

Ventromedial and dorsolateral prefrontal interactions underlie will to fight and die for a cause

Clara Pretus,^{1,2,10} Nafees Hamid,^{1,3} Hammad Sheikh,^{1,4} Ángel Gómez,^{1,5} Jeremy Ginges,^{1,4} Adolf Tobeña,^{1,2} Richard Davis,^{1,6,7} Oscar Vilarroya,^{1,2} and Scott Atran^{1,7,8,9}

¹Artis International, Scottsdale, AZ 85254, USA, ²Departament de Psiquiatria i Medicina Legal, Universitat Autònoma de Barcelona, Cerdanyola del Vallès 08193, Spain, ³Department of Security and Crime Science, University College London, London WC1H 9EZ, England, ⁴Department of Psychology, New School for Social Research, New York, NY 10011, USA, ⁵Departamento de Psicología Social y de las Organizaciones, Universidad Nacional de Educación a Distancia, Madrid 28040, Spain, ⁶School of Politics and Global Studies, Arizona State University, Tempe, AZ 85287, USA, ⁷The Changing Character of War Centre, Pembroke College, University of Oxford, Oxford OX1 1DW, England, ⁸Centre National de la Recherche Scientifique, Institut Jean Nicod–Ecole Normale Supérieure, Paris 75005, France, ⁹Gerald Ford School of Public Policy and Institute for Social Research, University of Michigan, Ann Arbor 48109, MI USA, and ¹⁰Fundació IMIM (Institut Municipal d'Investigacions Mèdiques), 08003 Barcelona, Spain

Correspondence should be addressed to Clara Pretus, Departament de Psiquiatria i Medicina Legal, Facultat de Medicina, Campus de Bellaterra, 08193, Cerdanyola del Vallès, Barcelona, Spain. E-mail: clara.pretus@gmail.com.

Abstract

Willingness to fight and die (WFD) has been developed as a measure to capture willingness to incur costly sacrifices for the sake of a greater cause in the context of entrenched conflict. WFD measures have been repeatedly used in field studies, including studies on the battlefield, although their neurofunctional correlates remain unexplored. Our aim was to identify the neural underpinnings of WFD, focusing on neural activity and interconnectivity of brain areas previously associated with value-based decision-making, such as the ventromedial prefrontal cortex (vmPFC) and the dorsolateral prefrontal cortex (dlPFC). A sample of Pakistani participants supporting the Kashmiri cause was selected and invited to participate in an functional magnetic resonance (fMRI) paradigm where they were asked to convey their WFD for a series of values related to Islam and current politics. As predicted, higher compared to lower WFD was associated with increased ventromedial prefrontal activity and decreased dorsolateral activity, as well as lower connectivity between the vmPFC and the dlPFC. Our findings suggest that WFD more prominently relies on brain areas typically associated with subjective value (vmPFC) rather than integration of material costs (dlPFC) during decision-making, supporting the notion that decisions on costly sacrifices may not be mediated by cost-benefit computation.

Key words: costly sacrifices; will to fight and die; sacred values; fMRI; functional connectivity

Introduction

In previous studies, our research group has studied beliefs and values embedded within protracted political conflict

across a number of contexts, such as the Israeli–Palestinian conflict (Sheikh *et al.*, 2013), the US–Iranian nuclear conflict (Dehghani *et al.*, 2009), the Indian–Pakistani conflict over

Received: 7 November 2018; **Revised:** 24 March 2019; **Accepted:** 29 April 2019

© The Author(s) 2019. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Kashmir (Sachdeva and Medin, 2009) and the frontline battle against Islamic State (ISIS) in Iraq (Gómez et al., 2017). Across these samples, their report high levels of commitment to their in-group and the values, often showing willingness to personally fight and die to defend them. Thus, the need emerged to develop a suitable measure to assess motivation for incurring extreme personal costly sacrifices including risking one's life and the lives of others to defend highly cherished beliefs. To this end, willingness to fight and die (WFD) was developed as a measure of extreme commitment in field studies of value-based conflicts, whether involving active supporters (Swann et al., 2009; Gómez et al., 2011; Atran et al., 2014a; Atran et al., 2014b) or frontline combatants (Gómez et al., 2017).

Behavioral field studies show that decisions over willingness to engage in costly sacrifices for the sake of a greater cause escape instrumental reasoning and cost-benefit computation (Ginges et al., 2011). For example, willingness to partake in political violence compared to non-violence was mainly predicted by perceived righteousness, with little contribution of perceived effectiveness, whereas support for military compared to diplomatic solutions to a conflict was insensitive to quantity in terms of number of saved lives and number of deterred enemy attacks in the future (Ginges and Atran, 2011).

Value-based choices are generally thought to be modulated by interactions between the dorsolateral prefrontal cortex (dlPFC) and the ventromedial prefrontal cortex (vmPFC) usually favoring adaptive behavior (Hare et al., 2009; Basten et al., 2010; Rudolf and Hare, 2014). Whereas the vmPFC has been proposed as a global value comparator that computes subjective value (Basten et al., 2010; Camille et al., 2011), the dlPFC is thought to modulate value encoding during decisions that integrate costs (Hare et al., 2009) or reweight stimulus attributes (Rudolf and Hare, 2014). However, the interplay between the dlPFC and the vmPFC during value-based decisions has not yet been investigated in the context of WFD. Although WFD should broadly recruit brain networks associated with value-based choice, such as the vmPFC and dlPFC (Plassmann et al., 2007; Hutcherson et al., 2012), decisions involving high compared to low WFD might be expected to show substantial differences in the extent to which brain regions associated with cost-benefit calculation are recruited.

