a Siemens TRIO 3-Tesla system with a 12-channel head coil. The functional MRI data was preprocessed and analyzed using SPM12 (Statistical Parametric Mapping, https://www.fil.ion.ucl.ac.uk/spm/software/spm12/). Data acquisition and preprocessing details are presented in the **Supplementary Methods**.

A two-level random effects general linear model (GLM) analysis was conducted on the fMRI signal for statistical analyses. For the threat acquisition phase, the first-level model included three boxcar regressors, one for each stimulus type: reinforced CS^+, non-reinforced CS^+ and CS^-. For the extinction phase, separate boxcar regressors for each stimulus type (CS^+ and CS^-) were defined at 4 stages: early run 1 (first half of run 1), late run 1 (second half of run 1), early run 2 (first half of run 2) and late run 2 (second half of run 2). Each regressor was convolved with a canonical hemodynamic response function with its time derivative to reduce unexplained noise and better fit for the model. The fixation cross epoch was used as an implicit baseline, and a high-pass filter of 128 seconds was applied to remove low frequency drifts. Other regressors of non-interest (nuisance variables) included 6 head motion parameters, 6 head motion parameters one time point before, and their squares (Friston 24-parameter model[57]), and indicator vectors for spikes in head movements identified based on frame-wise displacement (FD) > 0.5mm[58]. Single-subject contrast images were obtained and then modeled at the second-level random effects analysis. For the acquisition phase the following contrasts were examined: (1) overall non-reinforced

CS^+ versus CS^- condition, to examine the threat acquisition networks across all participants, and (2) group differences on the contrasts of non-reinforced CS^+ versus CS^- condition to control for pre-treatment group differences during threat acquisition.

Effects of LT on Extinction-related Brain Activity – Whole Brain Analyses

Specific effects of LT on extinction learning were next assessed using a whole-brain 3-way mixed ANOVA model with "phase" (early, late) and "run" (run 1, run 2) as within-subject factors and "treatment" (LT, PLC) entered as a between-subject factor with the corresponding CS^+ versus CS^- contrasts. To further disentangle significant interaction effects, we employed two different post hoc approaches to warrant both, high robustness and regional specificity. (1) A leave-one-subject-out (LOSO) cross-validation (CV) procedure[33] was used to examine the robustness of the LT effect. Briefly, for every participant the same group-level analyses were performed excluding the participant's data and the resultant clusters showing the significant interaction effect were used to obtain subject-specific masks, thus the remaining subjects served as an independent localizer for the subject left out. Next, individual mean beta values were extracted from corresponding contrasts within the subject-specific mask for each participant separately and appropriate t-tests were performed. This procedure ensured that the data used to define the subject-specific masks were independent from the data used for the group comparison. (2) To further validate our results and to examine the regional specificity of the treatment effect, additional whole-brain voxel-wise post hoc t-tests were employed.

Imaging group level analyses were performed using the Randomise Tool[59] implemented in FSL (FMRIB Software Library, http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/).. To correct for multiple comparisons (including the LOSO CV procedure) permutation-based inferences (10,000 permutations) were employed to obtain clusters satisfying $P < 0.05$, family-wise error (FWE)-corrected at a cluster-forming threshold of $P < 0.001$, two-tailed.

Effects of LT on Extinction-related Brain Activity–Independent Analyses Exploring Contributions of the vmPFC

Due to the critical role of the vmPFC in extinction learning[5,6,34,36,42] in humans and based-on our *a priori* regional hypothesis we explored (1) LT effects on stimulus-specific neural signal changes in the vmPFC in early extinction, and (2) the temporal patterns of LT effects during extinction using a single-trial analysis. To this end a vmPFC region of interest (ROI) was constructed independent of the neural activity patterns in the present study by centering a 9mm-radius sphere around the centroid coordinates of an anatomical defined vmPFC mask to increase regional-specificity (MNI_{xyz} = [0, 34, -15]) (mask displayed in **Figure 6c**). The anatomical vmPFC was created by combining the subcallosal cortex, frontal medial cortex, anterior cingulate cortex and paracingulate gyrus from the Harvard-Oxford Atlas (thresholded at 25% probability) and only contained regions inferior to the genu of the corpus callosum and medial to x = 20 and x = -20[60]. Note that we also thresholded the atlas at 0%

probability and the resultant centroid coordinates were similar (MNI_{xyz} = [0, 34, -16]) and the corresponding findings remained the same.

Exploration of Stimulus-specific Effects of LT on vmPFC Activity in Early Extinction

To specifically explore LT-treatment effect on vmPFC activity during early extinction, we conducted a 2 × 2 × 2 mixed ANOVA with the within-subject factors "stimulus" (CS^+, CS^-) and "run" (run 1, run 2), the between-subject variable "treatment" (LT, PLC) and the mean beta weights in the vmPFC ROI during early phase of extinction as dependent variable. Significant interaction effects were further disentangled using appropriate post hoc t-tests (Bonferroni corrected).

Exploratory Single Trial Analysis

Based on our hypothesis that LT would accelerate extinction leaning trial-wise vmPFC activations during CS^+ presentation were explored. To this end a GLM design matrix with separate regressors for each trial was constructed with each aligned to either the onset of the CS^+ or the CS^- stimuli. Considering that single trial estimates are highly sensitive to movement artifacts and noise that occur during one trial, we excluded trials with high variance inflation factors or grand mean β estimate and one PLC participant was excluded from this analysis (for details see **supplementary methods)**. To compute the mean beta values in the vmPFC ROI for each threat stimulus, the mean beta values in the same region across all of the CS^- trials were

subtracted away from each CS^+ trial as baseline for each run separately. Next the resultant beta values for each threat stimulus (across the two runs) were subjected to trial-specific two-sample t-tests. To control for the multiple comparisons, the *P* values were further false discovery rate (FDR) corrected. Due to limited reliability of the noisy single-trial SCRs (e.g., easily affected by stimulus unrelated movement and concurrent breathing)[61], we did not perform single trial SCR analyses.

