Behavioral/Cognitive

Shared Neural Representations of Cognitive Conflict and Negative Affect in the Medial Frontal Cortex

Luc Vermeylen,[1] David Wisniewski,[1] Carlos González-García,[1] Vincent Hoofs,[1] Wim Notebaert,[1] and Senne Braem[1,2]

[1]Department of Experimental Psychology, Ghent University, Ghent, 9000, Belgium, and [2]Department of Experimental and Applied Psychology, Vrije Universiteit Brussel, Elsene, 1050, Belgium

Influential theories of Medial Frontal Cortex (MFC) function suggest that the MFC registers cognitive conflict as an aversive signal, but no study directly tested this idea. Instead, recent studies suggested that nonoverlapping regions in the MFC process conflict and affect. In this preregistered human fMRI study (male and female), we used MVPAs to identify which regions respond similarly to conflict and aversive signals. The results reveal that, of all conflict- and value-related regions, only the ventral pre-supplementary motor area (or dorsal anterior cingulate cortex) showed a shared neural pattern response to different conflict and affect tasks. These findings challenge recent conclusions that conflict and affect are processed independently, and provide support for integrative views of MFC function.

Key words: **cognitive conflict; cognitive control; emotion; fMRI; MVPA; negative affect**

Significance Statement

Multiple theories propose that the MFC, and the dorsal ACC in particular, integrates information related to suboptimal outcomes from different psychological domains (e.g., cognitive control and negative affect) with the aim of adaptively steering behavior. In contrast to recent studies in the field, we provide evidence for the idea that cognitive control and negative affect are integrated in the MFC by showing that a classification algorithm trained on discerning cognitive control (conflict vs no conflict) can predict affect (negative vs positive) in the voxel pattern response of the dorsal ACC/pre-SMA.

Introduction

The MFC, and dorsal ACC (dACC) in particular, have been implicated in various psychological processes, such as cognitive control, somatic pain, emotion regulation, reward learning, and decision-making (Shackman et al., 2011; Ebitz and Hayden, 2016; Heilbronner and Hayden, 2016). In the domain of cognitive control, the dACC is consistently activated by cognitive conflict, that is, the simultaneous activation of mutually incompatible stimulus, task, or response representations (Botvinick et al., 2001). However, other studies also showed dACC's involvement during the evaluation of negative outcomes, such as reductions in reward (Gehring and Willoughby, 2002), negative feedback (Nieuwenhuis et al., 2004), and pain (Rainville, 2002). Therefore, more integrative accounts of the dACC started to redescribe its role in conflict monitoring as detecting a domain-general aversive learning signal that can bias behavior away from the source of conflict (Botvinick, 2007; Shackman et al., 2011; Shenhav et al., 2013, 2016). In other words, the dACC is thought to register "cognitive" conflict as an aversive event. In accordance with this idea, recent studies have supported the presence of a behavioral bias to avoid conflict, and have also shown that humans automatically evaluate conflict as negative (Dreisbach and Fischer, 2015; Inzlicht et al., 2015; Dignath et al., 2020).

If conflict is registered as an aversive event in the dACC, one intriguing possibility is that conflict and negative affect are encoded similarly in dACC ("shared or overlapping representations," Kragel et al., 2018). This conjecture also relates to a broader debate on the functional organization of the MFC in regard to the psychological domains of cognitive control, negative affect, and pain (Shackman et al., 2011; Inzlicht et al., 2015; Lieberman and Eisenberger, 2015). Toward the end of the last century, cognitive neuroscientists argued for the functional segregation of affect (ventral part of the ACC) and cognitive control (dorsal part of the

Received July 7, 2020; revised Sep. 8, 2020; accepted Sep. 18, 2020.

Author contributions: L.V., W.N., and S.B. designed research; L.V. and V.H. performed research; L.V., D.W., C.G.-G., and S.B. analyzed data; L.V. wrote the first draft of the paper; L.V., D.W., C.G.-G., V.H., W.N., and S.B. edited the paper.

The authors declare no competing financial interests.

This work was supported by Fonds Voor Wetenschappelijk Onderzoek–Research Foundation Flanders G.0660.17N to W.N. and S.B. and 11H5619N to L.V. C.G.-G. was supported by Ghent University Special Research Fund BOF.GOA.2017.0002.03. S.B. was supported by an ERC Starting grant (European Union's Horizon 2020 research and innovation programme, Grant agreement 852570). D.W. was supported by Fonds Voor Wetenschappelijk Onderzoek FWO.KAN.2019.0023.01 and European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant agreement 665501. All procedures applied in the present experiment were performed with adequate understanding and written consent of the subjects and are in accordance with the Declaration of Helsinki. We thank Tobias Egner for valuable comments on a previous draft of the manuscript.

Correspondence should be addressed to Luc Vermeylen at Luc.Vermeylen@ugent.be.

https://doi.org/10.1523/JNEUROSCI.1744-20.2020

ACC) in the MFC (Devinsky et al., 1995; Bush et al., 2000). In contrast, an influential meta-analysis by Shackman et al. (2011) seemed to contradict this idea by showing substantial overlap in the dorsal part of the ACC for the three-way conjunction of negative affect, conflict, and pain. Similarly, one previous study tried to investigate the overlap in activation between cognitive conflict and negative affect by using a repetition suppression procedure, and found that dACC showed an attenuated response to negative affect following cognitive conflict (Braem et al., 2017).