Accordingly, the goal of the present study was 2-fold. First, we sought to identify the neural correlates of WFD by comparing values receiving high vs low WFD ratings, especially focusing on activity in usual decision-making neural hubs such as the dlPFC and vmPFC. Second, we endeavored to establish a functional connectivity profile between these brain regions, respectively associated with integration of costs during decision-making (dlPFC) and subjective value computation (vmPFC), during proceeding of WFD. For this purpose, we recruited a sample of male participants of Pakistani origin living in Barcelona (Spain) who expressly supported the cause of Kashmiri resistance to Indian rule, and then scanned them while they expressed WFD for a series of values related to Islam and current politics. We then analyzed neural activity associated with high and low WFD, and compared functional connectivity between the dlPFC and the vmPFC in these two conditions using the CONN functional connectivity toolbox (The Gabrieli Lab, McGovern Institute for Brain Research). This toolbox allows quantifying the co-variance among previously selected neural regions, or between previously selected neural regions and the rest of the brain, in a given experimental condition.

Given that WFD for a value is a measure that reflects disposition to incur extreme costly sacrifices, we predicted high

compared to low WFD to be positively correlated with brain activity that is typically associated with subjective value during decision-making (vmPFC). Because the personal drive to fight and die for a cause often runs independently of risks and outcomes for the self (Sheikh et al., 2016), we also anticipated lower integration of costs, hence, reduced activity in the dlPFC, during high vs low WFD decisions. For the same reason, we expected functional connectivity between the dlPFC and the vmPFC to be weak or even negative in the high vs low WFD condition.

Methods

Participants

The study was approved by the Institutional Review Board Ethics Committee according to the Declaration of Helsinki guidelines and informed consent was obtained from all participants before taking part in the study. Data protection was a priority in this study and complete anonymity was guaranteed.

Participants were selected by means of field interviews and surveys with a sample of 146 male respondents of Pakistani origin living in migrant marginalized neighborhoods of Barcelona (Spain); the average age was 30.82 years old (range 18 to 62). All respondents were Muslim, and most self-identified as Sunni. In line with the recruitment criteria, all participants in the selected sample expressed (i) support for a jihadist group that fights for the Pakistani reclamation of Kashmir (ii) agreement with armed jihad (iii) agreement that Muslims must unite against the West by all means necessary and (iv) willingness to personally engage or facilitate violence in the name of armed jihad (to engage in violent protest, to financial support a non-state militant group, to join a non-state militant group or to fight and die on their own in support of armed jihad). From those $n=45$ respondents who met all three criteria in the field survey, $n=30$ agreed to participate in the fMRI study.

The final fMRI study sample consisted of male Muslim participants from Pakistani origin living in and around Barcelona. The average age was 30.10 years old (s.d. = 5.37, range from 18 to 39 years old): 51.7% of them were married, 82.7% finished secondary school and 10.3% held a bachelor's degree; 72.4% reported a well below average income; and 72.4% stated that religion was the most important thing in their lives. All but one agreed that 'India should have no right to ownership of Kashmir' and all of them reported a favorable or somewhat favorable opinion of Lashkar-e-Taiba (LeT or 'army of the pure'), one of the largest militant Islamist groups in Pakistan functioning since 1993 and declared a terrorist group by the US (2002) followed by the UN (2005). The organization fights for the liberation of Kashmir from Indian control and the unification of all Muslim majority regions beyond Pakistan, pointing at the US, Israel and India as existential enemies of Islam, as reported by the SATP (South Asia Terrorism Portal, www.satp.org). LeT has been attributed several violent attacks including the 2008 Mumbai bombings, which caused 200 deaths and 300 injured civilians, and the Indian parliament attacks in 2001 (see Box 1 for more details on the Kashmir conflict).

Box 1 The Kashmir conflict started in 1947 with India's independence and its partition from Pakistan. Although it began as a territorial dispute it has evolved into a clash between Hindu and Muslim religious identities especially salient for the Muslim community (Sachdeva and Medin, 2009). The first war between India and Pakistan over

Kashmir in 1947 caused mass killings of all ethnic groups and hundreds of thousands of refugees. In 1948 the UN Security Council imposed a ceasefire, forcing Pakistan to withdraw its troops, and promised a plebiscite for Kashmiris to democratically vote for their independence. Neither Pakistan obeyed nor did the plebiscite take place, creating a climate of escalating violence leading to the Sino-Indian war in 1962 and the Indo-Pakistani war in 1971. The Indian military occupation led to the formation of opposing Muslim militant groups, culminating in the Mujahideen insurgency in 1989, which has caused 70 000 deaths and 8000 missings to date. The most recent ceasefire resolution between Pakistan and India was signed in 2003, although there have been several violent attacks from both sides ever since. Currently, 43% of the territory is controlled by India, 37% by Pakistan and 20% by China. The general Pakistani view defends that Kashmir should be independent or annexed to Pakistan, according to its Muslim majority and responding to the Kashmiri uprising.

Procedure

Participants were invited to the fMRI facilities where they completed a pre-scan survey and an fMRI task described below. The pre-scan survey was used to determine which values were perceived as sacred (non-negotiable) and non-sacred (mundane values) by each participant in order to ensure sufficient variability in the WFD measure. A list of 25 candidate values was presented to participants, which was designed based on previous ethnographic fieldwork and included topics relevant to the sample population: 'Western military forces should be expelled from Muslim lands', 'Strict sharia should be applied in all Muslim countries', 'Armed jihad should be waged against the enemies of Muslims', 'India should have no right to ownership of Kashmir', 'Current Muslim countries should be unified in a single Caliphate', among others (see full list in Table S1). In line with previous work on sacred values (Baron and Spranca, 1997; Tetlock, 2003; Berns et al., 2012), a value was considered sacred if: (i) participants refused any material incentive to give it up, such as better economic conditions for their own community, that is, they responded 'No' instead of 'Yes' or 'Maybe' to the question 'Would you be willing to give up this value if adopting the opposite position entailed better economic conditions for your community?', and (ii) their stance was immune to social influence, that is, they responded 'No' instead of 'Yes' or 'Maybe' to the question 'If there was a democratic consensus over the opposite position, would you find it acceptable?' (Fiske and Tetlock, 1997; Baron and Leshner, 2000; Ginges et al., 2007; Berns et al., 2012; Sheikh et al., 2013). Participants had to respond 'No' to both questions in order for the value to be considered sacred. Next, participants ranked their sacred and non-sacred values by importance and the top 6 sacred values, as well as 6 randomly chosen sacred values were selected for the fMRI paradigm. There was a considerable amount of overlapping between each participants' sacred value and non-sacred value sets, with the most frequently employed sacred values being those including religious content (see Table S1). Measures of identity fusion with seven different groups (Swann et al., 2009), including 'Islam', 'Muslim Ummah', 'Pakistan', 'Spain', 'Borderless Muslim Caliphate', 'Federated Muslim Caliphate' and 'Close Friends', were also administered in the pre-scan survey. The identity fusion measure includes

five pairs of circles, one representing the self (a small circle) and another representing the group (a big circle), with varying degrees of overlap between the circles. From the five options, only the option representing the small circle fully inside the big circle denotes fusion. We added an additional response option to the original scale showing no contact at all between the two circles owing to the reluctance of some respondents to accept any contact between the circles in previous field studies (Gómez et al., 2017). These were intended to measure the extent to which respondents felt 'viscerally connected' to a reference group such as nation or religious community (Gómez et al., 2017).