Multivariate Voxel Pattern Analysis (MVPA)

Following Reddan, et al. [36], we used a linear support vector machine (C = 1) implemented in Canlab core tools (https://github.com/canlab/CanlabCore) to train multivariate pattern classifiers for threatening (CS^+) and non-threatening (CS^-) stimuli and developed a whole brain threat-predictive neural pattern with a LOSO CV procedure. Briefly, the pattern classifiers were trained on subject-wise univariate non-reinforced CS^+ > baseline and CS^- > baseline contrasts during threat acquisition to discriminate between threatening and non-threatening stimuli. To avoid overfitting, classification was evaluated by a leave-one-subject-out cross validation. In each fold, the optimal hyperplane was computed based on the multivariate pattern of 57 participants and evaluated by the remaining participant. This procedure was repeated with each participant serving as testing data once. The cross-validated neural threat-predictive pattern consists of the weights of each voxel in predicting the threatening or non-threatening stimulus presentation plus the intercept. Bootstrap tests (10,000 replacements) were conducted to provide two-tailed *P*-values for voxel weights in

order to threshold the classifier weights for displaying the voxels that made the most reliable contributions to the classification ($P_{FDR} < 0.05$, **Supplementary Figure 3a**). Next we used the pattern expression, generated by taking the dot-product of the unthresholded classifier weights with participant-specific brain activity maps, to evaluate effects of LT-treatment on the neural threat expression during the early extinction phase. To avoid circularity, the weight map from the cross-validation fold that held out the participant to be tested was used to assess threat pattern expression in that subject. That is, the pattern expression estimates for each individual were obtained from patterns trained on other participants' data[36]. For a given data image this procedure yields a single weighted average amount where the pattern defines the weights and it is known as "pattern expression"[62]. In line with a previous study[36] the pattern expression was designed to differentiate CS^+ versus CS^- and was thus considered as an index of the neural threat response . Effects of treatment were examined by comparing the magnitude of the neural threat-predictive pattern expression (CS^+ - CS^-) between the treatment groups using a two-sample t-test.

Voxel-wise Mediation Analysis

To test whether LT-enhanced extinction could be explained by the concomitantly observed increased vmPFC activity phase a voxel-wise mediation analysis was performed using Mediation Toolbox (https://github.com/canlab/MediationToolbox)[63,64]. Briefly, the mediation analysis tests whether the observed relationship between an independent variable (X) and a

dependent variable (Y) could be explained by a third variable (M). Significant mediation is obtained when inclusion of M in a path model of the effect of X on Y significantly alters the slope of the X–Y relationship. That is, the difference between total (path c) and direct (non-mediated, path c′) effects of X on Y (i.e., c − c′), which could be performed by testing the significance of the product of the path coefficients of path a × b, is statistically significant. In the current study, we used treatment (PLC versus LT) as the "X" variable; mean physiological threat response (SCR CS^+ - CS^-) during early extinction as the "Y" variable and activity from a candidate mediating voxel in the anatomical-defined vmPFC during early extinction as the "M" variable. Thus, the X–Y relationship (path c) reflects the linear association between treatment and response to the conditioned stimuli; the X–M relationship (path a) is the estimated linear effect of treatment on vmPFC activity and the M-Y (path b) is the slope of the vmPFC activity- response to the conditioned threat relationship controlling for treatment. Significance estimates were calculated at each voxel for paths a, b and a × b through bootstrapping (10,000 replacements). To apply multiple-comparisons correction, bootstrapped two-tailed *P*-values for each vmPFC voxel were corrected for multiple testing (i.e., small volume correction) using FDR correction. To test the robustness and visualize the mediation effect, we conducted a similar analysis using mean brain activity in the vmPFC ROI as a potential mediator (see **Figure 5b**).

Network-level Effects of LT: Functional Connectivity Analysis

To further address LT effects on the interplay of extinction-related regions, we conducted a generalized form of context-dependent psychophysiological interaction analyses (gPPI)[40]. Given that both human and animal studies strongly implicate vmPFC-mediated inhibition of the amygdala as the key of the extinction of conditioned threat[6-9], the analysis focused on the vmPFC-amygdala circuits and based on the phase-specific effects of LT on early extinction-related brain activity (see Results part), between group connectivity differences in this phase were examined. Similar to previous studies (see e.g., ref. [65]), we used a group-constrained subject-specific approach to define individual volume of interest (VOI) seed regions (three-voxel-radius, i.e., 9mm-radius, sphere centered at local maximum $CS^+ > CS^-$ activity in the early extinction within the significant vmPFC cluster showing LT effects). Four PLC participants didn't show greater activation for CS^+ stimulus relative to CS^- stimulus in any voxel within the cluster and so the VOIs were centered at the group-level maximum.

A two-sample t-test was then conducted using FSL's Randomise Tool to examine LT-effects on vmPFC functional connectivity for the CS^+ versus CS^- during early extinction. Based on our a priori hypothesis the voxel-wise FDR correction was adapted to the structurally-defined bilateral amygdala (Harvard-Oxford Subcortical Structural Atlas, thresholded at 25% probability).

Data availability

Unthresholded group-level statistical maps are available on NeuroVault (https://neurovault.org/collections/4722/) and code that supports the findings of this study is available from the corresponding author upon reasonable request.

References

50 Dahlöf, B. *et al.* Cardiovascular morbidity and mortality in the Losartan Intervention For Endpoint reduction in hypertension study (LIFE): a randomised trial against atenolol. *The Lancet* **359**, 995-1003 (2002).

51 Zakrocka, I. *et al.* Angiotensin II Type 1 Receptor Blockers Inhibit KAT II Activity in the Brain—Its Possible Clinical Applications. *Neurotoxicity research* **32**, 639-648 (2017).

52 Das, A., Dhanure, S., Savalia, A., Nayak, S. & Tripathy, S. Human bioequivalence evaluation of two losartan potassium tablets under fasting conditions. *Indian journal of pharmaceutical sciences* **77**, 190 (2015).

53 Feng, P., Becker, B., Feng, T. & Zheng, Y. Alter spontaneous activity in amygdala and vmPFC during fear consolidation following 24 h sleep deprivation. *NeuroImage* **172**, 461-469 (2018).

54 Spielberger, C., Gorsuch, R., Lushene, R., Vagg, P. & Jacobs, G. Manual for the state-trait inventory. *Palo Alto, CA: Consulting Psychologists* (1970).

55 Watson, D., Clark, L. A. & Tellegen, A. Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology* **54**, 1063 (1988).

