GABA Boosts Relief from Coercive Power: an fMRI study

Yawei Cheng
National Yang-Ming Chiao-Tung University
Róger Martínez
Taipei Medical University https://orcid.org/0000-0002-5536-5897
Yu-Chun Chen
Physical Education, National Taiwan University of Sport
Yang-Teng Fan
National Yang-Ming Chiao-Tung University
Chenyi Chen (viniverson@gmail.com)
Taipei Medical University

Article

Keywords: Lorazepam, Coercion, Milgram experiment, Anxiety, Hippocampus

Posted Date: April 12th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-151432/v1

Abstract

Individuals under coercive control frequently suffer from anxiety, with recent research asserting this situation as a catalyst for certain types of violence directed towards those suffering under the most serious and insidious forms of coercive power –such as domestic violence victims. Studies researching this matter have skewed towards dissembling manipulation, or participants' obedience levels, neglecting the fact that agents under coercive power are also victims of coercive violence. In this functional magnetic resonance imaging (fMRI) study, we investigate the effects of the anxiolytic $GABA_A$ (gamma-Aminobutyric acid) modulator, lorazepam, on behavioral and neural levels in response to coercive power. We used a virtual obedience to authority paradigm inspired in Milgram's renowned experiments of the same nature. An experimenter ordered a volunteer to press a handheld button to initiate actions that carry different moral consequences, including harming or helping others. Our results showed that lorazepam administration, relative to placebo, slowed down reaction times when initiating harming behaviors, but accelerated reaction times for helping actions, despite comparable subjective ratings regarding perceived coercion. Coercive harming significantly increased activation in the amygdala, hippocampus, orbitofrontal cortex, and dorsolateral prefrontal cortex (dlPFC). After lorazepam administration, activity in the amygdala and hippocampus decreased, but activity in the dlPFC and right temporoparietal junction increased. The lower activity in the hippocampus predicted higher subjective ratings for perceived coercion. Furthermore, lorazepam administration significantly decreased the functional connectivity of the hippocampus with the dlPFC during coercive harming. Our results shed light on the coping strategies against coercion beyond merely examining its effects.

Introduction

In 1963, Milgram published the findings of his now famous experiments on obedience to authority, with the majority of the participants seemingly obeying the instructions of an authority figure –albeit reluctantly– to harm another person by means of administering them with electric shocks (Milgram, 1963). Almost 20 years prior, the Nuremberg trials were being held, where those who orchestrated or aided in the organization of the Holocaust and other war crimes during World War II were then prosecuted (Fray et al., 1996). Many of the prosecuted claimed in their defense that they were not responsible for the crimes they were being imputed with, as they were “just obeying orders” –defense strategy later to be known as the Nuremberg defense (Arendt, 1994). Milgram’s experiments somehow echoed this defense, as the majority of the participants seemingly obeyed the instructions of an authority figure –albeit reluctantly– to harm another person by means of administering them with electric shocks(Milgram, 1963; Milgram, 1965). As a consequence of raised concerns and debates about the ethical maltreatment of participants as well as their welfare and dignity, no one has attempted a full replication of Milgram's procedure for several decades (Benjamin and Simpson, 2009; Blass, 2000, 2004, 2009; Miller, 2004; Miller et al., 1995). These studies, which have relied on a re-examination of Milgram's original data (Gilbert, 1981; Packer, 2008) or a partial-replication (Burger, 2009; Burger et al., 2011; Caspar et al., 2016), focus on authority power, dissembling manipulation and obedience levels.

While Milgram's and other studies reported widespread obedience to cause ordinary people to become callous and skewed towards dissembling obedience levels, they neglect the fact that agents under coercive pressure are also victims of coercive violence. In regards to the Milgram experiments, the high rate of participants completing the task may be excused due to the expert power displayed by the experimenter, as the participants were told that “although the shocks may be painful, they’re not dangerous”(Blass, 1999) (albeit 35% of the participants were still able to successfully disobey the orders to carry out the experiment to finalization). On the contrary, the fatal outcomes of the actions conducted by the war criminals during World War II were well known to them; and even then, there were still individual differences among those prosecuted in the trials –while some sat down during the whole event, others took their lives before the trials even began (an electronic publication of the Avalon et al., 1996).

However, not all participants were coerced into delivering harm, suggesting other factors at play. In real-life situations, individuals who obey coercive orders to harm exhibit differences in emotional responses to their immoral actions. Using again the Nuremberg trials as a historical example, some war criminals being processed took their own lives out of guilt-like anxiety even before the trials began. Moreover, while coercion-altered event-related potentials (ERPs) have been found associated with the auditory N1, induced by an implicit intentional binding paradigm (Caspar et al., 2016), how the individual variability in behavioral and neural reactivity to the anxiogenic properties of coercive power affect coping with coercion has never been addressed.

Anxiety or fear not only congeals cognitive resources and hampers cognitive functions –such as working memory–, but also changes moral judgements (Robinson et al., 2013). Clinical observations have shown that interpersonal situations are particularly strong at eliciting anxiogenic symptomatology, at least in humans (Association, 2013). As such, research using anxiolytic drugs has observed a dose-dependent increase in the participants' willingness to endorse responses that directly harm others in moral-personal dilemmas, regardless of whether the motivation for those harmful acts is selfish or utilitarian (Perkins et al., 2013). Thus, anxiety has a protective effect over participants. Nevertheless, it is interesting to ponder as to why the prospect of harming directly another person is to be experienced as a threat to oneself.

To address these questions, we investigated the behavioral and neural effects of the anxiolytic drug lorazepam on a virtual paradigm inspired on the Milgram experiments, in which an experimenter ordered a participant to inflict harm to a third party. Our study is not trying to give free way for Nuremberg-type defenses, rather than to put on display the wide spectrum of individual differences that exists at the moment of deciding (or not) to obey an order issued by an authoritative figure. The feeling of anxiety corresponding to the loss of sense of agency just accentuates that there is some sense of it left, and consequently also some responsibility. Lorazepam is a high-potency 3-hydroxy benzodiazepine prescribed for the relief of anxious symptomatology (Gould et al., 1997), as it binds to the gamma-aminobutyric acid (GABA) receptors, enhancing GABA release in the brain. Participants underwent a task to virtually simulate Milgram’s experiments, in which an experimenter ordered a subject to inflict harm to a third party through the perspective-taking of virtual actions. During fMRI scanning, participants watched

the first image of a morally-laden clip, after which they were coerced to press a handheld button in order for the successive images to play out. The last images of the clip carry different moral consequences, including harming and helping actions. The handheld button portion was designed not only to provide aid in engaging and taking the perspective of the virtual agent, but also to measure the participants' willingness to conform to the instructions of an authority figure and obey the coercive order to commit harming/helping actions. If it is true that anxiety is indeed involved in moral coercive behavior, then administering lorazepam should modulate the willingness of participants to commit (or not) certain moral acts. Specifically, our primary hypothesis was that lorazepam would significantly enhance our participants' willingness to fight against coercive power, as victims under coercive control experience anxiety. As an exploratory extension, we therefore sought to examine whether these two types of moral coercion, harming and helping others, would be differentially affected by lorazepam. Furthermore, instead of exploring the responsibility of the agents under coercion like previous studies have done already, we aim to understand the reasons that lead some people to successfully defeat the fear of authority (or counteract the effects of their coercive power) more readily than others.

Methods

Participants

To estimate the sample size needed for this placebo-controlled, crossover design study, we implemented G*power 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009). In order to have 95% power to detect a true difference based on the effect size f for the primary outcomes ranging from 0.2 to 0.22, seventy to eighty-four participants would be required with a 2-sided type I error of 0.05. Accordingly, eighty participants were enrolled, with one participant being excluded due to loss of follow-up, and other two participants being equally excluded due to excessive head motion (no. 62 & no. 37, see supplementary Fig. 1 and supplementary Table 1 for the results of sensitivity tests and the head motion descriptive statistics). The resulting seventy-seven healthy volunteers (40 males), aged between 21 and 31 (23.5 ± 2.2) years, participated in the study after providing written informed consent. Participants were screened for major psychiatric illnesses (e.g., general anxiety disorder, major depressive disorder, etc.) by the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I) (First and Gibbon, 2004), and excluded if there was evidence of any psychiatric disorder, or any comorbid neurological disorder (e.g., dementia, seizures, etc.), or any history of alcohol or substance abuse or dependence, including present or past episodes. Furthermore, participants who had any history of head injury or endocrinal disorder, including present or past episodes, were equally excluded from this study. All participants had normal or corrected-to-normal vision and were not taking any medication at the time of study. None of the female participants were taking oral contraceptives. This study was approved by the Ethics Committee of National Yang-Ming University (YM104041E) and conducted in accordance with the Declaration of Helsinki. Participants were recruited from local community colleges through printed advertisements. During the enrolment, the participants were told that this study was designed to determine whether the pill was effective in

changing emotional rating (such as guilty ratings) after playing an interactive computerized animation program. However, in each experimental session (lorazepam vs. placebo), they were blinded for drug type.