fMRI paradigm

The rapid event-related fMRI paradigm was designed using MATLAB 2012a (The MathWorks, Inc., Natick, Massachusetts, USA) with Psychtoolbox extensions (Kleiner et al., 2007). During brain scanning, participants rated their WFD for each value (one value per trial) by means of a 7-point Likert scale ranging from 1 (not willing) to 7 (extremely willing). Values were presented in Urdu (see Figure S1). Linguistic variations of each value produced 80 different trials (40 sacred value trials and 40 non-sacred value trials) with controlled length and sentence structure in order to increase the number of trials within the fMRI paradigm without wording repetitions. A *post hoc* analysis comparing stimulus length in Urdu showed no differences in stimulus length neither between sacred and non-sacred value trials (sacred values: $M=12.05$ words, $s.d.=\pm.28$; non-sacred values: $M=11.97$ words; $s.d.=\pm.6$, $t(28)=.598$, $P=.555$) nor between the conditions of interest, that is, high and low WFD (high WFD: $M=12.08$ words, $s.d.=\pm.36$; low WFD: $M=12.04$ words, $s.d.=\pm.76$, $t(28)=.204$, $P=.840$). Stimuli were presented randomly with jittered inter-trial intervals in two runs. Inter-trial interval durations varied randomly from 2.5 s up to 12.5 s in 2.5 s steps. Participants were instructed to read the value and, once they knew their response, convey their WFD making use of the 7-point Likert scale located below the value using left/right buttons. The participant's response was recorded based on their button presses by means of Psychtoolbox functions on Matlab. Pilot testing with non-selected candidates from the same population allowed establishing a response time of 10 s (reading + responding), which allowed all participants to respond to most trials. The values and the 7-point Likert scale were always presented in the same location of the screen and the cursor's starting point was always at 4 points, the center of the scale (if their response was actually a 4, they were instructed to go either left or right and then back to 4, otherwise the trial counted as a null). The value remained on screen until the 10 s were over. Keeping stimuli in place facilitated reading and reduced confusion. Before starting the fMRI paradigms, participants practiced the task in a computer until fully comprehending the exercise and mastering control of the response buttons.

fMRI acquisition

Images were acquired in a Siemens 3 T scanner. T1-weighted images were obtained using a fast spoiled gradient-echo (FSPGR) sequence (TR: 11.6 ms, TE: 4.8 ms, FA: 12, matrix size: 280×280 , 150 slices, slice thickness: 1.00 mm). An EPI-T2* sequence allowed obtaining the functional volumes, each comprising forty 3.4 mm thick slices (TR 2500 ms, TE: 27 ms, FA: 90, matrix size: 64×64 , 40 slices, slice thickness: 3.4 mm). Head motion was tracked by means of center of mass measurements during image

acquisition. None of the participants showed head movement above voxel size acquisition after a *post hoc* evaluation of the six head movement parameters.

Statistical Parametric Mapping (SPM12) (Wellcome Department of Imaging Neuroscience, London, UK) and the Analysis of Functional Neuroimaging software were employed for the functional MRI data analysis. Functional images were despiked Analysis of Functional NeuroImages (AFNI), corrected for motion-related artifacts (SPM12), normalized to MNI standard space (SPM12), smoothed with a 8 mm full-width-at-half-maximum Gaussian kernel (SPM12) and detrended (AFNI).

Behavioral analysis

The behavioral analysis included a paired sample t-test comparing WFD ratings in the high WFD condition and the low WFD condition. The effect of value sacredness on WFD ratings was computed by means of a repeated measures analysis of variance (ANOVA) assessing value sacredness as a within-subjects factor. Percentages of fused participants (6 out of 6 points) were calculated for each identity fusion scale and correlations between identity fusion and WFD ratings were conducted.

Neuroimaging analysis

Image processing was conducted using the SPM12 function package (Wellcome Trust, University College London, UK) on MATLAB. Trials were split in high vs low WFD as a function of each participant's average. The first-level general linear model (GLM) included two regressors for trials where participants conveyed either high (above each participant's mean) or low (below each participant's mean) WFD ratings. The average number of trials was 47.2 (s.d. = ± 8.82) in the high WFD condition, and 32.8 (s.d. = ± 8.82) in the low WFD condition. The duration of the events spanned from stimulus onset up to the participant's first button press, and thus, it varied from trial to trial. Each condition was parametrically modulated by a binary regressor with ones and zeros, including whether the value at stake was sacred (1) or not (0). Of note, the high and low WFD conditions did not always correspond to responses to sacred and non-sacred values, respectively. That is, the high WFD condition included, at average, 77.11% trials with sacred values (s.d. = 13.96%) and the low WFD condition included, at average, 12.81% trials with sacred values (s.d. = 13.94%). Because sacred values entail higher levels of attitude strength, familiarity, salience and emotional intensity than non-sacred values (Pretus et al., 2018; Hamid et al., 2019), the inclusion of a value sacredness parametric regressor was crucial to account for these confounding effects on the measures of interest (WFD). Missing trials (no button response throughout the trial), which were, at average, 3.62 trials per participant (s.d. = ± 5.44) were not modeled in the GLM, and thus, they were processed as part of the implicit baseline. In turn, in order to account for the confounding effects of motor preparation and response, we introduced an hrf-convolved button press regressor. Six non-convolved head movement regressors were also included.