In more recent years, however, other studies failed to provide evidence for such functional integration. For example, a number of recent studies and meta-analyses demonstrated that distinct, rather than overlapping, parts of the MFC are associated with cognitive conflict and pain processing (De La Vega et al., 2016; Jahn et al., 2016; Lieberman et al., 2016; Silvestrini et al., 2020). Similarly, one recent mega-analysis study reported a multivariate pattern analysis (MVPA) on full activation maps from 18 studies to assess the similarity of patterns evoked by different domains. This study also did not observe overlap between the activation patterns evoked by cognitive control, pain, and negative emotion in the MFC (Kragel et al., 2018). Together, current studies seem to be at odds with integrative views of MFC (Botvinick, 2007; Shackman et al., 2011; Shenhav et al., 2013, 2016; Calhoun and Hayden, 2015; Heilbronner and Hayden, 2016; Brown and Alexander, 2017), which aim to explain the various responses of dACC by one underlying process (e.g., avoidance learning, value estimation, surprise processing). However, these studies relied on group-averaged (often univariate) activation differences originating from distinct paradigms. For example, the mega-analysis by Kragel et al. (2018) used a context-insensitive "cognitive control" signal (i.e., across working memory, response inhibition, and conflict processing tasks) considering paradigms from multiple studies that differ in experimental control. Moreover, many of these previous studies made use of intense pain responses that could mask similarities with the arguably subtler affective evaluation of cognitive conflict.

Here, we took a different approach and developed a more targeted and well-controlled within-subjects test of shared neural representations of conflict and negative affect. Namely, by using multivariate cross-classification analyses, we assessed whether and where a classifier algorithm trained to discern conflict (incongruent vs congruent events) can successfully predict affect (negative vs positive events), and vice versa. Successful classification (i.e., classification above chance) would be indicative of a similarity between the neural pattern response, and thus a shared representational code between these two domains (Kaplan et al., 2015; Wisniewski, 2018).

Materials and Methods

Participants. The study was preregistered with the preregistration template from www.AsPredicted.org on the Open Science Framework (https://osf.io/p5frq/). As preregistered, 40 participants participated in our study. Two participants were excluded (one because of excessive head motion [>2.5 mm translation] and one aborted the scanning session). The average age of the remaining 38 participants (13 male, 25 female) was 23.71 years (SD = 3.53, minimum = 18, maximum = 33). Thirty-six participants were right-handed, one was left-handed, and one was ambidextrous (as assessed by the Edinburgh Handedness Inventory) (Oldfield, 1971). Every participant had normal or corrected-to-normal vision and reported no current or history of neurologic, psychiatric, or major medical disorder. Every participant gave their informed written consent before the experiment, and was paid 35 euros for participating afterward. The study was approved by the local ethics committee (University Hospital Ghent University).

Experimental design and paradigm. The experiment was implemented using Psychopy 2 version 1.85.2 (Peirce, 2007). On each trial, participants had to judge the color of a target stimulus in the center of the screen, using two MR-compatible response boxes (each box had two buttons) to indicate one of four possible response options (red, blue, green, and yellow). The key-to-color mapping was counterbalanced between participants. The exact features of the target stimulus varied blockwise, depending on one of four different task contexts. Specifically, participants either had to respond to the color of words ("color-word naming task") or respond to the color of circles ("color-circle naming task"), which both had a conflict and affective version (see Fig. 1*A*).

The conflict version of the color-word naming task was a Stroop task (Stroop, 1935), where the meaning of the words could either be congruent or incongruent with the actual color of the word. For example, participants could see the words "BLUE," "RED," "GREEN" or "YELLOW" (Dutch: "ROOD," "BLAUW," "GROEN" or "GEEL") presented in a blue, red, green, or yellow font. The conflict version of the color-circle naming task was essentially a color-based variant on the Eriksen flanker task (Eriksen and Eriksen, 1974), where the irrelevant feature consisted of a colored background square which could either be congruent or incongruent with the color of the circle. Here, participants could see blue, red, green, or yellow circles presented on a blue, red, green, or yellow background square. In both tasks, half of the trials were congruent (e.g., "RED" in a red font; a red circle presented on a red square background) while the other half of the trials were incongruent (e.g., "RED" in a blue font; a red circle on a blue square background).

The affect versions of the color-word naming and color-circle naming tasks made use of irrelevant affective words or pictures, respectively. In the color-word naming task, 16 positive and 16 negative words were presented (Moors et al., 2013) that were matched on arousal, power, age of acquisition, Dutch word frequency (Keuleers et al., 2010), word length, and grammatical category (Noun, Adjective, and Verbs). The affective picture distractors in the background of the color-circle naming task were retrieved from the OASIS database (Kurdi et al., 2017). Sixteen positive and 16 negative pictures were presented that were matched on semantic category (Animals, Objects, People, Scenery) and arousal. This resulted in a total of eight conditions: congruent, incongruent, positive, or negative trials, that either involved words or pictures/colored backgrounds.

Each trial started with a fixation sign (+) that was presented for 3-6.5 s (in steps of 0.5 s; mean = 3.5 s; drawn from an exponential distribution). Next, the target stimulus was presented for 1.5 s (fixed presentation time regardless of reaction time [RT]). In order to increase the saliency of the irrelevant dimension (conflict and affect), the onset of the word or picture preceded the presentation of the target feature by 200 ms during which the color of the target feature (word or circle) was white.

Participants performed five scanning runs; and during each run, the subjects performed each of the four task contexts in separate blocks. The order of the four blocked task contexts was fixed within participant but counterbalanced between participants. Each block hosted 32 trials (16 congruent/positive and 16 incongruent/negative), which were presented in a pseudo-random fashion with the following restriction: neither relevant nor irrelevant features of the target stimulus could be repeated from one trial to the next. This restriction was used to investigate confound-free congruency sequence effects (see Braem et al., 2019; Schmidt, 2019; but this was not the aim of the current study and will not be discussed further). In total, each participant made 640 trials (i.e., five runs of four blocks of 32 trials).