Procedures

In a double-blind, placebo-controlled, crossover design, participants received a single 0.5-mg dose of lorazepam (ATIVAN) on one day, and a single dose of placebo (i.e., vitamin C) on another day. A crossover study is a longitudinal study in which subjects receive a sequence of different treatments. Both lorazepam and placebo were administered orally. The experimental sequence of lorazepam and placebo administration was counter-balanced between participants through a Latin square design, which randomizes through having equal number of AB (lorazepam-placebo) and BA (placebo-lorazepam) sequences. Thus, half of the participants went first through the lorazepam session, and half of them went first through the placebo session. To coincide with the pharmacokinetics of lorazepam (Kyriakopoulos et al., 1978), fMRI scanning and behavior assessment took place approximately 2-hrs after treatment administration.

In this study, we designed a virtual obedience paradigm highly inspired by Milgram experiments on obedience to authority, in which an experimenter ordered a subject to inflict harm to a third party (Fig. 1A). During fMRI scanning, participants watched the first image of a morally-laden mini clip, then they were ordered (coerced) to press a handheld button in order to initiate the successive image sequence which show the actions taken to full completion and that carry different moral consequences, including harming, neutral, or helping actions. More specifically, before each coerced action, participants were told to “press the button to harm others” for harming condition, “press the button to help others” for helping condition, and “press the button to start the action” for neutral action, respectively. After MRI scanning, participants were ordered to undergo the same procedures that they did within the scanner. The visual stimuli were presented (5 trials for both harming and helping scenarios), and participants were asked to indicate how much the order to commit the action (coercion) would violate their own will. The ratings were on a 1 to 7 Likert scale, from “my will was not violated at all” to “my will was strongly violated”. Participants underwent the same experimental procedure in both placebo and lorazepam administrations with an at least two-weeks washout. The order of placebo and lorazepam administrations was counterbalanced across participants.

Visual stimuli

45 validated animations from previous fMRI studies were presented to participants in the coercive-helping and coercive-harming tasks (Chen et al., 2020a; Chen et al., 2020b). Each animation was comprised of three images, with no duration limit set for the 1st image, but a 200 milliseconds duration set for the 2nd image, and a 1000 milliseconds duration set for the 3rd image, and portraying the following scenarios: [1] a person who is alleviating physical pain from a suffering person (helping), [2] a person who is taking an action to physically harm another person (harming), and [3] a baseline stimuli depicting a person carrying out an action that is irrelevant to another person (neutral). The faces of the protagonists were not visible to ensure that no emotional reactions could be seen by the participants. The

order of scenarios was randomized across participants. Participants were explicitly primed to mentally simulate the agent, and forced to press the button to initiate the moral actions along with visual feedback of moral scenarios. More specifically, the participant would observe the first image of the animation, then would have to press the button to induce the remaining two images to play out.

fMRI acquisition, data processing and analysis

Participants underwent two sessions of fMRI scanning (placebo and lorazepam) in different days. Stimuli were presented with the E-prime software (Psychology Software Tools, Inc., Pittsburgh, PA) and an MRI compatible goggle (VisualStim Controller, Resonance Technology Inc.) in a three-level within-subject design of moral scenarios (harming vs. neutral vs. helping).

The scanning followed a block design (22.9 ± 0.6 s ON/ 13.2 ± 4.4 s OFF) and had two runs. Each run consisted of 6 ON blocks (2 harming, 2 helping, and 2 neutral scenarios) intermixed with 6 OFF blocks. Each ON block consisted of five trials, and five inter-stimulus intervals (duration 2200-ms each) with a fixation cross presented against a gray background. While the ITI was set as 2200-ms, the duration of each fMRI regressor was modeled with each participant's actual reaction times. Because the RT varies across trials and participants, the modeled duration self-served as jittering in nature, leaving the average length of each ON block 2294 ± 598 ms (mean ± SD). The sequence of the scenarios (harming, helping, neutral) was pseudo-randomized within each run. The order of runs was counterbalanced across participants.

Scanning was performed on a 3T Siemens Magnetom Trio-Tim magnet. For functional changes, changes in blood oxygenation level-dependent (BOLD) T2* weighted MR signal were collected along the AC–PC plane using a gradient echo-planar imaging (EPI) sequence (TR = 2200 ms, TE = 30 ms, FOV = 220 mm, flip angle = 90°, matrix = 64 × 64, 36 transversal slices, voxel size = 3.4 × 3.4 × 3.0 mm^3, no gap). High-resolution structural T1-weighted images were acquired using a 3D magnetization-prepared rapid gradient echo sequence (TR = 2530 ms, TE = 3.5 ms, FOV = 256 mm, flip angle = 7°, slice thickness = 1 mm, matrix = 256 × 256, no gap).

Functional images were processed with SPM12 (Wellcome Department of Imaging Neuroscience, London, UK) in MATLAB 9.0 (MathWorks Inc., Sherborn, MA, USA). Structural T1 images were coregistered to the mean functional images, and a skull-stripped image was created from the segmented gray matter, white matter, and Cerebrospinal Fluid (CSF) images. These segmented images were combined to create a subject-specific brain template. EPI images were realigned and filtered (128-s cutoff), then coregistered to these brain templates, normalized to Montreal Neurologic Institute (MNI) space, and smoothed (8 mm FWHM, full width at half maximum). The hemodynamic response function was time-locked to stimulus onset.

Data were input into a general linear model, with movement parameters as nuisance regressors. There was no significant difference in movement parameters between lorazepam and placebo. None of participants had movements greater than 2 mm of translation or 0.03 degrees of rotation (Supplementary

Table 1). A two-stage general linear model was used to examine the effect size of each condition. At the first level analysis, three conditions (harming, helping, neutral) were modeled separately with a duration of the participant's reaction time beginning at the onset of each ON block. The null event (fixation) was modeled with the duration 13.2 ± 4.4 seconds. Linear contrasts were applied to obtain parameter estimates. At the second-level analysis, images of parameter estimates from the first-level analysis (helping > neutral; harming > neutral) were collapsed into a repeated-measure factorial design with moral valence (helping vs. harming) and drug administration (lorazepam vs. placebo) as the within-subject variables. Groupwise effects for the following contrasts of whole brain activations were corrected for multiple comparisons family-wise error (FWE) rate at $P < .05$ [thresholded at uncorrected $P < .001$, cut-off, $t = 3.118$, and cluster extent of at least 10 contiguous voxels, determined by a Monte Carlo simulation conducted using 3dClustSim: https://afni.nimh.nih.gov/pub/dist/doc/program_help/3dClustSim.html]. An anatomically defined gray matter mask was created based on the MNI avg152T1 template and explicitly specified and applied to whole brain analysis.

To elucidate the lorazepam effect, regions of interest (ROIs) analyses in limbic areas were conducted for bilateral amygdala and hippocampus (Patin and Hurlemann, 2011; Schunck et al., 2010). Beyond existing literature on emotional processing, there may be additional cortical regions, which are pivotal in moral reasoning, modulated by anxiolytics. The coordinates for the right temporoparietal junction (rTPJ, 56, -50, 18) and dlPFC (42, 30, 26) were determined on the basis of neuroanatomical atlases and meta-analyses (Bzdok et al., 2012; Lamm et al., 2011). ROI data are reported for significant contrast image peaks within 10 mm of these a priori coordinates. Data extraction for the ROI analyses was performed using the MarsBaR toolbox (http://marsbar.sourceforge.net/) implemented in SPM12.