Group-level effects of the high vs low WFD contrast were evaluated by means of a t-test on individual level contrasts. In addition, we ran a second-level regression with identity fusion with Islam and the Muslim Ummah as a predictor of neural activity in the high vs low WFD contrast as well as in high and low vs baseline contrast, respectively. In order to be able to capture more subtle differences in neural activity associated with identity fusion, identity fusion was used as a continuous

variable in this analysis. Post hoc analyses included parameter estimate extraction of vmPFC and left and right dlPFC activity in the high WFD vs baseline and the low WFD vs baseline contrast using Marsbar in order to illustrate the amount of region of interest (ROI) activity in each condition. Reported results were corrected for multiple comparisons using family-wise error (FWE) correction at cluster level, single voxel $P < 0.001$.

Effective connectivity analysis

The functional connectivity analysis was conducted by means of the connectivity toolbox CONN on SPM12 (Whitfield-Gabrieli and Nieto-Castanon, 2012). To test the interplay between brain regions associated with executive control (dlPFC) and subjective value computation (vmPFC) during high and low WFD, a series of anatomical ROIs were defined using Automated Anatomical Labeling (aal) labels (Neurofunctional Imaging Group, Université de Bordeaux). These ROIs included the left dlPFC (left middle frontal gyrus), the right dlPFC (right middle frontal gyrus) and the vmPFC (bilateral medial orbitofrontal gyrus). For each participant, a generalized Psycho-Physiological Interaction (gPPI) analysis was conducted. The gPPI analysis tests the task-modulated effective connectivity between pairs of a previously defined set of ROIs (here: right dlPFC, left dlPFC and vmPFC) by means of a multiple regression model incorporating task effects, the ROIs timeseries and an interaction term (PPI term) between both parts. The effects of the high and low WFD conditions on effective connectivity between the bilateral dlPFC and the vmPFC were then computed and compared. Group-level gPPI results were corrected for multiple comparisons at seed level ($P < 0.05$ FWE).

Results

The average WFD rating in the high WFD condition was 6.67 (s.d. ± 0.48), whereas the average WFD rating in the low WFD condition was 3.94 (s.d. ± 1.02). The average response time was 5.93 s (s.d. = ± 1.03), and no differences in response time were detected between conditions (High WFD condition: $M = 5.02$, s.d. = ± 1.22 ; Low WFD condition: $M = 5.04$, s.d. = ± 1.38 ; $t(28) = -.089$, $P = .930$). In addition, response time did not correlate with WFD ratings, neither when analyzing all trials together ($r = -.032$, $P = .868$) nor when evaluating each condition separately (high WFD condition: $r = .070$, $P = .718$; low WFD condition: $r = .042$, $P = .828$). Indeed, value sacredness as a within-subjects factor had a significant effect on WFD ($F(1) = 168.612$, $P < .001$).

Seventy-five percent of the sample was fused with Islam, 72.4% with the Muslim Ummah, 93.1% with Pakistan, 31% with Spain, 21% with a Borderless Muslim Caliphate, 24% with a Federated Muslim Caliphate and 79.3% with Close Friends. Identity fusion with Pakistan was the only identity fusion measure that correlated with WFD ($r = 0.388$, $P < 0.038$), although this result did not survive Bonferroni correction.

fMRI signal contrasts between high vs low WFD appear in Table 1. The high compared to the low WFD condition was associated with greater activity in a cluster encompassing the anterior cingulate cortex and the left middle orbital gyrus ($T = 4.49$, $k = 446$, $P = 0.013$ FWEc) and lower activity in the right inferior parietal lobule ($T = 5.49$, $k = 1690$, $P < 0.001$ FWEc), right middle frontal gyrus ($T = 5.47$, $k = 424$, $P = 0.016$ FWEc), left superior medial frontal gyrus ($T = 5.37$, $k = 735$, $P = 0.001$ FWEc), left inferior parietal lobule ($T = 5.33$, $k = 1800$, $P < 0.001$ FWEc) and a cluster including the left inferior frontal

Table 1. Results of the high vs low WFD contrast thresholded at $T = 3.4$, $P < 0.05$ FWE cluster-level corrected, single level $P < 0.001$

Region label	k	T(28)	MNI coordinates			p-FWEc
			x	y	z	
High > Low WFD						
L ACC	446	4.49	-4	38	0	0.013
Location not in atlas	446	4.07	26	38	2	0.013
L Mid orbital gyrus	446	3.52	-2	62	-2	0.013
Low > high WFD						
R Supramarginal gyrus	1690	-5.49	56	-32	54	<0.001
R Angular gyrus	1690	-3.88	40	-72	48	<0.001
R Middle frontal gyrus	424	-5.47	48	36	32	0.016
L Superior medial gyrus	735	-5.37	0	24	46	0.001
L Inferior parietal lobule	1800	-5.33	-54	-40	54	<0.001
L Inferior parietal lobule	1800	-5.05	-36	-44	46	<0.001
L Superior parietal lobule	1800	-4.54	-26	-68	62	<0.001
L IFG (pars triangularis)	1230	-5.11	-40	20	6	<0.001
L IFG (pars triangularis)	1230	-4.99	-52	14	32	<0.001
L Middle frontal gyrus	1230	-3.50	-44	34	32	<0.001

gyrus (pars triangularis) and left middle frontal gyrus ($T = 5.11$, $k = 1230$, $P < 0.001$ FWEc; see [Figure 1A and B](#)). Post hoc parameter estimates analyses corroborated the increased vmPFC activity ($t(28) = 3.25$, $P = 0.003$) and decreased left PFC ($t(28) = -2.19$, $P = 0.037$) and right dlPFC ($t(28) = -3.00$, $P = 0.006$) activity in the high WFD vs baseline contrast compared to the low WFD vs baseline contrast, although the left dPFC comparison did not survive Bonferroni correction (see [Figure 1B](#)). Differences in WFD scores in the high vs low WFD conditions correlated with neural activity differences in the left dlPFC ($r = -0.578$, $P = 0.001$) and right dlPFC ($r = -0.456$, $P = 0.013$), but not the vmPFC, between conditions after Bonferroni correction (see [Figure 1C](#)).