In each task context (block), we also included one catch trial (at random, but not in the first two or last two trials of each block). In these catch trials, the presentation of the task-irrelevant word, picture, or colored square would not be followed by the presentation of the target color, and remain on screen for 3 s. Participants were instructed that, during these catch trials, when no color information was present in the relevant dimension, their goal was to judge the irrelevant dimension depending on the cognitive domain. In the conflict domain, participants had to respond to the meaning of the word ("RED," "BLUE," "GREEN," or "YELLOW") or to the color of the background square (red, blue, green, or yellow) by using the respective key that would be used to judge the relevant dimension. In the affective domain, participants had to judge the affective word or background picture as either positive or

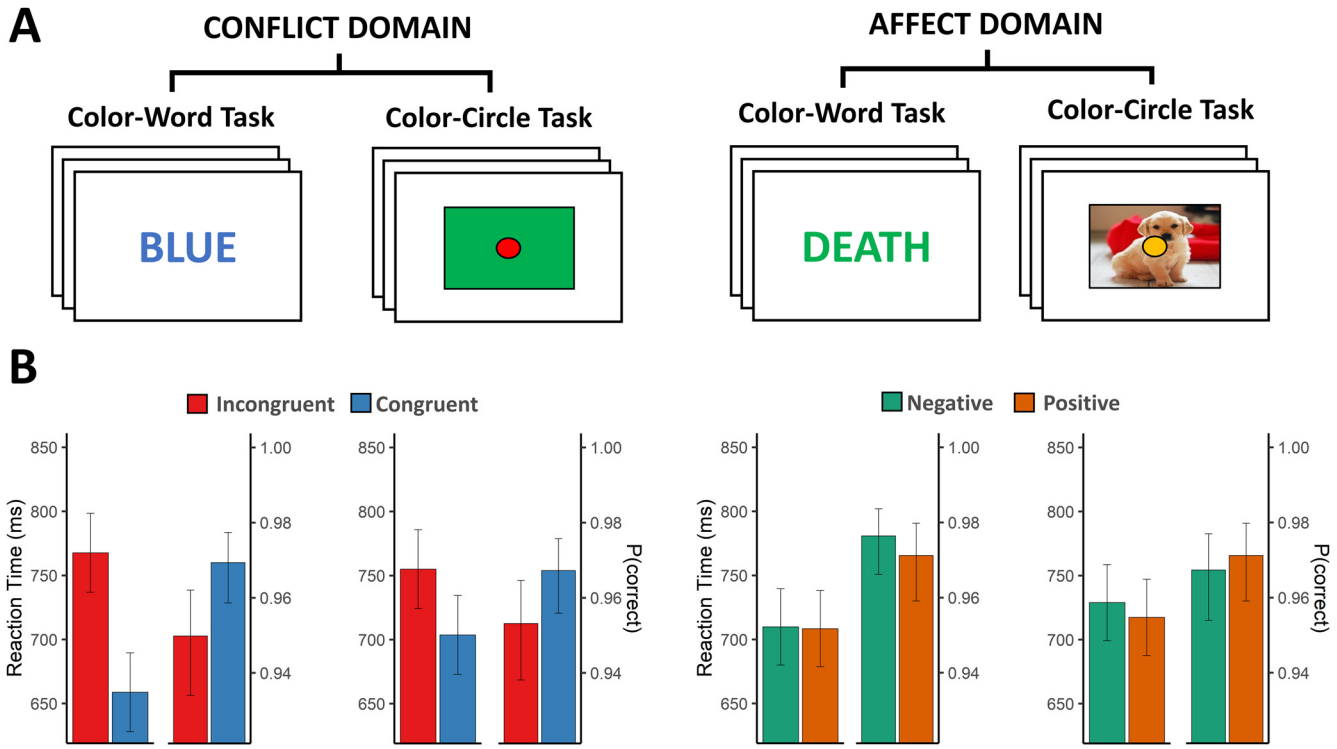

Figure 1. Task design and behavioral data. ***A***, Task design. Subjects judged either the color of words or the color of circles. In the conflict domain, the color either matched or mismatched with word meaning or background color, creating congruent or incongruent conditions, respectively. In the affective domain, positive or negative words and pictures were used to create the respective conditions. Regardless of the domain, subjects always had to judge the color of the word or the circle. These four task contexts were presented blockwise (with order counterbalanced) in each run. ***B***, RT and accuracy for the corresponding four task contexts. Error bars indicate 95% CI.

negative by pressing all keys once or twice (response mapping for positive and negative stimuli counterbalanced between participants). The purpose of these catch trials was to increase the saliency of the irrelevant dimension.

Before the scanning session, participants were welcomed and instructed to read the informed consent after which they started practicing the experimental paradigm. After the scanning sessions, participants performed an unannounced recognition memory test on old and new affective words and pictures. Here, participants had to indicate whether they had previously seen the word or picture in the experiment (old/new judgment). The new words were matched with the old words in terms of valence, arousal, power, age of acquisition, word length, frequency, and grammatical category. The new pictures were matched on valence, arousal and semantic category. In both a behavioral ($n = 20$) and fMRI pilot ($n = 20$), we already established that participants showed adequate performance on both the main task and the recognition memory task. Finally, participants completed four questionnaires (Need for Cognition, Behavioral Inhibition/Activation Scale, Positive and Negative Affect Schedule, Barret Impulsivity Scale) and were thanked for their participation. No significant correlations between these questionnaire scales and cross-classification accuracies were found.