Functional connectivity analysis

Based on our whole-brain results and prior studies (Arce et al., 2006; Schunck et al., 2010), the psychophysiological interaction (PPI) analysis was seeded in the left hippocampus (-30, -12, -18) to estimate how lorazepam administration altered the functional connectivity of the hippocampus during the unwilling coercive harming condition (harming vs. neutral). The time series of the first eigenvariates of the BOLD signal were temporally filtered, mean corrected, and deconvolved to generate the time series of the neuronal signal for the source region, i.e., the left hippocampus, as the physiological variable in the PPI. The PPI analysis assesses the hypothesis that the activity in one brain region can be explained by an interaction between cognitive processes and hemodynamic activity in another brain region. As the hippocampus was selected as the PPI source region, the physiological regressor was denoted by the activity in the left hippocampus. Coercive harming (harming vs. neutral) was the psychological regressor. The interaction between the first and second regressors represented the third regressor. The psychological variable was used as a vector coding for the specific task (1 for harming, -1 for neutral) convolved with the hemodynamic response function. The individual time series of the left hippocampus was obtained by extracting the first principle component from all raw voxel time series in a sphere (4 mm radius) centered on the coordinates of the subject-specific hippocampus activations. These time series were mean-corrected and high-pass filtered to remove low-frequency signal drifts. PPI analyses were then carried out

for each subject by creating a design matrix with the interaction term, the psychological factor, and the physiological factor as regressors. PPI analyses were performed for each session separately (lorazepam and placebo) to identify brain regions showing significant changes in functional coupling with the hippocampus during coercive harming in relation to lorazepam administration. Subject-specific contrast images were then entered into random effects analyses at FWE of $P < .05$ (thresholded at $P < .001$, uncorrected, $k = 10$).

Results

Reaction times and Subjective ratings

Due to the reaction time data not being normally distributed, the RTs were base-10 LOG-transformed before further analyses (see supplementary Fig. 2, supplementary Table 2, and supplementary Table 3 for the results of sensitivity tests, normality plots, tests, and descriptive statistics of RTs during each condition). In order to contrast out the general sedation effect (slowing) of lorazepam, a difference score of RTs for helping/harming actions was derived from subtracting participant's LOG-RTs in the neutral condition from LOG-RTs in the helping or harming condition (Supplementary Table 4, Supplementary Table 5). A 2 (administration: placebo vs. lorazepam) x 2 (scenario: harming vs. helping) repeated ANOVA was then applied to test the lorazepam effect on the speed of moral behaviors under coercion. There was a main effect of scenario ($F_{1, 76} = 23.61$, $P < .001$, $\eta^2 = 0.19$) as well as an interaction between administration and scenario ($F_{1, 76} = 9.4$, $P = .003$, $\eta^2 = .11$). Follow-up analyses indicated that the reaction times in the helping condition (-0.054 ± 0.011, mean ± SE) were shorter, as compared to the harming condition (-0.018 ± 0.009, $P < .001$). The lorazepam effect had opposite directions depending on the factor of scenario. Lorazepam administration slowed down harming (lorazepam vs. placebo: -0.002 ± 0.015 vs. -0.035 ± 0.012, $t_{76} = 1.715$, $P = .09$, Cohen's d = 0.196), whereas it accelerated helping (-0.059 ± 0.014 vs. -0.049 ± 0.013) (Fig. 1B). Considering the motor inhibitory effects of Lorazepam, which is a benzodiazepine that slows motor and cognitive processes through its inhibitory mechanisms of action, we further explore whether the increasing RT is depending on the factor of moral valence, or immersive in each condition. We thoroughly re-examined the RTs data by using both parametric and non-parametric analyses on the LOG RTs (both with and without creation of RT difference scores) and raw RTs (both with and without RT difference scores), respectively (supplementary Results).

A 2 (administration: placebo vs. lorazepam) x 3 (scenario: harming vs. helping vs. neutral) repeated ANOVA using LOG RTs (before the creation of RT difference score) further confirmed the unique interaction between lorazepam effect and moral content ($P = 0.03$) in which lorazepam administration slowed down harming (lorazepam vs. placebo: 3.02 ± 0.02 vs. 2.99 ± 0.02), whereas it accelerated helping (2.96 ± 0.02 vs. 2.98 ± 0.02), and did not affect the neutral actions (3.02 ± 0.02 vs. 3.02 ± 0.02) (supplementary Results). The post-hoc tests were only conducted for the sake of revealing how the interaction functioned. There were no significant post-hoc simple main effect.

Regarding to the non-normative raw RTs data analyses, while both Friedman (χ^2 = 15.343, P = 0.002) and Kendall's W (Kendall's W = 0.066, P = 0.002) non-parametric tests showed significant differences between conditions, lorazepam moderated RTs depending on the factor of moral valence and did not increase RTs in every condition (supplementary Table 2–5)

The subjective ratings of coercion, derived from the violation of free will, were subject to a 2 (administration: placebo vs. lorazepam) x 2 (scenario: harming vs. helping) repeated ANOVA to test the lorazepam effect. There was a main effect of scenario ($F_{1, 76}$ = 49.19, P < .001, η^2 = 0.39), indicating that, under coercion, participants were less willing (i.e., more self-reported violation to their own will) to do harming (5.15 ± 0.13) than to helping (4.53 ± 0.14). Neither the administration ($F_{1, 76}$ = 0.15, P = .70) nor its interaction with scenario ($F_{1, 76}$ = 0.13, P = .72) reached significance.

fMRI results

Table 1 lists the brain regions showing a significant hemodynamic change to coercive harming and helping after placebo and lorazepam administration. In response to coercive harming (vs. neutral), both lorazepam and placebo administration showed the activation in the amygdala, hippocampus, putamen, anterior insula, temporal pole, thalamus, orbitofrontal cortex, dorsomedial prefrontal cortex, and dorsolateral prefrontal cortex (dlPFC) (Fig. 1C). Regions with greater activity during coercive harming vs. helping showing significant hemodynamic increase were the anterior insula, amygdala, hippocampus, orbitofrontal cortex, dorsomedial prefrontal cortex, and dlPFC. The reverse contrast (coercive helping vs. harming) showed increased signal in the dlPFC, rTPJ, and posterior cingulate.

Table 1

Brain regions showing significant BOLD activities to coercive harming and helping after placebo and lorazepam administration. Pooled group results for all participants ($N = 77$). All clusters are significant at voxel-wise FWE-corrected $P < .05$, except those marked with an asterisk, which are taken from a priori predefined ROIs and significant at uncorrected $P < .05$. Abbreviations: R, Right; L, left; dlPFC, dorsolateral prefrontal cortex; dmPFC, dorsomedial prefrontal cortex; vmPFC, ventromedial prefrontal cortex; rTPJ, right temporoparietal junction.

		MNI coordinates			
Brain regions	Side	x	y	z	*t*
Coercive Harming vs. Neutral					
Temporal pole	R	48	8	-24	4.03
vmPFC	-	0	50	-20	3.69
Amygdala	R	28	-2	-20	5.89
Amygdala	L	-30	-4	-14	5.06
Anterior insula	R	28	16	-18	4.29
Anterior insula	L	-28	16	-18	4.15
Hippocampus	R	32	-10	-14	6.77
Hippocampus	L	-32	-10	-14	6.34
Orbitofrontal cortex	R	40	36	-14	3.51
Orbitofrontal cortex	L	-38	32	-8	4.12
Putamen	R	26	4	-4	5.34
Putamen	L	-30	4	-4	5.95
Thalamus	L	-20	-28	0	8.76
Thalamus	R	20	-20	2	8.16
Middle occipital gyrus	R	36	-88	0	10.43
dlPFC	L	-46	18	14	5.41
dlPFC	R	54	20	28	4.18
dmPFC	-	0	62	24	6.64
Supramarginal gyrus	L	-54	-26	30	8.19
Postcentral gyrus	R	56	-22	30	7.13
Midcingulate gyrus	-	0	0	38	3.69
Precentral gyrus	L	-14	-12	74	3.31

		MNI coordinates			
Coercive Helping vs. Neutral					
Temporal pole	R	50	10	-24	3.83
Amygdala	L	-30	-4	-20	3.14
Insula	L	-36	-14	-4	3.92
Thalamus	R	22	-24	-2	8.57
Thalamus	L	-16	-28	2	7.34
Inferior frontal gyrus	L	-52	40	4	3.53
Middle occipital gyrus	R	28	-90	6	12.3
rTPJ	R	64	-44	12	3.35
Cingulate gyrus	R	12	2	28	3.73
dlPFC	R	50	22	32	4.37
dlPFC	L	-44	28	38	4.18
dmPFC	L	-4	54	42	3.97
Midcingulate gyrus	L	-2	-34	48	3.65
Postcentral gyrus	R	30	-46	66	7.2
Postcentral Gyrus	L	-4	-50	70	3.46
Coercive Harming vs. Helping					
Anterior insula	R	32	14	-20	5.59
Amygdala	L	-26	-8	-12	3.27
Hippocampus	L	-30	-14	-12	3.46
Fusiform	L	-26	-70	-10	4.1
Orbitofrontal cortex	L	-34	32	-8	4.01
Inferior temporal gyrus	L	-50	-66	-4	4.61
Anterior insula	L	-44	14	-2	4.15
dmPFC	R	6	64	28	4.99
Postcentral gyrus	L	-56	-24	28	5.58
Cuneus	L	-16	-86	36	4.25
dlPFC	L	-56	6	38	3.98