56 Schiller, D., Raio, C. M. & Phelps, E. A. Extinction training during the reconsolidation window prevents recovery of fear. *Journal of visualized experiments: JoVE* (2012).

57 Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S. & Turner, R. Movement - related effects in fMRI time - series. *Magnetic resonance in medicine* **35**, 346-355 (1996).

58 Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L. & Petersen, S. E. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage* **59**, 2142-2154 (2012).

59 Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M. & Nichols, T. E. Permutation inference for the general linear model. *Neuroimage* **92**, 381-397 (2014).

60 Koenigs, M. *et al.* Distinct regions of prefrontal cortex mediate resistance and vulnerability to depression. *Journal of Neuroscience* **28**, 12341-12348 (2008).

61 Gerlicher, A., Tüscher, O. & Kalisch, R. Dopamine-dependent prefrontal reactivations explain long-term benefit of fear extinction. *Nature communications* **9**, 4294 (2018).

62 Woo, C.-W., Chang, L. J., Lindquist, M. A. & Wager, T. D. Building better biomarkers: brain models in translational neuroimaging. *Nature neuroscience* **20**, 365 (2017).

63 Wager, T. D., Davidson, M. L., Hughes, B. L., Lindquist, M. A. & Ochsner, K. N. Prefrontal-subcortical pathways mediating successful emotion regulation. *Neuron* **59**, 1037-1050 (2008).

64 Wager, T. D. *et al.* Brain mediators of cardiovascular responses to social threat, part II: Prefrontal-subcortical pathways and relationship with anxiety.

NeuroImage **47**, 836-851 (2009).

65 Nielsen, J. D. *et al.* Working memory modulation of frontoparietal network connectivity in first-episode schizophrenia. *Cerebral Cortex* **27**, 3832-3841 (2017).

Figure legends

Figure 1. Skin conductance results. (a) Physiological threat responses (CS^+ - CS^-) during acquisition. (b) Mean SCR for CS^- presentations during acquisition. (c) Physiological threat responses (CS^+ - CS^-) during extinction learning. Physiological threat responses during early extinction (across two runs) are presented in the inset. (d) Mean SCR for CS^- presentations during extinction learning. †$P < 0.05$, one-tailed; *$P < 0.05$, two-tailed, error bars represent standard errors. The filled curve indicates the null-hypothesis distribution of the difference of means (Δ) and the 95% confidence interval of Δ is illustrated by the black line.

Figure 2. Losartan treatment effects on brain activity (CS^+ - CS^-) during extinction learning. (a) vmPFC activity showed significant treatment by phase interaction effect. (b) Losartan specifically increased vmPFC activity during early extinction learning. All images were thresholded at $P < 0.05$, cluster-level family-wise error (FWE)-corrected with a cluster-forming threshold of $P < 0.001$, two-tailed. Examples of unthresholded patterns are presented in the insets; small squares indicate voxel statistical weight; red-outlined squares indicate significance at $P_{\text{clusterFWE}} < 0.05$. vmPFC, ventromedial prefrontal cortex.

Figure 3. Leave-one-subject-out (LOSO) cross-validation (CV) procedure revealed that losartan treatment specifically increased vmPFC activity (CS^+ - CS^-) during the early extinction phase. (a) Overlay of all 59 LOSO CV ROIs. All ROIs were created leaving out one subject at the group-level statistic ($p < 0.05$, cluster-level family-wise error (FWE)-corrected with a cluster-forming threshold of P

< 0.001, two-tailed). (b) Mean vmPFC activity (CS^+ - CS^-) extracted from the ROIs depicted in (a) showed that losartan increased vmPFC activity during early, but not late, extinction learning. n.s. represents not significant; *$P < 0.05$. The filled curve indicates the null-hypothesis distribution of the difference of means (Δ) and the 95% confidence interval of Δ is illustrated by the black line. vmPFC, ventromedial prefrontal cortex.

Figure 4. Losartan treatment specifically increased vmPFC activity to threat stimulus (CS^+) during early extinction learning. (a) Losartan increased vmPFC activity to CS^+, but not CS^-, in the early extinction phase in both runs. n.s. represents not significant; *$P < 0.05$; and ***$P < 0.001$, error bars represent standard errors. (b) Single trial analysis confirmed that losartan increased vmPFC activity to CS^+ in early trials during extinction learning. *$q < 0.05$, FDR corrected. Data are represented as group mean ± SEM. (c) vmPFC ROI. vmPFC, ventromedial prefrontal cortex.

Figure 5. vmPFC activity mediated losartan treatment effect on accelerated extinction learning. (a) Sagittal slice showing regions whose activity increased response to the losartan treatment in yellow (path a), regions whose activity significant negative correlated with physiological threat responses while controlling for the treatment effect in green (path b), and regions whose activity showed significant mediation (a × b) effect in blue. All images were thresholded at $q < 0.05$, FDR corrected within the vmPFC mask. (b) Mediation path diagram with the brain activity in the vmPFC ROI. (c) Examples of each path in the mediation path diagram. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$. The filled curve indicates the null-hypothesis

distribution of the difference of means (Δ) and the 95% confidence interval of Δ is illustrated by the black line. vmPFC, ventromedial prefrontal cortex.

Figure 6. Losartan treatment effect on vmPFC-amygdala functional coupling. (a) Sagittal slice showing that losartan increased functional connectivity between vmPFC and BLA. The image was thresholded at $q < 0.05$, FDR corrected within the amygdala mask. (b) Extracted gPPI parametric estimates in the bilateral BLA showed that losartan treatment specifically enhanced vmPFC-bilateral BLA functional pathway during fear-associated stimulus presentation. n.s. represents not significant; *$P < 0.05$. The filled curve indicates the null-hypothesis distribution of the difference of means (Δ) and the 95% confidence interval of Δ is illustrated by the black line. vmPFC, ventromedial prefrontal cortex; BLA, basolateral amygdala.