Study participants were predominantly fused with Islam and the Muslim Ummah; however, the available variability was negatively associated with left middle frontal gyrus activity in the high WFD vs baseline contrast ($T = 5.01$, $k = 489$, $P = 0.024$ FWEc) and in the high vs low WFD contrast ($T = 4.38$, $k = 71$, $P = 0.045$ FWEc after small volume correction with the left dlPFC ROI, see [Figure 2](#)). However, identity fusion with Islam and the Muslim Ummah did not significantly correlate with average WFD scores ($r = 0.141$, $P = 0.467$) nor with WFD scores in the high ($r = 0.058$, $P = 0.766$) or low ($r = 0.066$, $P = 0.735$) WFD condition, probably due to a limited variability of identity fusion in this sample.

The gPPI analysis revealed a negative effective connectivity between the vmPFC as a seed region and the bilateral dlPFC ($F(2)(27) = 2.98$, Intensity = 4.25, $P = 0.01$ FWE seed-level corrected) when comparing the high vs low WFD conditions (see [Figure 3](#)). Results of the reverse contrast (low > high WFD) did not survive multiple comparisons correction.

Discussion

The present study aimed to identify the neural underpinnings of WFD for the sake of values related to Islam in a sample of Pakistani participants supporting the cause of Kashmiri resistance to Indian rule. As anticipated, decisions involving high compared to low WFD (with respect to each participants' average) correlated with higher ventromedial prefrontal activity (vmPFC), a brain region typically associated with subjective value computation during decision-making ([Plassmann et al., 2007](#); [Bartra et al., 2013](#)). Moreover, consistent with behavioral field studies indicating that decisions over costly sacrifices run independently from

cost-benefit computation, high compared to low WFD negatively correlated with brain activity associated with integration of costs and stimulus reweighting, that is, the dlPFC ([Hare et al., 2009](#); [Rudolf and Hare, 2014](#)). In addition, the effective connectivity analysis revealed a stronger negative correlation between the vmPFC as a seed region and the bilateral dlPFC in the high vs low WFD condition.

Other neuroimaging studies on decision-making report interactions between the vmPFC and the dlPFC during value-based choices. For example, [Baumgartner et al. \(2011\)](#) observed increased dlPFC activity and higher functional connectivity between the dlPFC and the vmPFC during decisions where fairness and self-interest were in conflict (i.e. in participants who more frequently made fair decisions vs participants who made more selfish decisions, but no between-group differences in non-conflicting decisions). Similarly, [Hare et al., 2009](#) found higher dlPFC activity and increased dlPFC-vmPFC connectivity during decisions about food consumption in dieting 'self-controllers' (participants evaluating both taste and health for each food item) vs dieting 'non-self-controllers' (evaluating taste only); and there was a correlation between vmPFC activity and food value regardless of self-control.

Nevertheless, none of these prior studies investigated the relationship between these brain areas in regard to decisions involving costs in human lives. Here, we found that decisions over WFD for a cause were modulated by interactions between the vmPFC and the dlPFC. Particularly, the vmPFC was more active, and the dlPFC, together with other fronto-parietal areas, such as the left inferior frontal gyrus and the bilateral parietal lobe, was less active during decisions involving high WFD for a cause. Several studies point at the vmPFC as an integration hub that reflects the subjective value of a given choice 'all things considered' ([Chib et al., 2009](#); [Peters and Büchel, 2010](#); [Bartra et al., 2013](#); [Shenhav and Greene, 2014](#)). Moreover, the dlPFC has been associated with a myriad of functions related to executive control, such as stimulus reweighting ([Rudolf and Hare, 2014](#)) and integration of costs ([Hare et al., 2009](#)). More generally, the fronto-parietal network, including areas found to be less active in the high vs low WFD conditions (inferior frontal gyrus and the parietal lobules), has been associated with cognitive control enabling rapid and flexible goal-oriented behavior ([Marek and Dosenbach, 2018](#)). From this perspective, our results suggest

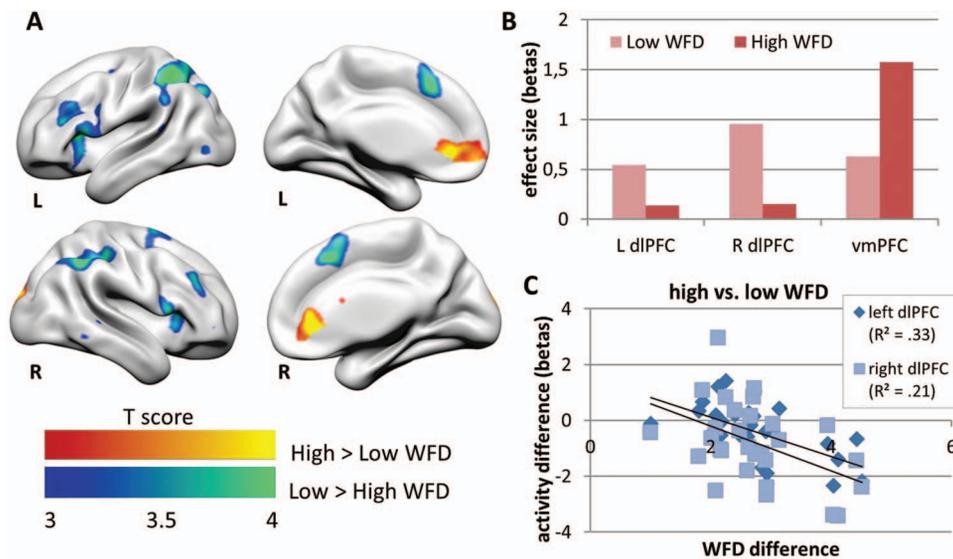


Fig. 1. Results of the high vs low WFD contrast. (A) The high vs low WFD condition yielded higher activity in a cluster including the anterior cingulate cortex (ACC) and the vmPFC (in red-yellow) and lower activity in the superior medial frontal gyrus, the right middle frontal gyrus, Left inferior frontal gyrus (pars triangularis) and bilateral inferior parietal lobule (in blue-green). Results thresholded at $T = 3.4$, $P < 0.05$ FWEc, single level $P < 0.001$. (B) Parameter estimates of vmPFC activity were higher in the high WFD vs baseline condition compared to the low WFD vs baseline condition ($t(28) = 3.25$, $P = 0.003$), while those of the bilateral dlPFC were lower (left dlPFC: $t(28) = 2.19$, $P = 0.037$; right dlPFC: $t(28) = 3.00$, $P = 0.006$) in the same contrast. (C) Between-conditions difference in WFD ratings correlated with between-conditions left dlPFC ($r = -0.578$, $P = 0.001$) and right dlPFC ($r = -0.456$, $P = 0.013$), but not with vmPFC activity difference, after Bonferroni correction for three tests (one per ROI).