Behavioral data analysis. Behavioral analyses were performed in R (RStudio version 1.1.463, www.rstudio.com). For the RT analyses, we removed incorrect, premature (<150 ms), and extreme responses (RTs outside 3 SDs from each condition mean for each participant). This resulted in an average of 94.42% of the trials left for the RT analyses (SD = 3.18, min = 84.22, max = 98.28). We conducted a repeated-measures ANOVA on the RT and accuracy measure with the within-subject factors Condition (conflict domain: congruent vs incongruent, affective domain: positive vs negative) and Task (color-word naming vs color-circle naming). We also assessed postscanning recognition memory of affective stimuli with a probit generalized linear mixed effects model on the probability to say that the stimulus was "old" with fixed effects for Experience (old vs new), Valence (positive vs negative), and Task Type (word vs picture) and crossed random effects for Participant and Item. We also preregistered some exclusion criteria based on behavioral performance. Participants with a mean RT outside 3 SDs from the sample mean or a hit rate <3 SDs or 60% (chance level = 25%) from the sample mean were excluded. Participants that performed poorly on the postscanning recognition memory test, that is, hit rate or false alarm rate outside 3 SDs of the sample mean were also excluded. In the end, no exclusions based on task performance had to be made. While performance on catch trials was not a preregistered exclusion criterion, we found that 2 participants responded on chance level in the catch trials of the affective domain (chance level = 50%, positive vs negative judgment). Excluding these participants did not change our conclusions.

fMRI data acquisition. fMRI data were collected using a 3T Magnetom Prisma MRI scanner system (Siemens Medical Systems), with a 64-channel radiofrequency head coil. A 3D high-resolution anatomic image of the whole brain was acquired for coregistration and normalization of the functional images, using a T1-weighted MPRAGE sequence (TR = 2250 ms, TE = 4.18 ms, TI = 900 ms, acquisition matrix = 256 × 256, FOV = 256 mm, flip angle = 9°, voxel size = 1 × 1 × 1 mm). Furthermore, a field map was acquired for each participant, to correct for magnetic field inhomogeneities (TR = 520 ms, TE1 = 4.92 ms, TE2 = 7.38 ms, image matrix = 70 × 70, FOV = 210 mm, flip angle = 60°, slice thickness = 3 mm, voxel size = 3 × 3 × 2.5 mm, distance factor = 0%, 50 slices). Whole-brain functional images were collected using a T2*-weighted EPI sequence (TR = 1730 ms, TE = 30 ms, image matrix = 84 × 84, FOV = 210 mm, flip angle = 66°, slice thickness = 2.5 mm, voxel size = 2.5 × 2.5 × 2.5 mm, distance factor = 0%, 50 slices) with slice acceleration factor 2 (simultaneous multislice acquisition). Slices were oriented along the AC-PC line for each subject.

fMRI data analysis. fMRI data analysis was performed using MATLAB (version R2016b 9.1.0, The MathWorks) and SPM12 (www.fil.ion.ucl.ac.uk/spm/software/spm12/). Raw data were imported according to BIDS standards (http://bids.neuroimaging.io/), and functional data were subsequently realigned, slice-time corrected, normalized (resampled voxel size 2 mm^3), and smoothed (FWHM of 8 mm). The preprocessed data were then entered into a first-level GLM analysis, and

Table 1. Whole-brain searchlight decoding results

Anatomical area	Hemisphere	MNI coordinates			Voxels	T_{max}
		x	*y*	*z*		
Within-affect within-task decoding						
Left middle occipital gyrus extending into right middle occipital gyrus	L	−38	−74	−4	6700	10.79
Within-affect cross-task decoding						
No cluster-corrected activations						
Within-affect overall decoding						
Middle occipital gyrus	L	−32	−78	−8	2730	10.81
Fusiform gyrus	R	36	−42	−20	108	7.91
Middle occipital gyrus/middle temporal gyrus	R	46	−68	−8	602	7.72
Insula	R	38	26	2	208	6.63
Within-conflict within-task decoding						
Inferior parietal lobule	L	−28	−54	42	467	7.36
Middle frontal gyrus	L	−44	6	32	779	6.88
Middle occipital gyrus	L	−26	−94	8	226	5.99
Pre-SMA	L	−12	16	54	75	6.13
Within-conflict cross-task decoding						
Superior parietal lobule	L	−20	−62	48	471	8.61
Inferior frontal gyrus	L	−34	22	24	42	5.96
Within-conflict overall decoding						
Superior parietal lobule	L	−26	−56	48	2096	8.97
Middle frontal gyrus extending into pre-SMA	L	−44	4	38	1389	8.46
Superior occipital gyrus	R	26	−70	38	240	6.43
Cross-domain cross-task decoding						
No cluster-corrected activations						
Cross-domain overall decoding						
No cluster-corrected activations						

subsequently into an MVPA (Cox and Savoy, 2003; Kriegeskorte et al., 2006; Haxby, 2012; Haynes, 2015). Results were analyzed using a mass-univariate approach. Although we preregistered that we would not normalize and smooth the data for our classification analyses, we found that temporal signal-to-noise ratio (tSNR) in the primary motor cortex significantly increased with wider smoothing (FWHM, $F_{(1,74)} = 1503$, $p < 0.001$). In addition, an independent classification analysis (classifying left vs right responses in primary motor cortex) showed that decoding accuracies were significantly higher with wider smoothing ($F_{(1,74)} = 12.54$, $p < 0.001$). Knowing that decoding information in the PFC is notoriously difficult as decoding accuracies are close to chance (relative to decoding in occipitotemporal cortex) (Bhandari et al., 2018), and the finding that smoothing can and does often improve SNR and decoding performance (Hendriks et al., 2017; Kamitani and Sawahata, 2010; Op de Beeck, 2010), we decided to optimize our MVPA analyses by decoding on normalized and smoothed data. For completeness, however, changing the smoothing parameters did not chance our main conclusion that dACC/pre-SMA shows above-chance level cross-domain cross-task classification (0 mm FWHM: Wilcoxon t-test [V] = 380, $p = 0.001$, Bayes factor [BF] = 9.81, 4 mm FWHM: V = 396, $p = 0.006$, BF = 2.19, 8 mm FWHM: V = 330, $p = 0.007$, BF = 8.43). As tSNR and thus reliability increased significantly with smoothing, we chose to report the most reliable data (FWHM 8 mm).