Table 1 Participant demographics and neuropsychological performance

Measure	Time	LT Group (n = 30)	PLC Group (n = 29)	$t_{(57)}$	P
Age (years)	/	20.50 (1.80)	20.86 (1.68)	-0.80	0.43
BMI	/	22.33 (2.51)	21.34 (2.52)	1.51	0.14
Systolic Pressure	**Pre drug administration**	110.03 (6.42)	107.59 (5.49)	1.57	0.12
	Pre extinction	111.17 (6.07)	108.79 (6.18)	1.49	0.14
	Post extinction	111.00 (5.69)	109.72 (5.76)	0.86	0.40
Diastolic Pressure	**Pre drug administration**	71.70 (7.86)	71.10 (5.95)	0.33	0.74
	Pre extinction	72.70 (7.51)	70.17 (4.88)	1.53	0.13
	Post extinction	72.50 (8.57)	72.34 (4.15)	0.09	0.93
Heart Rate	**Pre drug administration**	76.93 (9.27)	76.03 (10.09)	0.36	0.72
	Pre extinction	75.33 (10.06)	74.17 (10.04)	0.44	0.66
	Post extinction	77.27 (9.44)	76.76 (10.93)	0.19	0.85
STAI State Anxiety	**Pre extinction**	41.90 (8.88)	41.24 (10.51)	0.26	0.80
	Post extinction	37.23 (8.48)	38.14 (8.16)	-0.42	0.68
PANAS Negative Affect Scale	**Pre extinction**	16.17 (6.95)	16.34 (9.08)	-0.08	0.93
	Post extinction	14.43 (6.46)	14.59 (7.60)	-0.08	0.93
PANAS Positive Affect Scale	**Pre extinction**	22.73 (7.55)	23.07 (6.83)	-0.18	0.86
	Post extinction	23.70 (7.67)	22.34 (6.02)	0.75	0.45

STAI, Spielberger Trait State Anxiety Inventory; PANAS, Positive Affect Negative Affect Scale; BMI, Body Mass Index

Figure 1. Skin conductance results.

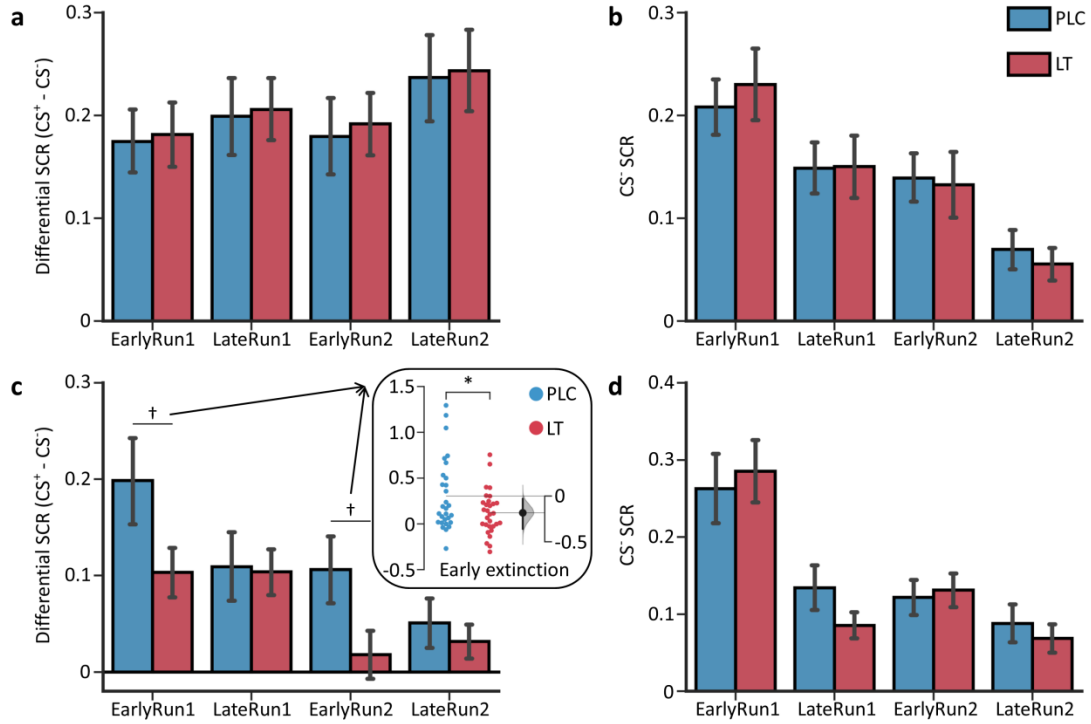


Figure 2. Losartan treatment effects on brain activity (CS^+ - CS^-) during extinction learning.

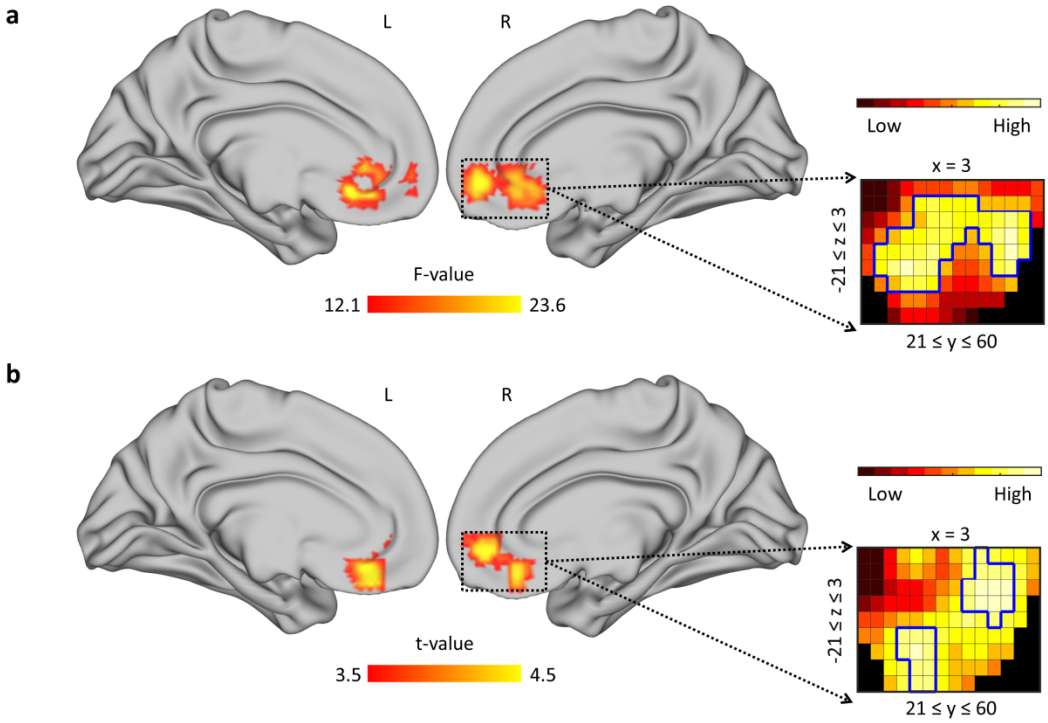


Figure 3. Leave-one-subject-out (LOSO) cross-validation (CV) procedure revealed that losartan treatment specifically increased vmPFC activity (CS^+ - CS^-) during the early extinction phase.