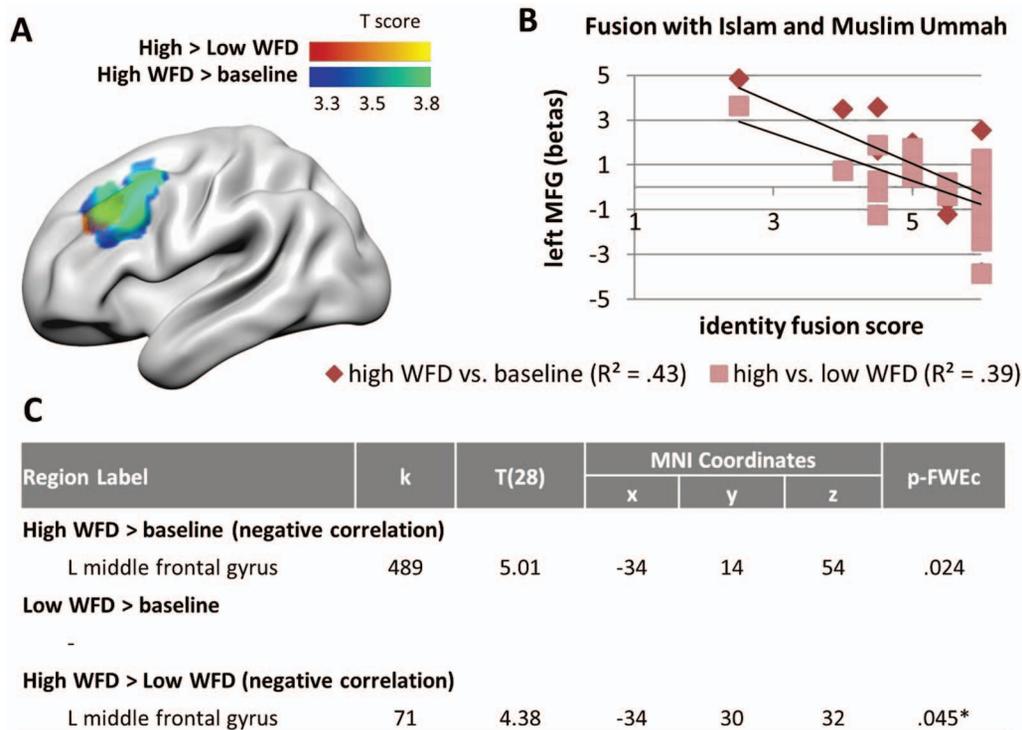


Fig. 2. Correlations between neural activity and identity fusion measures. (A) Surface projection of the left middle frontal gyrus activity predicted by fusion with Islam and the Muslim Ummah in the high WFD vs baseline contrast ($T = 5.01$, $k = 489$, $P = 0.024$ FWEc) and in the high vs low WFD contrast ($T = 4.38$, $k = 71$, $P = 0.045$ FWEc after small volume correction with the left dlPFC ROI). (B) Scatterplot of the correlation between the fusion score and left middle frontal gyrus parameter estimates found in the high WFD vs baseline contrast ($r = -0.659$, $P < 0.001$) and the high vs low WFD contrast ($r = -.627$, $P < 0.001$). (C) Results table of the second-level GLM using identity fusion with Islam and the Muslim Ummah as predictor of brain activity in the high and low WFD vs baseline contrasts, respectively, and the high WFD vs low WFD contrast.

that high (vs low) WFD choices could involve higher subjective value, with little contribution of brain regions associated with integration of costs and cognitive control (including the

dlPFC). In addition, high compared to low WFD was associated with negative effective connectivity between the vmPFC as a seed region and the bilateral dlPFC. This finding suggests