First-level GLM analyses consisted of five identically modeled sessions (i.e., the five runs). Each session consists of eight regressors of interest (for the eight conditions, see above), four block regressors (to account for the blocked presentation of each combination of word vs picture versions of the conflict vs affect tasks), two nuisance regressors (that model performance errors and catch trials), and six movement regressors. The regressors were convolved with the canonical HRF. The modeled duration of the regressors of interest (the eight conditions) and nuisance regressors (errors, catch trials) was zero, while the modeled duration of the block regressors was equal to the length of the blocks.

Next, the β images from the first-level GLM were submitted to leave-one-run-out decoding scheme with the Decoding Toolbox (Hebart et al., 2015) using a linear support vector classification algorithm (C = 1). We performed whole-brain searchlight decoding (sphere radius: 3 voxels; Table 1) as well as ROI decoding (see below for ROI methods). Cross-validation decoding was conducted within the affective (positive vs negative) and conflict (congruent vs incongruent) domain for each task separately ("within-domain within-task classification"). To assess the generalizability of the classifier within the domain, we also conducted cross-classification analyses where we trained the classifier on one task and tested its performance on the other task for each task type combination (from color-circle naming to color-word naming and vice versa) separately ("within-domain cross-task classification"). To investigate the generalizability of these classifiers across the domain (our main hypothesis), we trained the classifier in the conflict domain and tested its performance in the affective domain, and vice versa. We conducted these analyses cross-task-type combinations (i.e., from color-circle naming to color-word naming, or from color-word naming to color-circle naming) to further control for low-level task features, following the same reasoning as the within-domain cross-task classification analyses ("cross-domain cross-task classification"). For each of these three decoding analyses, we also ran ANOVAs to evaluate whether the result differed depending on the task (e.g., color-circle naming vs color-word naming) or task-to-task direction (i.e., from color-circle naming to color-word naming, or from color-word naming to color-circle naming). Finally, we also report an "overall decoding" analysis, where the classifier was trained across the two task types at once, thereby ignoring whether the event featured words or pictures/colored backgrounds.

Each classification analysis resulted in "accuracy minus chance" decoding maps for each subject. These maps were then entered into a group second-level GLM analysis in SPM12. Here, a one-sample *t* test determined which voxels show significant accuracy above-chance level.

Importantly, counterbalancing schemes that are not problematic for traditional analyses can introduce problems for the assessment of chance performance in MVPA (Görgen et al., 2018). This can be the case when the counterbalancing is "broken" by splitting the runs in training and test sets given that the counterbalancing and cross-validation are situated on the same level (e.g., the within-subject level of the runs) (Görgen et al., 2018). However, in our study, the order of the four tasks within a scanning run was counterbalanced between participants but fixed within participant (which is where the leave one run out cross-validation occurs). Therefore, counterbalancing conditions were always matched

Table 2. Whole-brain univariate results

Anatomical area	Hemisphere	MNI coordinates			Voxels	T_{max}
		x	*y*	*z*		
Conflict domain across both tasks (incongruent > congruent)						
Pre-SMA/dACC	L	−4	14	50	780	8.92
Inferior frontal gyrus	L	−38	22	24	1040	8.38
Superior parietal lobule/precuneus	L	−26	−58	48	524	8.08
Thalamus	L	−16	−2	16	114	6.51
Cerebellum	R	36	−54	−30	167	7.42
Conflict in the color-word naming task						
Middle/inferior frontal gyrus	L	−52	10	40	907	8.48
Superior parietal lobule/precuneus	L	−28	−58	50	614	8.17
Inferior parietal lobule	L	−38	−38	42	198	6.85
Pre-SMA/dACC	L	−6	14	50	190	7.51
Middle frontal gyrus/precentral gyrus	L	−28	2	68	164	6.50
Conflict in the color-circle naming task						
SMA	L	−6	4	60	2	5.57
Affective domain across tasks (negative > positive)						
Left fusiform gyrus/inferior temporal gyrus	L	−40	−44	−16	84	7.14
Middle occipital gyrus/middle temporal gyrus	L	−46	−76	2	453	8.31
Right inferior temporal gyrus	R	48	−72	−2	148	6.97
Affect in the color-word naming task						
No cluster-corrected activations						
Affect in the color-circle naming task						
Fusiform gyrus/inferior temporal gyrus	L	−42	−44	−16	105	8.23
Middle temporal gyrus/middle occipital gyrus	R	50	−74	2	244	7.76
Middle temporal gyrus/middle occipital gyrus	L	−48	−76	8	542	8.87

between training and testing runs. In addition, we modeled the blocked structure of the scanning runs by adding block regressors in the first-level GLMs (before entering the β images to the classification analyses). Therefore, it is unlikely that our counterbalancing could have introduced problems for the assessment of chance level. Nonetheless, for our main findings, this was verified by inspecting the null distributions obtained via permutation testing. Further, we did not remove univariate difference (more specifically, response differences of the same sign), as is sometimes done. Univariate differences are often a useful source of information and should not necessarily be viewed as irrelevant for multivariate analyses (Hebart and Baker, 2018). Further, we do not intend to make a special claim about fine-grained subtle patterns that go beyond univariate response differences.

Next to MVPA, we also conducted classic univariate analyses. Here, we constructed a set of contrasts subtracting (A) positive from negative conditions and (B) congruent from incongruent conditions for (1) each task separately as well as across both tasks. These contrast images were then entered into a second-level analysis in which a one-sample *t* test determined which voxels show significant activation for each contrast. We applied a statistical threshold of $p < 0.001$ (uncorrected) at the voxel level, and $p < 0.05$ (familywise error corrected) at the cluster level on all analyses (Table 2).