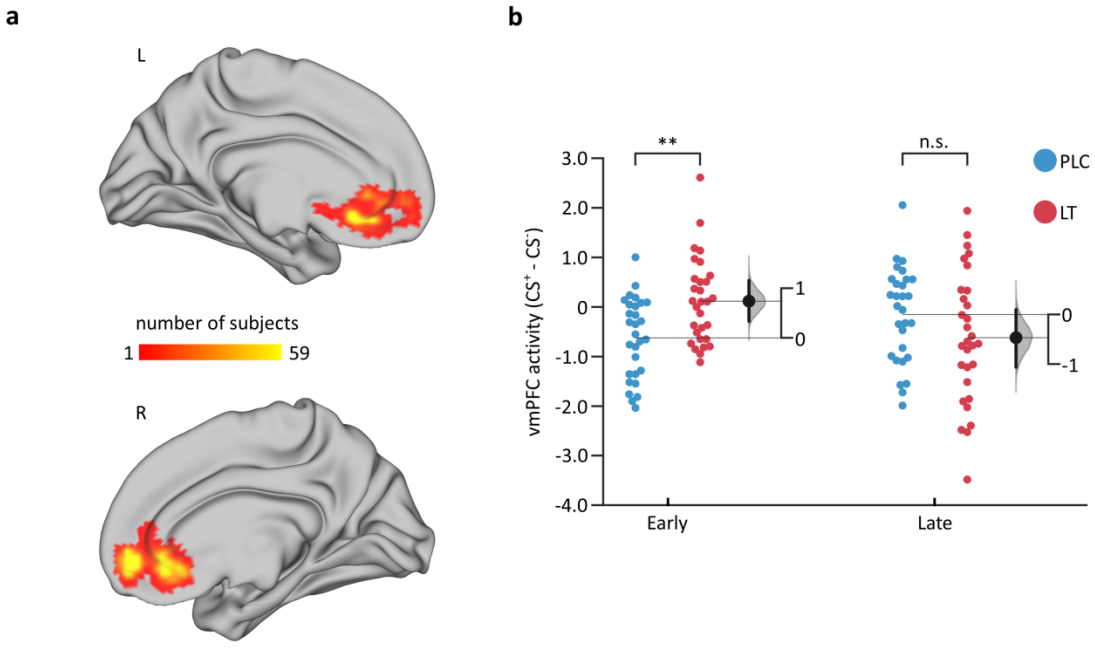


Figure 4. Losartan treatment specifically increased vmPFC activity to threat stimulus (CS^+) during early extinction learning.

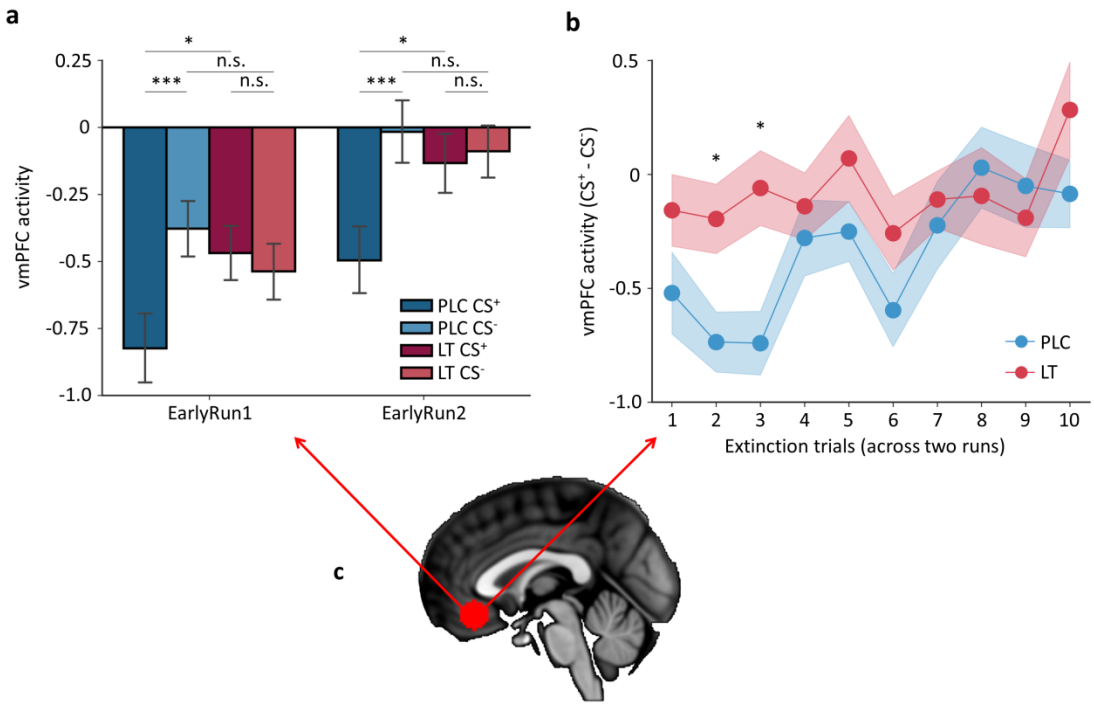


Figure 5. vmPFC activity mediated losartan treatment effect on accelerated extinction learning.