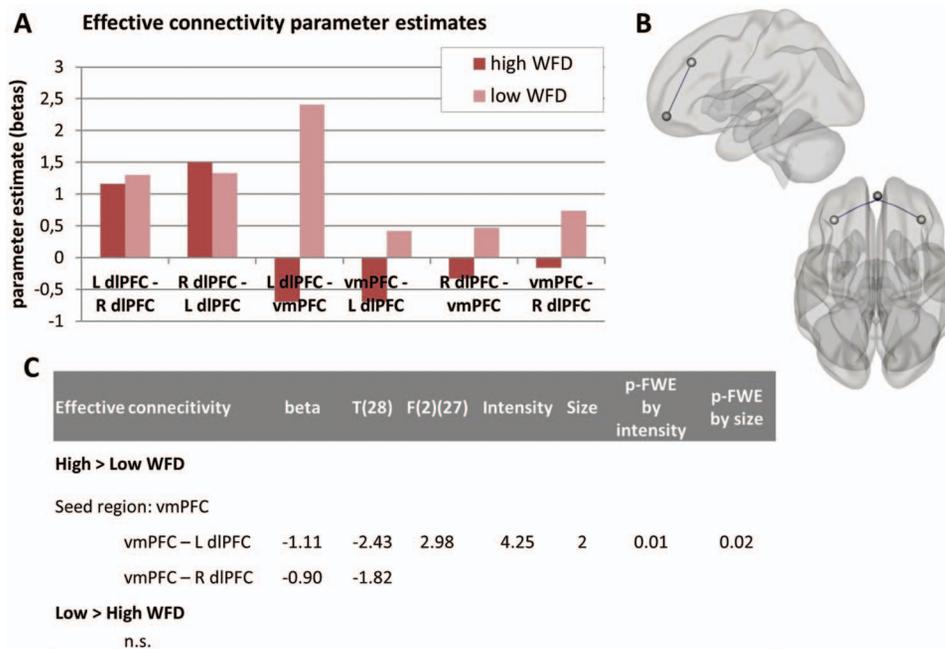


Fig. 3. Effective connectivity analysis results. (A) Parameter estimates of the high WFD effects (dark red) and the low WFD effects (light red) on the effective connectivity between each pair of ROIs including the left and right dlPFC and the vmPFC ('seed region'–[target region]). (B) 3D display of the effective connectivity differences between the high and low WFD conditions after FWE seed-level correction (left lateral and superior view). (C) Effective connectivity results table comparing high > low WFD including a negative correlation between the vmPFC as a seed region and the bilateral dlPFC; no significant results were found in the reverse contrast (low > high WFD).

that a greater vmPFC response contributes to bilateral dlPFC downregulation during decisions involving high WFD for a cause. In this vein, future research should clarify whether stimulating dlPFC activity effectively decreases WFD ratings for causes that would otherwise be associated with high WFD responses.

Finally, most participants were fused with Pakistan, Islam and the Muslim Ummah, in line with studies showing identity fusion in frontline fighters, who establish family-like bonds with fellow combatants (Whitehouse et al., 2014; Gómez et al., 2017). Available variability revealed a positive correlation between identity fusion with Islam and the Muslim Ummah and decreased dlPFC recruitment during the high WFD condition (both compared to baseline and compared to the low WFD condition), although this did not entail higher WFD ratings. Identity fusion is associated with commitment and costly sacrifices in defense of in-group values (Swann et al., 2014). Identity fusion is one of the components of the 'devoted actors model', which characterizes individuals whose personal identity is merged or 'fused' with that of their social group and to the group's most cherished (sacred) values (Atran et al., 2014a; Atran, 2016; Sheikh et al., 2016; Gómez et al., 2017). For similar levels of in-group commitment (similar WFD ratings), fused participants showed a lower reliance on brain areas typically associated with integration of costs during decisions involving costly sacrifices (left dlPFC). This finding suggests that the effect of identity fusion on dlPFC activity alone cannot account for the observed high WFD making use of the present experimental design.

Our results are consistent with behavioral studies showing that decisions over willingness to engage in costly sacrifices for the sake of a greater cause escape instrumental reasoning and cost-benefit computation (Ginges and Atran, 2011). In those prior studies, decisions over willingness to resort to political

violence (military action) compared to non-violence (diplomacy) was mainly predicted by perceived righteousness, with little contribution of perceived effectiveness, whereas support for military compared to diplomatic solutions to a conflict was insensitive to quantity in terms of number of saved lives and number of deterred future attacks by the enemy (Ginges and Atran, 2011). Participants were ready to support a military intervention to resolve a hypothetical international conflict if it meant saving the life of a single hostage or deterring one future attack. By contrast, diplomatic solutions were favored only if they involved saving more than 100 hostages or 60 future attacks (Ginges and Atran, 2011).

The present study has several limitations. On one hand, values in the high and low WFD conditions differed in a number of ways, including emotional intensity, familiarity and salience. This owes to the fact that sacred values (higher emotional intensity, familiarity and salience) were more often than not included in the high WFD condition, whereas the opposite was true for non-sacred values. This confound was addressed by including a binary parametric regressor in the high and low conditions indicating whether the value at stake was sacred or not. However, we understand that if a value is worth fighting and dying for, it will almost certainly entail higher levels of emotional intensity and salience. Hence, we tend to think that such qualities are intrinsic (rather than a confound) to high levels of WFD. On the other hand, it would have been interesting to compare WFD and willingness to pay for goods in the same individuals in order to evaluate the degree of overlap between the functional clusters underlying value-based choice in the social cause vs material goods condition. Further research is needed to specifically test whether the pattern of dlPFC–vmPFC interactions is consistent across social and material values, or whether discrepancies modulated by individual traits, such as identity fusion, exist.

Finally, our study is limited by the fact that it focuses on a population involved in a particular conflict. Future research should compare behavioral and neuronal processes associated with WFD to fight and die across communities involved in different world conflicts.

In sum, our findings support the notion that WFD relies on brain activity that is generally associated with subjective value computation (vmPFC). Moreover, in line with behavioral data showing that decisions on costly sacrifices run independently from cost-benefit ponderation (Ginges *et al.*, 2011), high compared to low WFD was characterized by decreased recruitment of brain regions previously associated with integration of costs during decision-making, such as the dlPFC. Accordingly, our data also reveal a negative effective connectivity between the vmPFC and the bilateral dlPFC associated with high compared to low WFD. By providing a neurobiological perspective via fMRI measures and brain functional connectivity, these findings are potentially relevant to identifying later stages of radicalization that involve WFD for a cause.

Funding

This work was supported by the Minerva Program and the Air Force Office of Scientific Research of the U.S. Department of Defense (AFOSR FA9550-14-1-0030 DEF) and the BIAL Foundation (Grant #163/14) and the Ministerio de Ciencia, Innovación y Universidades (RTI2018-093550-B-I00).

Conflict of interest. None declared.

Acknowledgements

We thank everyone who participated in this study.