ROI analyses. As part of our preregistered main analysis plan, we conducted ROI decoding analyses. We set out to study the amygdala, ACC, dACC/pre-SMA, anterior insula (AI), parietal cingulate cortex (PCC), ventral striatum (VS), and the ventromedial PFC (vmPFC). Our preregistration noted that all ROIs would be obtained from the Harvard-Oxford cortical and subcortical structural atlases, thresholded at 25%. However, as the dACC ROI was not defined in the Harvard-Oxford atlas, we decided to retrieve this ROI from Neurosynth (Yarkoni et al., 2011) by entering "dacc" as search term (returning 162 studies reporting 4547 activations). Although this ROI was based on the "dacc" search term, the peak effect of studies reporting dACC activity actually lies more dorsally than the cingulate gyrus, overlapping with the pre-SMA (Lieberman and Eisenberger, 2015). Therefore, we refer to this ROI as the dACC/pre-SMA. Next, we built a 10 mm sphere around the peak activation point in this activation map (association map). Because the dACC ROI was spherical (in contrast to the other six atlas ROIs), we decided to retrieve all ROIs from NeuroSynth for comparability. However, because our preregistration mentioned that the ROIs would be obtained from the Harvard-Oxford cortical and subcortical structural atlases, we also reanalyzed our results by replacing all non-dACC/pre-SMA Neurosynth ROIs with their Harvard-Oxford analog. None of these Harvard-Oxford ROIs showed significant cross-domain cross-task decoding ($p > 0.066$, BF < 0.94), nor overall cross-domain decoding ($p > 0.319$, BF < 0.25).

In addition to the preregistered ROI analyses, which were based on anatomically determined ROIs, we also ran another set of ROI analyses with functionally informed ROIs. Namely, we created 10 mm sphere ROIs for all conflict-sensitive regions based on the most recent and inclusive meta-analysis we could find on cognitive conflict (Chen et al., 2018, their Table 2).

Each ROI decoding analysis returned one accuracy-minus-chance value per ROI and participant. We tested whether these values were significantly higher than zero (one-tailed) with the nonparametric Wilcoxon signed-rank test and a Bayesian *t* test (using the default priors from the BF package in R; Cauchy prior width: $r = 0.707$). We report the BF that quantifies the evidence for the alternative hypothesis (i.e., decoding accuracy is higher than zero). For our main hypothesis (cross-domain cross-task classification), we also report permutation tests. Here, the labels of the classification categories were shuffled (within-run only) and classification was performed for all possible permutations leading to first-level null distributions (i.e., for each subject) of accuracy-minus-chance values. A second-level null distribution was retrieved by calculating 10,000 group averages (where the value for each subject is a random draw from its respective null distribution). We retrieved a *p* value by dividing the number of samples above the empirical accuracy-minus-chance level by the total number of samples in the null distribution. The procedure of obtaining the second-level null distribution and *p* value was repeated 1000 times, after which we averaged the resulting *p* values. This led to stable *p* values that are not influenced by stochasticity in the procedure.

For our main set of ROI analyses (i.e., successful cross-domain cross-task classification), we report the uncorrected *p* values and BFs (always one-tailed — is accuracy-minus-chance larger than 0?), as well as the Bonferroni-corrected *p* values that control for the fact that we analyzed seven ROIs (i.e., multiplying the uncorrected *p* value by 7). Although our literature study above focuses on the MFC, we wanted to ensure in

our preregistration that other regions could be evaluated as well. For the other (secondary) set of ROI analyses, we present uncorrected p values, but these results can be more strictly evaluated in light of the adjusted Bonferroni α level for seven tests ($\alpha = 0.00714$).

Finally, we investigated whether the significant cross-task cross-domain classification accuracy correlated with the following behavioral indices: postscanning affective recognition memory (d'), congruency sequence effects in RT, and error rate and congruency sequence effects in RT and error rates (p values of reported correlations are Holm-corrected for five tests).

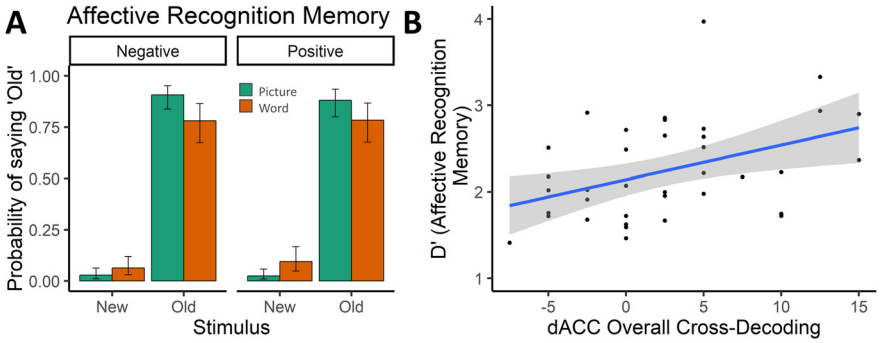

Figure 2. Postscanning incidental recognition memory. ***A***, After the experiment, subjects performed an unannounced recognition memory test on the affective stimuli. Subjects demonstrated high hit rates and low false alarm rates, which were further modulated by affective stimulus type (Pictures > Words). Data are mean ± 95% CI. ***B***, *Post hoc* correlation analyses also revealed a significant Spearman correlation between the sensitivity index (d') from the postscanning recognition memory for affective stimuli and overall cross-classification accuracy in the dACC/pre-SMA ROI ($Rs = 0.43$, $p = 0.006$, BF = 5.02), surviving correction for multiple comparisons (see Materials and Methods). Gray band represents 95% CI.