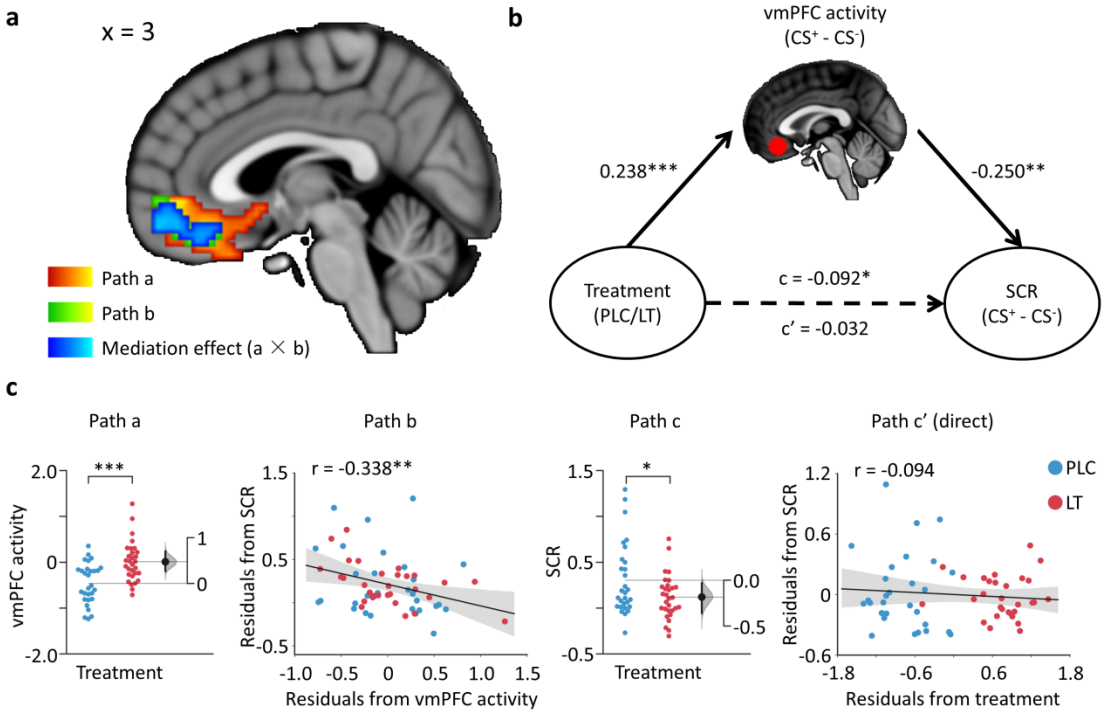
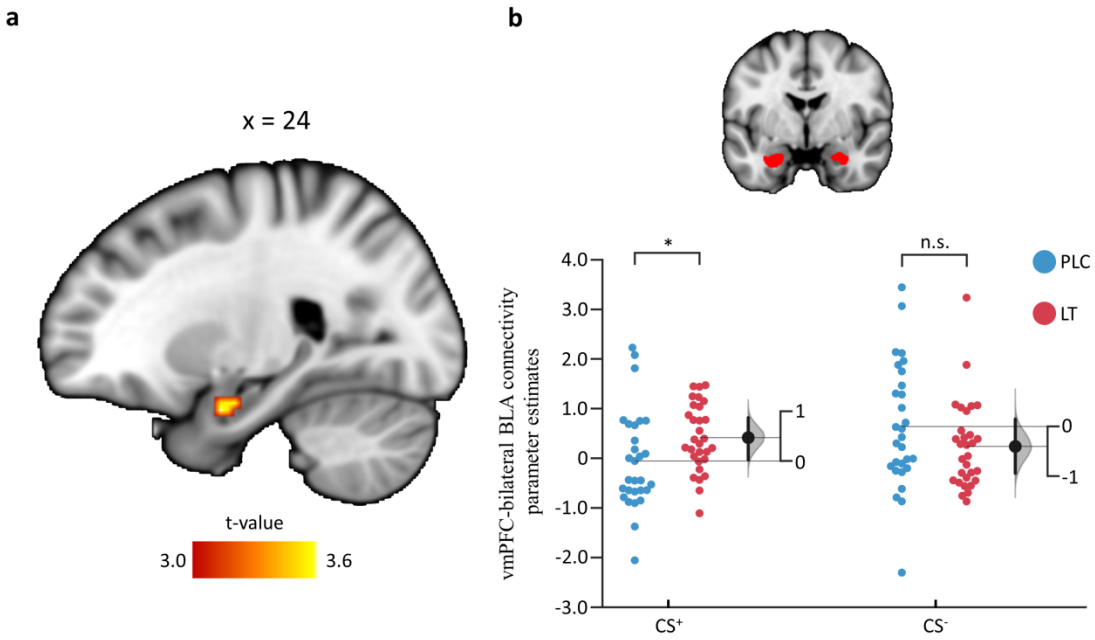


Figure 6. Losartan treatment effect on vmPFC-amygdala functional coupling.



Human extinction learning is accelerated by an angiotensin antagonist via ventromedial prefrontal cortex and its connections with basolateral amygdala

Zhou et al.,

Correspondence: ben_becker@gmx.de

Supplementary Methods

Unconditioned Stimulus and Physiologic Measurement

A mild electric shock (2ms duration) generated by a Biopac stimulator module STM100C and a STIMSOC adapter (Biopac Systems, Inc.) served as US. Shocks were delivered via two MRI-compatible electrolyte gel supported Ag/AgCl electrodes attached to the subject's right wrist. Immediately before the start of the experiment, shock intensity levels were adjusted on an individual level by delivering gradually increasing shocks until the shock reached the level that participants reported as "highly uncomfortable, but not painful". SCRs were assessed at a sampling rate of 1kHz from pre-gelled Ag/AgCl laminated carbon snap electrodes (EL508, Biopac Systems, Inc.) attached to the first and second fingers of the left hand between the first and second phalanges via Biopac Module EDA100C-MRI module attached to a MP150 (Biopac Systems Inc.).

Skin Conductance Response Analysis

The level of a skin conductance response per se was defined as the maximum of the low-pass filtered (0.1Hz) conductance signal during a time window beginning 1 second and ending 5 seconds after stimulus onset minus baseline (the mean

conductance in the 2s immediately before the onset of the stimulus). Responses below 0.02μS were encoded as zero[1-3]. To account for individual variability, the raw SCR scores were square root transformed and range corrected by dividing each response by the mean square root transformed US response during initial acquisition[3,4]. Following previous procedures[1,3,4], we divided both CS^+ and CS^- stimuli of each run into early (first half) and late (last half) phases and the mean SCR for CS^- stimulus was subtracted as baseline from CS^+. The differential SCR response (CS^+ - CS^-) was regarded as the physiological threat response to the conditioned fear stimulus.