References

- Atran, S. (2016). The devoted actor: Unconditional commitment and intractable conflict across cultures. *Current Anthropology*, *57*, S192–203.
- Atran, S., Sheikh, H., Gomez, A. (2014a). Devoted actors sacrifice for close comrades and sacred cause. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, 17702–3.
- Atran, S., Sheikh, H., Gomez, A. (2014b). For cause and comrade: devoted actors and willingness to fight. *Clodynamics: The Journal of Quantitative History and Cultural Evolution*, *5*, 41–57.
- Baron, J., Leshner, S. (2000). How serious are expressions of protected values. *Journal of Experimental Psychology: Applied*, *6*, 183–94.
- Baron, J., Spranca, M. (1997). Protected values. *Organizational Behavior and Human Decision Processes*, *70*, 1–16.
- Bartra, O., McGuire, J.T., Kable, J.W. (2013). The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage*, *76*, 412–27.
- Basten, U., Biele, G., Heekeren, H.R., *et al.* (2010). How the brain integrates costs and benefits during decision making. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 21767–72.
- Baumgartner, T., Knoch, D., Hotz, P., *et al.* (2011). Dorsolateral and ventromedial prefrontal cortex orchestrate normative choice. *Nature Neuroscience*, *14*, 1468–74.
- Berns, G.S., Bell, E., Capra, C.M., *et al.* (2012). The price of your soul: neural evidence for the non-utilitarian representation of sacred values. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*, 754–62.
- Camille, N., Griffiths, C.A., Vo, K., *et al.* (2011). Ventromedial frontal lobe damage disrupts value maximization in humans. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*, 7527–32.
- Chib, V.S., Rangel, A., Shimojo, S., *et al.* (2009). Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. *Journal of Neuroscience*, *29*(39), 12315–20.
- Dehghani, M., Iliev, R., Atran, S., *et al.* (2009). Emerging sacred values: the Iranian nuclear program. *Judgment and Decision Making*, *4*, 990–3.
- Fiske, A.P., Tetlock, P.E. (1997). Taboo trade-offs: reactions to transactions that transgress the spheres of justice. *Political Psychology*, *18*, 255–98.
- Ginges, J., Atran, S. (2011). War as a moral imperative (not just practical politics by other means). *Proceedings of the Royal Society of London B: Biological Sciences*, *278*, 2930–8.
- Ginges, J., Atran, S., Medin, D., *et al.* (2007). Sacred bounds on rational resolution of violent political conflict. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 7357–60.
- Ginges, J., Atran, S., Sachdeva, S., *et al.* (2011). Psychology out of the laboratory: the challenge of violent extremism. *American Psychologist*, *66*, 507–19.
- Gómez, Á., Brooks, M.L., Buhrmester, M.D., *et al.* (2011). On the nature of identity fusion: insights into the construct and a new measure. *Journal of Personality and Social Psychology*, *100*, 918–33.
- Gómez, Á., López-Rodríguez, L., Sheikh, H., *et al.* (2017). The devoted actor's will to fight and the spiritual dimension of human conflict. *Nature Human Behaviour*, *1*, 673–9.
- Hamid, N., Pretus, C., Vilarroya, O., *et al.* (2019) Neuroimaging "will to fight" for sacred values: an empirical case study with supporters of an Al Qaeda affiliate (under review). *Royal Society Open Science*. Submitted.
- Hare, T.A., Camerer, C.F., Rangel, A. (2009). Self-control in decision-making involves modulation of the vmPFC valuation system. *Science*, *324*, 646–8.
- Hutcherson, C.A., Plassmann, H., Gross, J.J., *et al.* (2012). Cognitive regulation during decision making shifts behavioral control between ventromedial and dorsolateral prefrontal value systems. *Journal of Neuroscience*, Vol. *32*, pp. 13543–54.
- Kleiner, M., Brainard, D., Pelli, D., *et al.* (2007). What's new in psychtoolbox-3. *Perception*, *36*, 1–16.
- Marek, S., Dosenbach, N.U.F. (2018). The frontoparietal network: function, electrophysiology, and importance of individual precision mapping. *Dialogues in Clinical Neuroscience*, *20*, 133–40. Neurofunctional Imaging Group, Université de Bordeaux. AAL/AAL2, [WWW]. Available: <http://www.gin.cnrs.fr/en/tools/aal-aal2/>. Last accessed: 2019 May 17.
- Peters, J., Büchel, C. (2010). Neural representations of subjective reward value. *Behavioural Brain Research*, *213*, 135–41.
- Plassmann, H., O'Doherty, J., Rangel, A. (2007). Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *27*, 9984–8.
- Pretus, C., Hamid, N., Ginges, J., *et al.* (2018). Neural and behavioral correlates of sacred values and vulnerability to violent extremism. *Frontiers in Psychology*, *9*, 2462.
- Rudolf, S., Hare, T.A. (2014). Interactions between dorsolateral and ventromedial prefrontal cortex underlie context-

- dependent stimulus valuation in goal-directed choice. *Journal of Neuroscience*, **34**, 15988–96.
- Sachdeva, S., Medin, D. (2009). Group identity salience in sacred value based cultural conflict: an examination of the Hindu-Muslim identities in the Kashmir and Babri mosque issues. In: *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, pp. 3111–4. Cognitive Science Society.
- Sheikh, H., Ginges, J., Atran, S. (2013). Sacred values in the Israeli-Palestinian conflict: resistance to social influence, temporal discounting, and exit strategies. *Annals of the New York Academy of Sciences*, **1299**, 11–24.
- Sheikh, H., Gómez, Á., Atran, S. (2016). Empirical evidence for the devoted actor model. *Current Anthropology*, **57**, S204–9.
- Shenhav, A., Greene, J.D. (2014). Integrative moral judgment: dissociating the roles of the amygdala and ventromedial prefrontal cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, **34**, 4741–9.
- Swann, W.B., Buhrmester, M.D., Gómez, A., et al. (2014). What makes a group worth dying for? Identity fusion fosters perception of familial ties, promoting self-sacrifice. *Journal of Personality and Social Psychology*, **106**, 912–26.
- Swann, W.B., Gómez, A., Seyle, D.C., et al. (2009). Identity fusion: the interplay of personal and social identities in extreme group behavior. *Journal of Personality and Social Psychology*, **96**, 995–1011.
- Tetlock, P.E. (2003). Thinking the unthinkable: sacred values and taboo cognitions. *Trends in Cognitive Sciences*, **7**, 320–4.
- Whitehouse, H., McQuinn, B., Buhrmester, M., et al. (2014). Brothers in arms: Libyan revolutionaries bond like family. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 17783–5.
- Whitfield-Gabrieli, S., Nieto-Castanon, A. (2012). Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain Connectivity*, **2**, 125–41.