Results

We first analyzed the behavioral data to check whether the conflict task showed the typical congruency effects, and subjects processed the affective stimuli in the affective task. The behavioral data from the conflict tasks (Fig. 1*B*, left panels) showed typical congruency effects in RT ($F_{(1,37)}$ = 149.81, $p < 0.001$, BF10 > 100) and accuracy ($F_{(1,37)}$ = 11.72, $p = 0.002$, BF = 52), which were larger in the color-word task (mean = 109 ms, $T_{(68)}$ = 13.39, $p < 0.001$) relative to the color-circle task (mean = 51 ms, $T_{(68)}$ = 6.31, $p < 0.001$) for the RT measure (interaction between task and congruency effect: $F_{(1,37)}$ = 35.55, $p < 0.001$, BF > 100). In the affective tasks, we found slower RTs with negative relative to positive background stimuli ($F_{(1,37)}$ = 5.74, $p = 0.025$, BF = 0.38), but this was only the case for the color-circle task (mean = 12 ms, $T_{(73)}$ = 3.12, $p = 0.003$) and not the color-word task (mean = 1 ms, $T_{(73)}$ = 0.37, $p = 0.710$; interaction between task and valence effect: $F_{(1,37)}$ = 4.27, $p = 0.046$, BF = 0.37) (Fig. 1*B*, right panels). However, the BF ($p = 0.025$ vs BF = 0.38) leads us to interpret this effect as marginal. No effects on accuracy were found in the affective tasks (F values < 3.12, p values > 0.085). In the conflict tasks, catch trials were used to draw attention to the conflicting nature of the stimuli. Here, subjects had to judge the irrelevant, rather than the relevant, dimension (see Materials and Methods). We observed above-chance catch trial performance (chance level = 25%), which did not differ between the two conflict tasks ($\chi^2_{(1)}$ = 0.10, $p = 0.755$, BF = 0.12; color-circle task, 90.7%; color-word task, 89.9%). In the affective tasks, catch trials (where subjects had to make a valence judgment instead of a color judgment) and a postexperiment incidental memory test were used to inform processing of the (task-irrelevant) affective stimuli. We observed above-chance catch trial performance (chance level = 50%), which did not differ between the two affect tasks ($\chi^2_{(1)}$ = 0.19, $p = 0.664$, BF = 0.12; color-circle task, 86.4%; color-word task, 87.8%). In the incidental recognition memory test, subjects demonstrated high hit rates and low false alarm rates ($\chi^2 = 442$, $p < 0.001$), which were further modulated by affective stimulus type (Pictures > Words; $\chi^2 = 16.49$, $p < 0.001$) (Fig. 2*A*).

In a first set of MVPAs, we trained and tested a classifier within-task (within the Stroop or flanker task; Fig. 3*A*, left panels — which regions respond to conflict within tasks?), in each of our preregistered ROIs (for analysis details, see Materials and Methods). Within-task ROI analyses in the conflict domain (congruent vs incongruent) revealed evidence for above chance-level decoding in the ACC (Wilcoxon V = 388, uncorrected $p = 0.003$, BF = 15.61, Bonferroni-adjusted α = 0.00714), dACC/pre-SMA (V = 327, $p = 0.009$, BF = 8.48), and PCC (V = 346, $p = 0.009$, BF = 4.57) but not in in any of the other regions (all $p > 0.145$, BF < 0.61) (Fig. 3*A*, right). The decoding accuracies did not differ by task (ACC: $F_{(1,37)}$ = 0.01, $p = 0.915$, BF = 0.18, dACC/pre-SMA: $F_{(1,37)}$ = 0.72, $p = 0.400$, BF = 0.24, PCC: $F_{(1,37)}$=0.51, $p = 0.476$, BF = 0.22).

A second set of MVPAs evaluated whether we could also cross-classify conflict signals cross-task (train and test on different tasks; which regions respond similarly to conflict independent of specific task features? Fig. 3*B*, left panels). To our knowledge, no study has shown a voxel pattern response to conflict that is independent of conflict task in the ACC or pre-SMA (e.g., Jiang and Egner, 2014). Here, our within-conflict cross-task ROI analyses revealed above-chance level conflict decoding across tasks in the dACC/pre-SMA (V = 283, $p = 0.012$, BF = 5.57) and PCC (V = 328, p = 0.023, BF = 3.67) (Fig. 3*B*, right). Decoding accuracy did not differ between cross-task combination for the dACC/pre-SMA ($F_{(1,37)}$ = 0.89, $p = 0.352$, BF = 0.26) but was larger in the Stroop to flanker decoding relative to the flanker to Stroop decoding for the PCC ($F_{(1,37)}$ = 4.35, $p = 0.043$, BF = 1.21). These results were also replicated in an overall decoding approach where the classifier was trained and tested in the whole domain regardless of task. Within the affective domain (positive vs negative), we also performed similar within- and cross-task decoding analyses. However, while some of these analyses showed evidence for affect information in the AI (the within-affect overall decoding: V = 407, $p = 0.011$, BF = 5.00; all other, $p > 0.086$, BF < 0.92) and vmPFC (the within-affect within-task decoding: V = 376, $p = 0.001$, BF = 39.10; all others, $p > 0.073$, BF < 1.26), they did not show evidence for decoding in the ACC, dACC/pre-SMA, or PCC (but see follow-up analyses below).

Finally, we investigated our main hypothesis by training a classifier on discerning conflict (incongruent vs congruent) and testing its performance on discerning affect (negative vs positive), and vice versa. For this main set of analyses, we report the uncorrected and Bonferoni-corrected p values, and focused on the cross-domain cross-task decoding (train and test in different domains on different tasks) as this analysis controls for low-level features shared between the two tasks (Fig. 3*C*, left). The cross-domain cross-task ROI decoding revealed evidence for cross-classification in the dACC/pre-SMA (V = 330, $p = 0.007$, BF = 8.43; Fig. 3*C*, right) and vmPFC (V = 277, $p = 0.045$, BF = 1.56), which did not differ by cross-task combination (dACC/pre-