		MNI coordinates			
Coercive Helping vs. Harming					
dlPFC	R	26	18	48	4.51
Middle frontal gyrus	R	36	54	2	3.77
Angular gyrus	R	42	-64	48	4.54
Posterior cingulate	R	8	-48	38	3.34
rTPJ	R	52	-50	20	3.26
Placebo vs. Lorazepam					
Hippocampus*	L	-28	-40	4	2.25*
dlPFC*	R	48	32	30	2.44*
(Coercive Harming – Coercive Helping)|Placebo > (Coercive Harming – Coercive Helping)|Lorazepam					
Hippocampus*	L	-32	-10	-18	2.18*
Amygdala*	R	30	0	-14	1.8*
(Coercive Harming – Coercive Helping)| Lorazepam > (Coercive Harming – Coercive Helping)| Placebo					
rTPJ*	R	56	-46	20	1.88*
dmPFC*	R	42	30	28	2.29*

Regarding the ROI results (Fig. 2), significant interactions between administration (lorazepam vs. placebo) and scenario (harming vs. helping) were observed in the amygdala ($F_{1, 76}$ = 5.15, P = .026, η^2 = 0.06), hippocampus ($F_{1, 76}$ = 4.89, P = .03, η^2 = 0.06), and dlPFC ($F_{1, 76}$ = 5.87, P = .018, η^2 = 0.07). The rTPJ had a marginal effect with a medium effect size ($F_{1, 76}$ = 3.70, P = .058, η^2 = 0.05). The follow-up analyses indicated that the administration effect in the amygdala, hippocampus, dlPFC, and rTPJ had opposite directions depending on the factor of scenarios (coercive harming vs. helping). Acute administration of lorazepam relative to placebo, the activity in the amygdala (-0.004 ± 0.04 vs. 0.117 ± 0.035) and hippocampus (0.003 ± 0.028 vs. 0.082 ± 0.027) was decreased, whereas the activity in the dlPFC (0.007 ± 0.06 vs. -0.203 ± 0.063) and rTPJ (0.009 ± 0.065 vs. -0.175 ± 0.071) was increased.

To examine the association between subjective experience of coercion and neural responses, we did the whole-brain correlation analysis when subjective ratings were computed as a continuous variable with FWE rate at P < .05 (Fig. 3A). After lorazepam administration, the activity in the hippocampus to coercive harming was significantly negatively correlated with subjective ratings of coercion (r = -0.3, P = .01).

Functional connectivity

Lorazepam triggered distinct patterns in the functional coupling (Fig. 3B). After the lorazepam administration, the PPI analysis seeded in the hippocampus (-30, -12, -18) showed a significant negative coupling with the dlPFC (20, 22, 48; 40, 28, 52), orbitofrontal cortex (20, 36, -18), temporal pole (56, 8, -14), anterior cingulate cortex (4, 22, -6), postcentral gyrus (-50, -20, 58), superior temporal gyrus (-54, -26, 4), supplementary motor area (-10, -22, 50), medial frontal gyrus (20, 60, 0), and superior temporal gyrus (50, -22, 6). Whereas after placebo administration, the left hippocampus showed significantly positive connectivity with the dlPFC (52, 40, 2; -40, 44, 32), middle occipital gyrus (-44, -78, 4), middle frontal gyrus (-26, -2, 50), and supplementary motor area (-6, 4, 76). Importantly, lorazepam relative to placebo administration significantly decreased the coupling of the left hippocampus with the dlPFC (48, 48, 2), postcentral gyrus (14, -58, 64), and inferior temporal gyrus (54, -62, -10) during coercive harming.

Discussion

With the present fMRI study, we aimed to elucidate the effects of lorazepam on obedience to authority under coercion, by means of a virtual obedience to authority paradigm inspired on Milgram’s experiments of the same nature. We found that reaction times to initiate moral behaviors have an interaction with drug administration and scenario.

It is important to note that reaction times have ever been considered as an objective proxy measure for sedation, which can be a side effect accompanying lorazepam's anxiolytic power (Curran et al., 1998; Mintzer and Griffiths, 2003; Vermeeren et al., 1995), thus impairing psychomotor performance and subsequently increasing reaction times (Smiley, 1987; Van Ruitenbeek et al., 2010). This makes it necessary to verify that any changes caused by lorazepam are not merely the results of such side effect. Our rationale for disentangling the effect of lorazepam due to sedation or due to an anxiolytic effect, was that if the reaction time could be ascribed to the former, participants would exhibit increased reaction times no matter which moral valence they would be responding to. Here, the lorazepam effect was found to have opposite directions depending on the factor of moral valence.

When it comes to the sense of agency, previous literature uncovered explicit and implicit approaches to measure such subjective phenomenon. The implicit index, as measured by the intentional binding effect, was based on the subjective perception of time compression between an action and its effects (Haggard, 2017; Moore, 2016); whereas for the explicit measures for the sense of agency, responsibility and guilt ratings were used (Caspar et al., 2018; Caspar et al., 2020). Nevertheless, the speed in which the decision of a moral action is made, as measured by reaction times, could equally help illuminate an agent’s underlying moral character (Critcher et al., 2013). Agents who make an immoral decision quickly (vs. slowly) are frequently evaluated more negatively by others. Conversely, agents who arrive at a moral decision quicker (vs. slower) receive particularly positive moral character evaluations. Quicker decisions carry this signal value as they are assumed to indicate certainty of behavior, reflecting unambigious motives driving at the backdrop of such actions. On the contrary, the longer an agent takes to act during such task, the more mixed the agent’s motivation is assumed to be, hence the polarized moral character evaluations. Consequently, reaction times may index the subjects’ implicit sense of agency or the

willingness to conform to the instruction of an authority figure, whereas subjective ratings of coercion represent its explicit assessment. The results showed that, under the effects of lorazepam, reaction times showed an increase when harming others but a decrease when helping despite of comparable self-reported ratings. It is reasonable to advance our hypothesis that the anxiolytic drug could help free participants from coercive control, enable them to recover their sense of agency, and follow their own willing as to slow down their harming actions and accelerate those behaviors devoted for helping.

To explain the behavior of the agents under coercion, previous studies have set the players distributed in a rather simple model (Caspar et al., 2016; Caspar et al., 2018). An agent is given an order by an authority figure which dilutes the agent's sense of agency, and leading the agent to perform a harmful behavior to another person (Fig. 4A). We thus see necessary to put forward a new model (Fig. 4B), where an authority figure will issue an order to an agent, triggering anxiety in the latter, as well as the corresponding loss of sense in agency, and making the agent conduct a harming act towards another person. It is in the anxiety portion of the model where lorazepam would let its effect be felt.

The question as to why harming another person would be experienced as a threat to oneself still needs to be resolved. Alongside the above findings, we also observed that coercive harming vs. helping recruited the activity in the amygdala and insula. This finding is in line with previous reports which posit that directly harming others generates an aversive emotional "gut" reaction (Greene et al., 2004). Conflicting with others is a common trigger for, or worsens anxiety (Grupe and Nitschke, 2013; Steimer, 2002). The experimenters' authority during Milgram's experiments, based on French and Raven's classic categories of the bases of social power (French Jr and Raven, 1959) were perceived not only as forms of expert and legitimate power, but also as coercive power (so much that there was no significant difference between the three) (1999; Blass, 1999). As such, the experimenters were seen as capable of administrating punishment (threat) to the participants if they didn't stick to their orders –an anxiogenic source.