MRI Data Acquisition and Preprocessing

MRI was acquired using a Siemens TRIO 3-Tesla system with a 12-channel head coil. Functional MRI data was acquired using a T2*-weighted echo-planar imaging (EPI) pulse sequence (33 transverse slices, repetition time = 2s, echo time = 30ms, slice thickness = 3mm, gap = 0.6mm, field of view = 200 × 200mm, resolution = 64 × 64, flip angle = 90°, voxel size = 3.1 × 3.1 × 3mm). To improve spatial normalization and exclude participants with apparent brain pathologies a high-resolution T1-weighted image was acquired using a magnetization-prepared rapid gradient echo (MPRAGE) sequence (176 sagittal slices, repetition time = 1900ms, echo time = 2.52ms, slice thickness = 1mm, field of view = 256 × 256mm, acquisition matrix = 256 × 256, flip angle = 9°, voxel size = 1 × 1 × 1mm).

Functional MRI data was preprocessed using SPM12. The first five volumes of each run were discarded to allow MRI T1 equilibration. The remaining volumes were

spatially realigned to the first volume and unwarped to correct for nonlinear distortions possibly related to head motion or magnetic field inhomogeneity, co-registered to the structural image, normalized to the Montreal Neurological Institute (MNI) space using a two-step procedure implementing segmentation of the T1-weighted image and application of the resulting deformation parameters to the functional images (interpolated to 3 × 3 × 3mm voxel size), and spatially smoothed using an 8-mm full-width at half maximum gaussian kernel. For the multivariate voxel pattern analysis unsmoothed data was used.

Exploratory Single Trial Analysis

Given that single trial analysis are susceptible to acquisition artifacts that occur during a given trial the grand mean β estimate across whole brain for each trial was computed for every subject. Trials with mean β estimate > 3 or < 3 standard deviation (SD) from the grand mean or variance inflation factor > 4 were excluded[5,6]. One subject (PLC) was excluded from the single trial analysis given the high number of excluded trials (14) with > 3 SD from the mean. Number of excluded trials for the remaining subjects was small and did not significantly differ between the treatment groups (LT, mean ± SD = 3.233 ± 1.942; PLC, mean ± SD = 2.448 ± 1.765, $P > 0.1$).

Mediation analysis

To explore whether the contribution of the differential (CS^+ - CS^-) vmPFC partial threat pattern expression (M) to the treatment (X) induced acceleration of early

extinction learning on the psychophysiological threat response (CS^+ - CS^-) SCR (Y) a mediation analysis was employed (Canlab Mediation Toolbox using10,000 bootstrap samples).

Supplementary Results

Losartan treatment enhanced threat-signal specific vmPFC activity

A more detailed examination of the treatment effects by means of extraction of the condition-specific neural signal from the independently (structurally) defined vmPFC ROI (details see Methods) additionally revealed significant main effects of stimulus type ($F_{(1, 57)} = 15.630$, $P < 0.001$, partial $\eta^2 = 0.215$, vmPFC activity was decreased to CS^+ compared to CS^-) and run ($F_{(1, 57)} = 16.848$, $P < 0.001$, partial $\eta^2 = 0.228$ with increased vmPFC activity in run 2 in relative to run 1), as well as a significant treatment × stimulus type interaction effect ($F_{(1, 57)} = 17.353$, $P < 0.001$, partial $\eta^2 = 0.233$) during early extinction. Post hoc comparisons between the treatment groups demonstrated that the interaction effect during early extinction was driven by a LT-induced selective increase of vmPFC responses to the CS^+ ($t_{(57)} = 2.777$, $P = 0.007$, d = 0.723), in the absence of significant effects on the CS^- ($t_{(57)} = -1.098$, $P = 0.277$, d = -0.286). Post hoc comparisons between the stimulus types showed that in the PLC group the BOLD signal in response to the CS^+ was decreased relative to the CS^- ($t_{(28)} = -5.685$, $P < 0.001$, d = -1.056). In contrast, LT-treated subjects exhibited a comparable vmPFC BOLD response to the CS^+ compared to the CS^- ($t_{(29)} = 0.152$, $P = 0.881$, d = 0.028).

vmPFC Partial Threat Pattern Expression Mediated Losartan Treatment Effect on Extinction Acceleration

Mediation analysis demonstrated that LT-treatment significantly reduced the vmPFC partial threat pattern expression (path a, b = -0.007, Z = -3.41, $P < 0.001$) and that lower vmPFC pattern expression led to reduced physiological threat expression with treatment as an adjustor (path b, b = 9.511, Z = 4.323, $P < 0.001$). More important, a significant mediation effect was found (a × b, b = -0.066, Z = -3.715, $P < 0.001$). Together with the total effect of LT-treatment on decreasing physiological threat responses (path c, b = -0.092, Z = -2.121, $P = 0.034$), these results suggested that LT enhanced early extinction learning through vmPFC processing.

Visualization

Statistical maps were visualized using the Connectome Workbench (https://www.humanconnectome.org/software/connectome-workbench) and Mango (http://ric.uthscsa.edu/mango/). Behavioral data were plotted using Seaborn (https://seaborn.pydata.org/) and Dabest (https://github.com/ACCLAB/DABEST-python)[7].

References

1 Taschereau-Dumouchel, V. *et al.* Towards an unconscious neural reinforcement intervention for common fears. *Proceedings of the National Academy of Sciences* **115**, 3470-3475 (2018).

2 Marstaller, L., Burianová, H. & Reutens, D. C. Adaptive contextualization: a new role for the default mode network in affective learning. *Human brain mapping* **38**, 1082-1091 (2017).

3 Schiller, D. *et al.* Preventing the return of fear in humans using reconsolidation update mechanisms. *Nature* **463**, 49 (2010).

4 Koizumi, A. *et al.* Fear reduction without fear through reinforcement of neural activity that bypasses conscious exposure. *Nature human behaviour* **1**, 0006 (2017).