Here, it is likely to make a case for lorazepam as a freer of cognitive power –letting participants to reason more clearly. Lorazepam has been suggested to increase ruthlessness, as it was observed that its administration drove participants to endorse harmful behaviors due to the drug's ability to reduce threat intensity during moral dilemmas (Perkins et al., 2013). Anxiety is an emotion which impairs goal-directed actions (Alvares et al., 2014), and that must be overcome by controlled cognitive processes. This notion has been supported by the finding that cognitive load selectively increased reaction times during cognitive reflection (Greene et al., 2008; Paxton et al., 2012), but decreased reaction times when participants were required to respond quickly (Suter and Hertwig, 2011). Anxiolytic drugs are presumed to cause their effects by altering anxiety and/or through cognitive modulation (Richter et al., 2010). Here, we did find that the lorazepam effect did not only decrease the activity in the amygdala and hippocampus, but also increased the activity in the dlPFC and rTPJ. The dlPFC has been consistently implicated in the cognitive control of motor behaviors, monitoring ongoing actions in alignment to internal goals (Miller, 2000; Morris et al., 2014). A number of meta-analyses have demonstrated rTPJ engagement in computational processes associated with sense of agency (Decety and Lamm, 2007; Decety and Sommerville, 2003). It is not surprising to see that lorazepam can increase the dlPFC and rTPJ activity, as

a way to regain cognitive control and sense of agency, which, in turn, slows down coercive harming. Furthermore, the activation in the anterior insular cortex, midcingulate cortex, amygdala, and putamen during the unwilling harming condition, may suggest that the neural responses associated with empathy and guilt processing could be linked to the prolonged RTs or less motivation for harming.

Notably, after lorazepam administration, the weaker activity in the hippocampus during coercive harming predicted higher subjective ratings of coercion (see Fig. 3A). Based on a role-playing game, the two-dimensional geometric model of social relationships, a "social space" framed by power and affiliation, predicted the activity in the hippocampus, suggesting its critical involvement in a map for social navigation (Tavares et al., 2015). Moreover, we also found the negative coupling of the hippocampus with the dlPFC during coercive harming after lorazepam administration. Accordingly, the lorazepam effect could interfere hippocampus activity in response to coercive harming. This would render participants to resist authority and fight against coercive control, in regards with more subjective feelings of coercion, and to express how much the action would violate their own will to a larger degree.

Some limitations of this study should be acknowledged. First, this study may have a confined ecological validity due to the use of a virtual obedience paradigm in its experimental design, as the paradigm could be affected by individual differences regarding the willingness to please the experimenter or to conduct voluntary harmful acts, and which may also affect moral motivation under coercion and the respective RTs. However, due to ethical concerns, as well as for neuroimaging purposes, this was the best course of action that could be thought of and consequently devised. Nevertheless, future research targeting on such individual variation is also warranted. Second, post-session questionnaires to assess the perceived interpersonal behavior of the experimenter might be helpful to evaluate the effects of response expectancy to the intervention in drug-placebo studies (Gaab et al., 2019). Third, although reaction times for moral actions (as mentioned above) may very well index an individual’s sense of agency, the relationship between these two elements warrants further investigation. In our case, the participants may or may not be aware of their delay on the unwilling harming actions or the lorazepam effect of RTs.

In conclusion, the present findings –which incorporate multimodal indices, including functional neuroimaging, neuropharmacological intervention, and behavioral assessments– provide evidence to corroborate the notion that agents under coercive pressure suffer from anxiety, and that the anxiolytic drug lorazepam might help in unveiling the power of authority and assist in the emergence of prosocial behavior. This study sets the base for further research taking into consideration the victims living under serious and insidious coercive violence, e.g., the victims of domestic and occupational violence. By identifying key factors (which may very likely be remorse/anxious feelings resulting from the tug of war between the fear of authority and self-conscience) leading to these individual differences, such research could help turn over the toll of coercion. The latter would be much more helpful and constructive for mankind than purely studying the reasons why some people lost moral conscience under coercion while others did not.

Declarations

Data availability

The data generated in this study are available from the corresponding author on reasonable request.

Acknowledgements

We thank Chun-Ning Hung for assisting with the data collection. We thank Professor Jean Decety for the morally-laden scenarios used as the stimuli, and over which the paradigm was built upon, without his generous contribution such endeavor wouldn't have been possible. The study was funded by the Ministry of Science and Technology (MOST 108-2410-H-010-005-MY3; 108-2410-H-009-020-MY3; 108-2636-H-038-001-; 109-2636-H-038-001-), National Yang-Ming Chiao-Tung University Hospital (RD2019-003), Taipei Medical University (DP2-108-21121-01-N-03-03), and Health Department of Taipei City Government (11001-62-039).

Competing interest

None of the authors have any conflicts of interest to declare.

Author Contributions

Y.C. and C.C. conceived and conceptualized the study. Y.-C.C. and C.C. collected and analyzed the data. Y.C., R.M.M., and C.C. conducted the necessary literature reviews and drafted the first manuscript. Y.T.F. provided critical feedback and helped shape the manuscript. All authors contributed towards the revision and writing the final draft. All authors contributed towards the writing and revision of the final draft.

References

1999. The Milgram Paradigm after 35 years: Some things we now know about obedience to authority. Blackwell Publishing, United Kingdom, pp. 955-978.

Alvares, G.A., Balleine, B.W., Guastella, A.J., 2014. Impairments in goal-directed actions predict treatment response to cognitive-behavioral therapy in social anxiety disorder. PLoS One 9, e94778.

an electronic publication of the Avalon, P., William, C.F., Lisa A. Spar, C.-D., 1996. The Avalon Project at the Yale Law School : documents in law, history and diplomacy. New Haven, Conn. : The Avalon Project, c1996-.

Arce, E., Miller, D.A., Feinstein, J.S., Stein, M.B., Paulus, M.P., 2006. Lorazepam dose-dependently decreases risk-taking related activation in limbic areas. Psychopharmacology (Berl) 189, 105-116.

Arendt, H., 1994. Eichmann in Jerusalem : a report on the banality of evil. Revised and enlarged edition. New York, N.Y., U.S.A. : Penguin Books, 1994.

Association, A.P., 2013. Diagnostic and statistical manual of mental disorders, 5th ed. American Psychiatric Association, Arlington.

Benjamin, L.T., Jr., Simpson, J.A., 2009. The power of the situation: The impact of Milgram's obedience studies on personality and social psychology. Am Psychol 64, 12-19.

Blass, T., 1999. The Milgram Paradigm After 35 Years: Some Things We Now Know About Obedience to Authority1. Journal of Applied Social Psychology 29, 955-978.

Blass, T., 2000. The Milgram paradigm after 35 years: Some thighs we now know about obedience to authority. In: Blass, T. (Ed.), Obedience to authority: Current perspectives on the Milgram paradigm. Lawrence Erlbaum, Mahwah, NJ, pp. 33-59.

Blass, T., 2004. The man who shocked the world: The life and legacy of Stanley Milgram. Basic Books, New York, NY.

Blass, T., 2009. From New Haven to Santa Clara: A historical perspective on the Milgram obedience experiments. Am Psychol 64, 37-45.

Burger, J.M., 2009. Replicating Milgram: Would people still obey today? Am Psychol 64, 1-11.

Burger, J.M., Girgis, Z.M., Manning, C.C., 2011. In their own words: Explaining obedience to authority through an examination of practicipants' comments. Social Psychological and Personality Science 2, 460-466.

Bzdok, D., Schilbach, L., Vogeley, K., Schneider, K., Laird, A.R., Langner, R., Eickhoff, S.B., 2012. Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of mind, and empathy. Brain Struct Funct 217, 783-796.

Caspar, E.A., Christensen, J.F., Cleeremans, A., Haggard, P., 2016. Coercion Changes the Sense of Agency in the Human Brain. Curr Biol 26, 585-592.

Caspar, E.A., Cleeremans, A., Haggard, P., 2018. Only giving orders? An experimental study of the sense of agency when giving or receiving commands. PLoS One 13, e0204027.

Caspar, E.A., Ioumpa, K., Keysers, C., Gazzola, V., 2020. Obeying orders reduces vicarious brain activation towards victims' pain. Neuroimage 222, 117251.

Chen, C., Martinez, R.M., Chen, Y., Cheng, Y., 2020a. Pointing fingers at others: The neural correlates of actor-observer asymmetry in blame attribution. Neuropsychologia 136, 107281.

Chen, C., Martinez, R.M., Cheng, Y., 2020b. The key to group fitness: The presence of another synchronizes moral attitudes and neural responses during moral decision-making. Neuroimage 213, 116732.