5 Atlas, L. Y., Bolger, N., Lindquist, M. A. & Wager, T. D. Brain mediators of predictive cue effects on perceived pain. *Journal of Neuroscience* **30**, 12964-12977 (2010).

6 O'brien, R. M. A caution regarding rules of thumb for variance inflation factors. *Quality & quantity* **41**, 673-690 (2007).

7 Ho, J., Tumkaya, T., Aryal, S., Choi, H. & Claridge-Chang, A. Moving beyond P values: Everyday data analysis with estimation plots. *bioRxiv*, 377978 (2018).

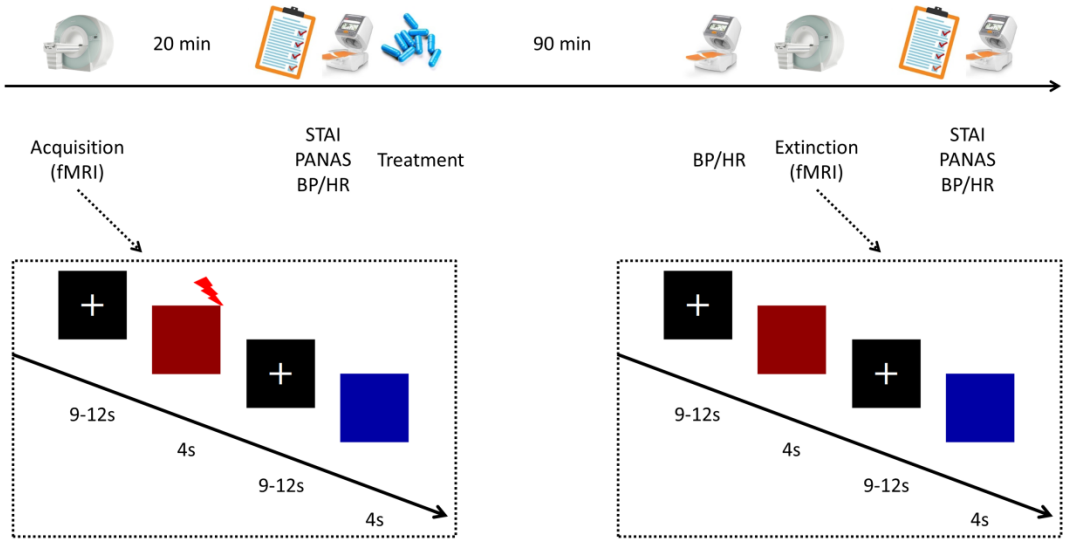
Figure legends

Supplemental Figure 1. Experimental timeline and schematic synopsis of the fear acquisition and extinction fMRI paradigm

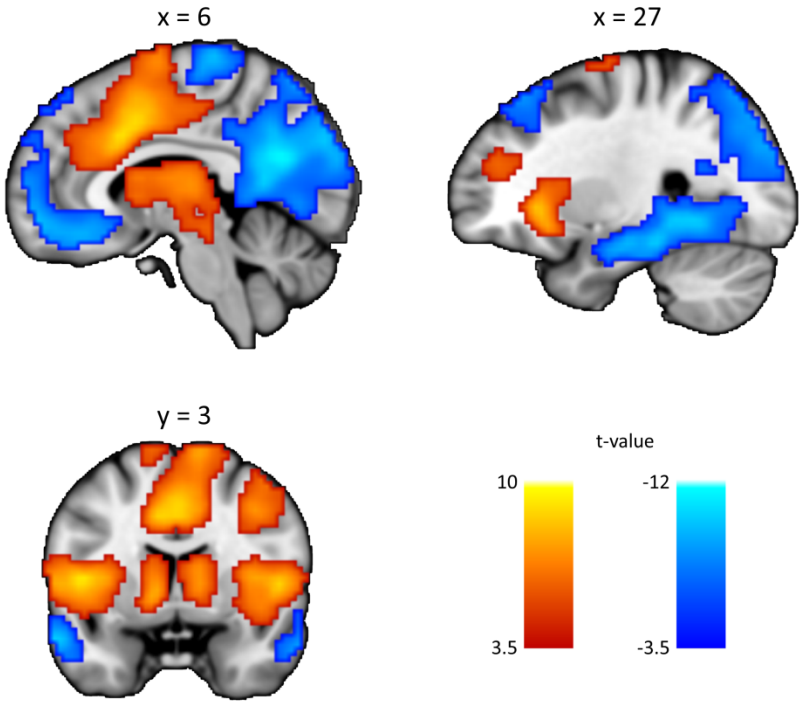
Supplemental Figure 2. Non-reinforced CS^+ versus CS^- BOLD responses during acquisition, across both experimental groups. Images displayed at $P < 0.05$, cluster-level family-wise error (FWE)-correction with a cluster-forming threshold of $P < 0.001$, two-tailed. Hot color indicates $CS^+ > CS^-$, whereas cold color indicates $CS^+ < CS^-$.

Supplemental Figure 3. Multivariate neural threat-predictive pattern results. (a) Neural threat-predictive pattern, consisting of voxels in which activity reliably predicted threatening (unreinforced CS^+) versus non-threatening ($CS^–$) stimuli during threat acquisition. The map shows weights that exceed a threshold ($q < 0.05$, FDR corrected based on bootstrapped 10,000 samples) for display only. dACC, dorsal anterior cingulate cortex; vmPFC, ventromedial prefrontal cortex. Hot color indicates positive weights and cold color indicates negative weights. (b) ROC plot. The neural threat-predictive pattern yielded a classification accuracy of 87.93% in a leave-one-subject-out cross-validation (LOSO CV) procedure. (c) Losartan treatment reduced the partial threat pattern expression of the vmPFC. $^{**}P < 0.01$. The filled curve indicates the null-hypothesis distribution of the difference of means (Δ) and the 95% confidence interval of Δ is illustrated by the black line.

Supplemental Figure 1. Experimental timeline and schematic synopsis of the fear acquisition and extinction fMRI paradigm



Supplemental Figure 2. Non-reinforced CS^+ versus CS^- BOLD responses during acquisition, across both experimental groups.



Supplemental Figure 3. Multivariate neural threat-predictive pattern results.