Critcher, C.R., Inbar, Y., Pizarro, D.A., 2013. How quick decisions illuminate moral character. Social Psychological & Personality Science 4, 308-315.

Curran, H.V., Pooviboonsuk, P., Dalton, J.A., Lader, M.H., 1998. Differentiating the effects of centrally acting drugs on arousal and memory: an event-related potential study of scopolamine, lorazepam and diphenhydramine. Psychopharmacology (Berl) 135, 27-36.

Decety, J., Lamm, C., 2007. The role of the right temporoparietal junction in social interaction: how low-level computational processes contribute to meta-cognition. Neuroscientist 13, 580-593.

Decety, J., Sommerville, J.A., 2003. Shared representations between self and other: a social cognitive neuroscience view. Trends Cogn Sci 7, 527-533.

First, M.B., Gibbon, M., 2004. The Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I) and the Structured Clinical Interview for DSM-IV Axis II Disorders (SCID-II). In: Hilsenroth, M.J., Segal, D.L. (Eds.), Comprehensive handbook of psychological assessment. John Wiley & Sons Inc., pp. 134–143.

Fray, W.C., Spar, L.A., School, Y.L., 1996. The Avalon Project at the Yale Law School: documents in law, history and diplomacy. The Avalon Project., Now Haven, Conn.

French Jr, J.R.P., Raven, B., 1959. The bases of social power. Studies in social power. Univer. Michigan, Oxford, England, pp. 150-167.

Gaab, J., Kossowsky, J., Ehlert, U., Locher, C., 2019. Effects and Components of Placebos with a Psychological Treatment Rationale - Three Randomized-Controlled Studies. Sci Rep 9, 1421.

Gilbert, S.J., 1981. Another look at the Milgram obedience studies: The role of the graduated series of shocks. Personality and Social Psychology Bulletin 7, 690-695.

Gould, R.A., Otto, M.W., Pollack, M.H., Yap, L., 1997. Cognitive behavioral and pharmacological treatment of generalized anxiety disorder: A preliminary meta-analysis. Behavior Therapy 28, 285-305.

Greene, J.D., Morelli, S.A., Lowenberg, K., Nystrom, L.E., Cohen, J.D., 2008. Cognitive load selectively interferes with utilitarian moral judgment. Cognition 107, 1144-1154.

Greene, J.D., Nystrom, L.E., Engell, A.D., Darley, J.M., Cohen, J.D., 2004. The neural bases of cognitive conflict and control in moral judgment. Neuron 44, 389-400.

Grupe, D.W., Nitschke, J.B., 2013. Uncertainty and anticipation in anxiety: an integrated neurobiological and psychological perspective. Nat Rev Neurosci 14, 488-501.

Haggard, P., 2017. Sense of agency in the human brain. Nat Rev Neurosci 18, 196-207.

Kyriakopoulos, A.A., Greenblatt, D.J., Shader, R.I., 1978. Clinical pharmacokinetics of lorazepam: a review. J. Clin. Psychiatry 39, 16-23.

Lamm, C., Decety, J., Singer, T., 2011. Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. Neuroimage 54, 2492-2502.

Milgram, S., 1963. Behavioral Study of obedience. The Journal of Abnormal and Social Psychology 67, 371-378.

Milgram, S., 1965. Some conditions of obedience and disobedience to authority. Human Relations 18,, 57-76.

Miller, A.G., 2004. What can the Milgram obedience experiments tell us about the Holocaust? Generalizing from the social psychology laboratory. In: Miller, A.G. (Ed.), The social psychology of good and evil. Guilford, New York, NY, pp. 193-239.

Miller, A.G., Collins, B.E., Brief, D.E., 1995. Perspectives on obedience to authority: The legacy of the Milgram experiments. Journal of Social Issues 51, 1-19.

Miller, E.K., 2000. The prefrontal cortex and cognitive control. Nat Rev Neurosci 1, 59-65.

Mintzer, M.Z., Griffiths, R.R., 2003. Lorazepam and scopolamine: A single-dose comparison of effects on human memory and attentional processes. Exp Clin Psychopharmacol 11, 56-72.

Moore, J.W., 2016. What Is the Sense of Agency and Why Does it Matter? Front Psychol 7, 1272.

Morris, R.W., Dezfouli, A., Griffiths, K.R., Balleine, B.W., 2014. Action-value comparisons in the dorsolateral prefrontal cortex control choice between goal-directed actions. Nat Commun 5, 4390.

Packer, D.J., 2008. Identifying Systematic Disobedience in Milgram's Obedience Experiments: A Meta-Analytic Review. Perspect Psychol Sci 3, 301-304.

Patin, A., Hurlemann, R., 2011. Modulating amygdala responses to emotion: evidence from pharmacological fMRI. Neuropsychologia 49, 706-717.

Paxton, J.M., Ungar, L., Greene, J.D., 2012. Reflection and reasoning in moral judgment. Cogn Sci 36, 163-177.

Perkins, A.M., Leonard, A.M., Weaver, K., Dalton, J.A., Mehta, M.A., Kumari, V., Williams, S.C., Ettinger, U., 2013. A dose of ruthlessness: interpersonal moral judgment is hardened by the anti-anxiety drug lorazepam. J Exp Psychol Gen 142, 612-620.

Richter, A., Grimm, S., Northoff, G., 2010. Lorazepam modulates orbitofrontal signal changes during emotional processing in catatonia. Human Psychopharmacology: Clinical and Experimental 25, 55-62.

Robinson, O.J., Vytal, K., Cornwell, B.R., Grillon, C., 2013. The impact of anxiety upon cognition: perspectives from human threat of shock studies. Front Hum Neurosci 7, 203.

Schunck, T., Mathis, A., Erb, G., Namer, I.J., Demazieres, A., Luthringer, R., 2010. Effects of lorazepam on brain activity pattern during an anxiety symptom provocation challenge. J Psychopharmacol 24, 701-708.

Smiley, A., 1987. Effects of minor tranquilizers and antidepressants on psychomotor performance. J Clin Psychiatry 48 Suppl, 22-28.

Steimer, T., 2002. The biology of fear- and anxiety-related behaviors. Dialogues Clin Neurosci 4, 231-249.

Suter, R.S., Hertwig, R., 2011. Time and moral judgment. Cognition 119, 454-458.

Tavares, R.M., Mendelsohn, A., Grossman, Y., Williams, C.H., Shapiro, M., Trope, Y., Schiller, D., 2015. A Map for Social Navigation in the Human Brain. Neuron 87, 231-243.

Van Ruitenbeek, P., Vermeeren, A., Riedel, W.J., 2010. Memory in humans is unaffected by central H1-antagonism, while objectively and subjectively measured sedation is increased. Eur Neuropsychopharmacol 20, 226-235.

Vermeeren, A., Jackson, J.L., Muntjewerff, N.D., Quint, P.J., Harrison, E.M., O'Hanlon, J.F., 1995. Comparison of acute alprazolam (0.25, 0.50 and 1.0 mg) effects versus those of lorazepam 2 mg and placebo on memory in healthy volunteers using laboratory and telephone tests. Psychopharmacology (Berl) 118, 1-9.

Figures

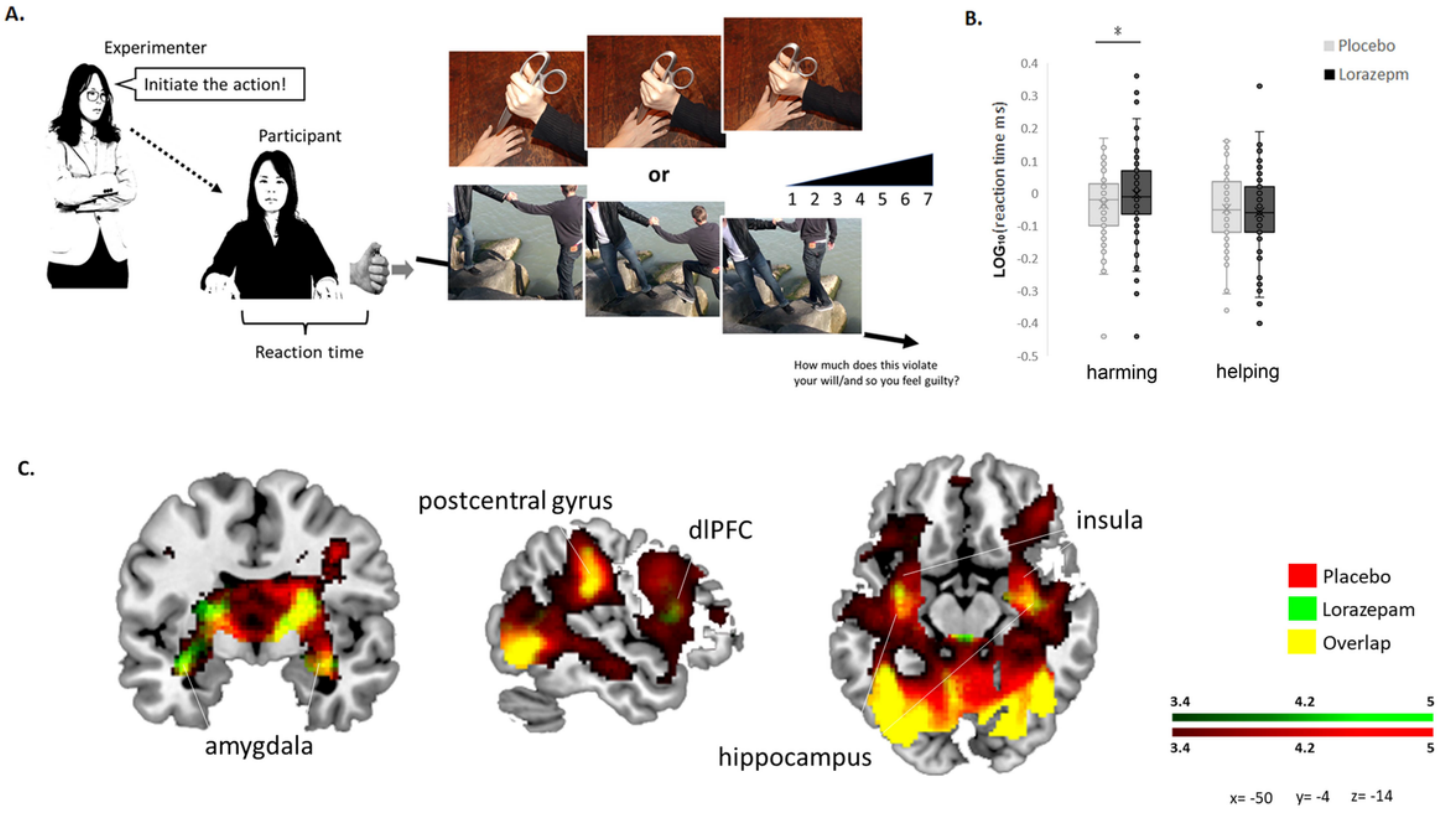

Figure 1

Experimental setup and lorazepam impact on moral decision-making under coercive pressure. A. Schematic representation of the paradigm for coercive commands. The experimenter ordered the participant to initiate moral behaviors by pressing a trigger button in a virtual computerized program along with visual feedback of moral scenarios. B. The reaction time in helping was shorter than that in neutral ($p < .001$) and harming ($p < .001$). Acute lorazepam administration slowed down the harming [lorazepam vs. placebo (mean ± SE): -0.002 ± 0.015 vs. -0.035 ± 0.012, $t76 = 1.715$, $P = .09$, Cohen's $d = 0.196$], whereas it accelerated the helping (-0.059 ± 0.014 vs. -0.049 ± 0.013). C. Lorazepam impact on the whole-brain hemodynamic responses to coercive harming. Results from the whole-brain contrast thresholded at $P < .001$ and cluster extent $k > 10$ for viewing.

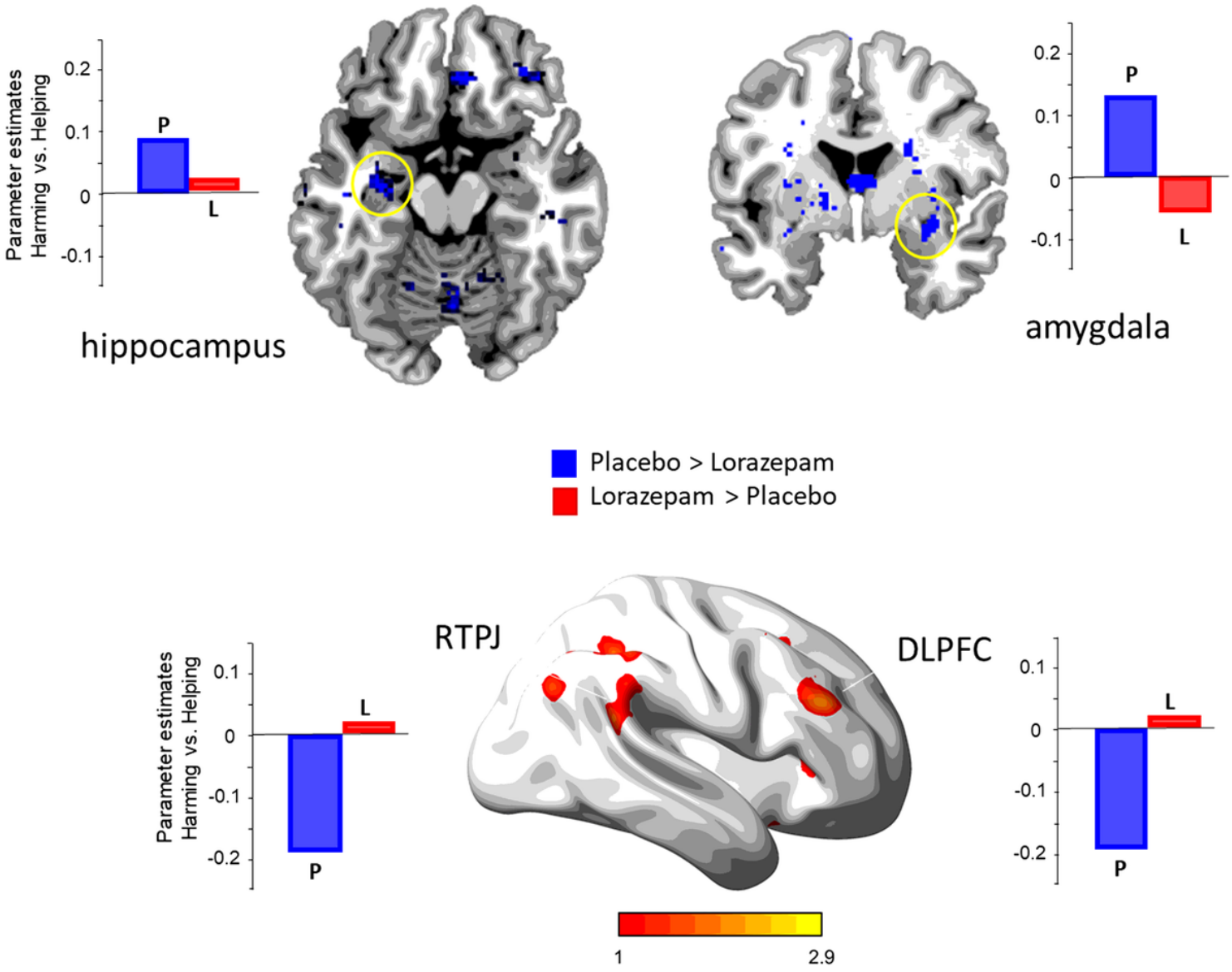

Figure 2

Lorazepam impact on the brain regions involved in coercive harming. The regions of interest (ROIs) included the amygdala (x 30, y 0, z -14), hippocampus (-32, -10, -18), rTPJ (56, -46, 20), and dlPFC (42, 30, 28). Acute lorazepam administration modulated the regions, depending on the factor of scenarios (harming vs. helping). The activity in the amygdala and hippocampus was reduced, whereas the activity in the dlPFC and rTPJ was increased. Clusters from the whole-brain contrast thresholded at P < .05 for viewing.

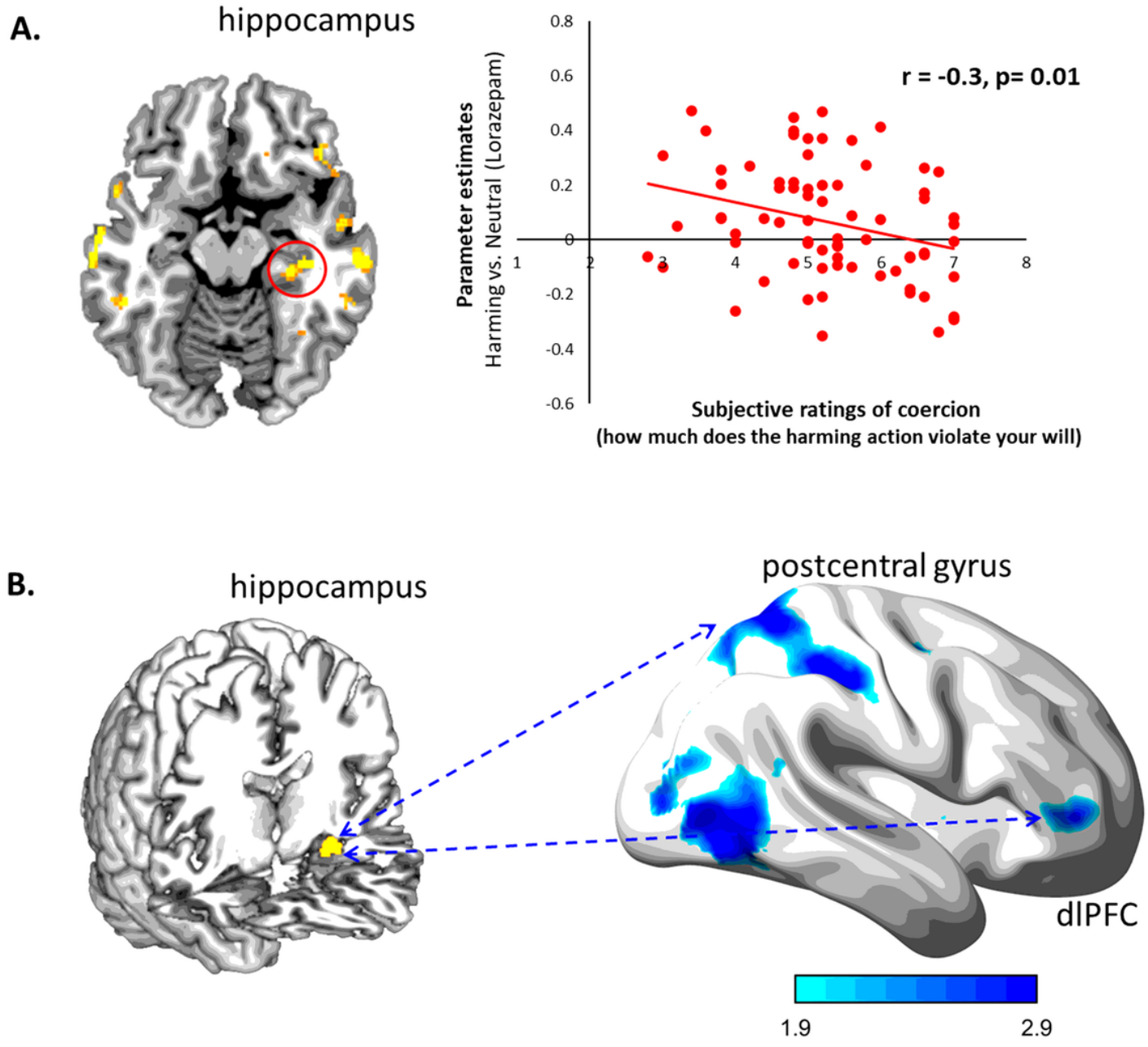

Figure 3

Lorazepam impact on the neural correlates and functional connectivity for subjective experience of coercive harming. A. The subjective experience of coercion was assessed by the violation of free will. After lorazepam administration, less activity in the hippocampus (30, -28, -12) predicted higher subjective ratings of coercion (r = -0.3, P = .01). Clusters from the whole-brain contrast thresholded at P < .05 for

viewing. B. Lorazepam relative to placebo administration significantly reduced the coupling of the hippocampus (-30, -12, -18) with dlPFC (48, 48, 2), postcentral gyrus (14, -58, 64), and inferior temporal gyrus (54, -62, -10) in response to coercive harming. Results from the whole-brain contrast thresholded at $P < .001$ and cluster extent $k > 10$ for viewing.

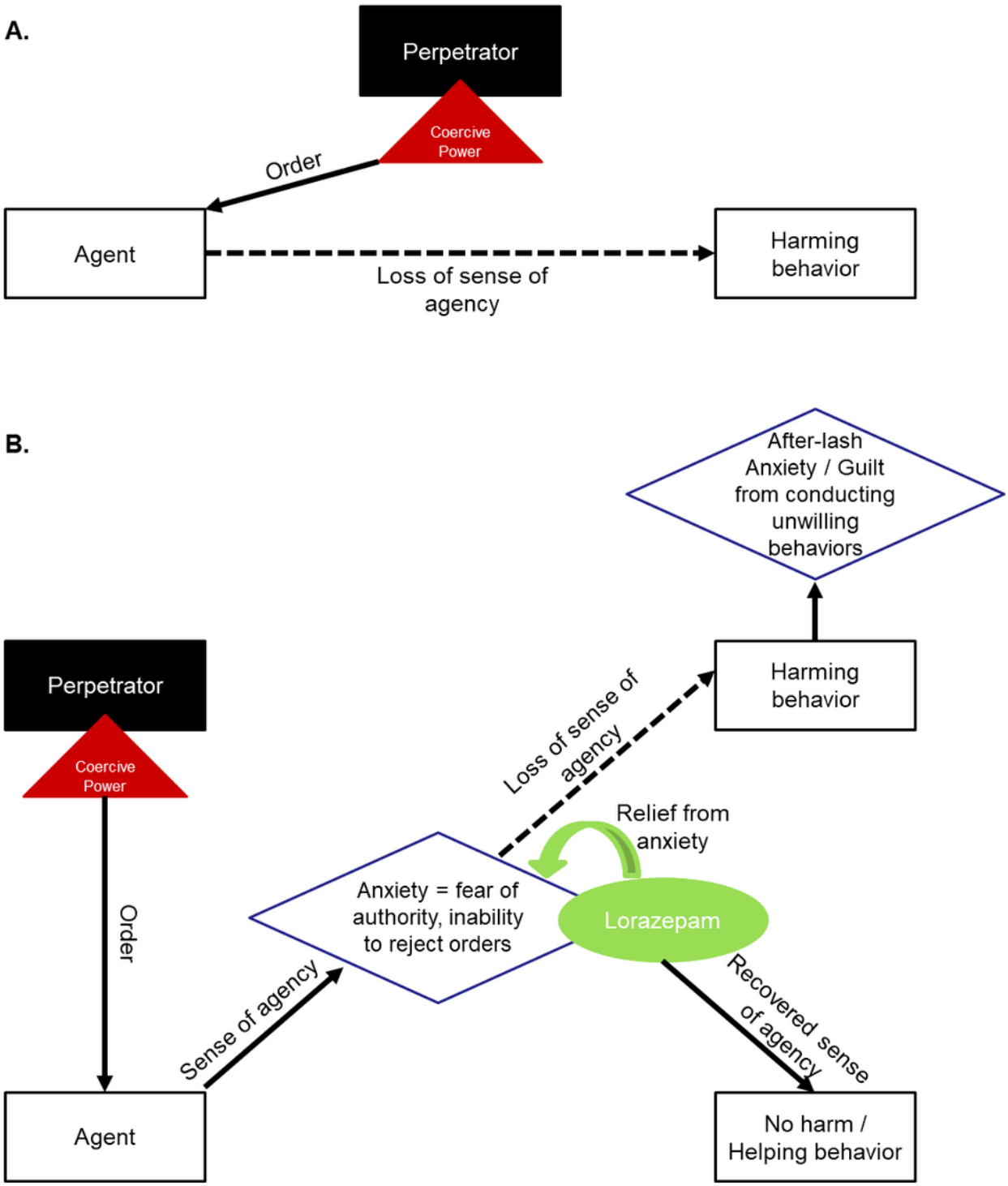

Figure 4

The framework models for obedience to authority under coercion. A. Depicts previous model, where an agent is given a coercive order by an authoritative figure, which dilutes the agent's sense of agency, leading the agent to perform a harmful act towards another person. B. Depicts the new model, where an agent is given a coercive order by an authoritative figure, triggering anxiety in the agent and diminishing the agent's sense of agency, thus, the agent proceeds to perform a harmful act towards another person. After Lorazepam administration, the anxiolytic drug would reduce anxiogenic symptoms produced by coercion, letting the agent free cognitive power in order to reappraise its decisions and act with freedom from authority.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterials01022021.docx