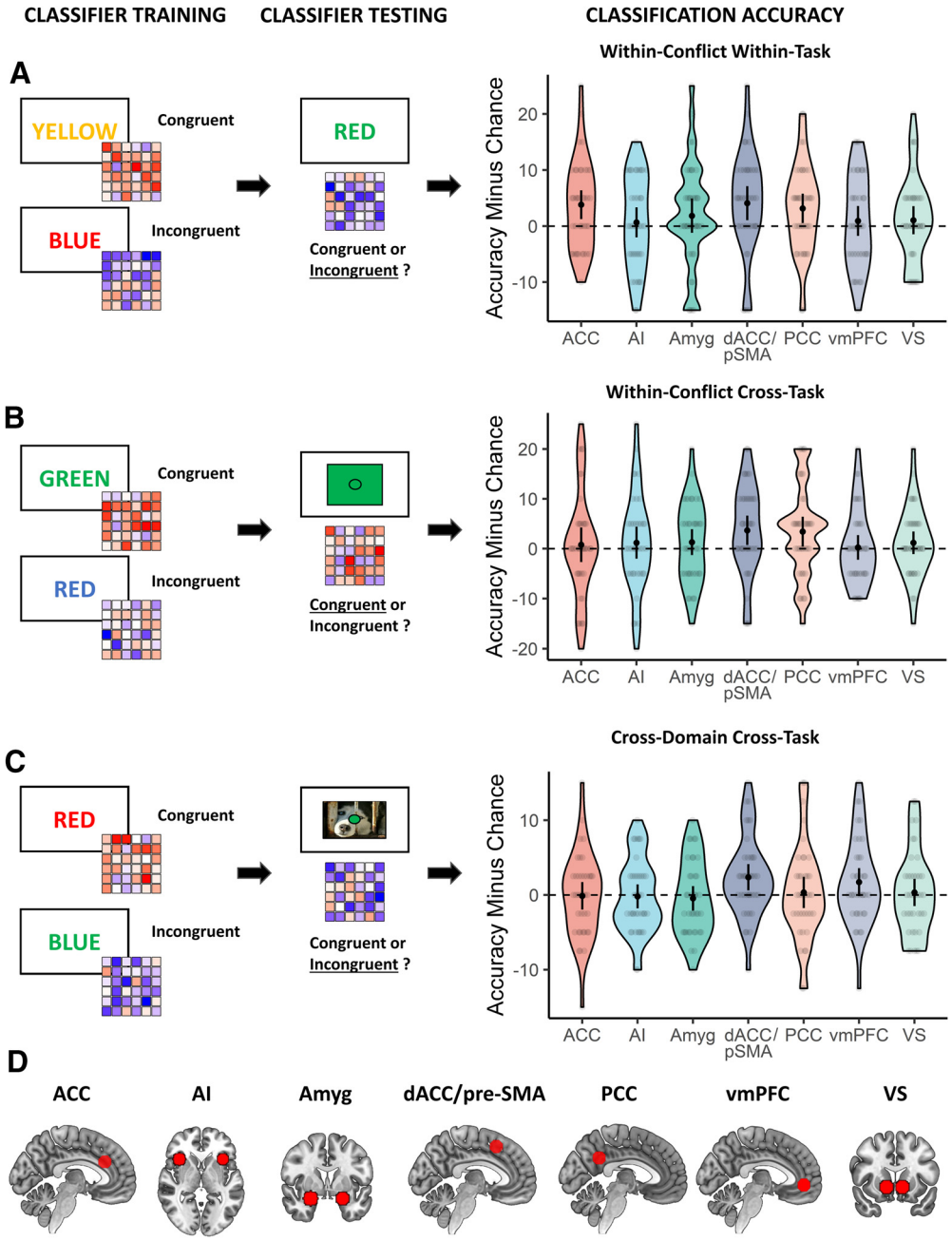

Figure 3. Main results. ***A***, Training and testing the classifier within the conflict domain, within the same task. ***B***, Training the classifier on one conflict task and testing its performance on another conflict task. ***C***, Training the classifier to discern conflict and testing its performance on classifying affect in another task (and vice versa). ***D***, ROIs: ACC, AI, amygdala (Amyg), dACC/pre-SMA, posterior cingulate cortex (PCC), vmPFC, and VS. Black dots represents mean. Error bars indicate ±95 CI. Transparent dots represent individual data points. The shape of the violin shows the distribution of the data.

SMA: $F_{(1,37)} = 0.36$, $p = 0.551$, BF = 0.20, vmPFC: $F_{(1,37)} = 0.02$, $p = 0.880$, BF = 0.18). Bonferroni-corrected p values controlling for the use of seven ROIs showed that this result was still significant for the dACC/pre-SMA ($p_{corr} = 0.049$), but not for the vmPFC ($p_{corr} = 0.316$). None of the other ROIs reached significance (all p values > 0.444, BF values < 0.24). Using the overall decoding approach (training on both tasks in one domain and testing on both tasks in the other domain), we were only able to replicate successful cross-domain decoding in dACC/pre-SMA ($V = 449$, $p = 0.021$, BF = 4.65; all other, $p > 0.319$, BF < 0.22). *Post hoc* correlation analyses also revealed a significant Spearman correlation between the sensitivity index (d') from the postscanning recognition memory for affective stimuli (independent of valence) and overall cross-domain classification accuracy in the dACC/pre-SMA ROI (R values = 0.43, $p = 0.006$, BF = 5.02, Fig. 2*B*), surviving correction for multiple comparisons (see Materials and Methods), suggesting that subjects with better recognition memory were more attentive to the affective stimuli which might have made cross-classification more successful.

In what follows, we report three additional sets of control analyses. A first set was designed to evaluate the extent to which our analysis choices influenced our findings (Botvinik-Nezer et al., 2020). A second set was geared toward ruling out alternative hypotheses. Finally, a third set further zoomed in on the above observation that we could not decode affect within or across

tasks in the dACC/pre-SMA, but did observe cross-domain decoding.

First, a number of control analyses using different analysis choices further confirmed our main finding. We replicated our results using permutation testing (see Materials and Methods), as classification rates might not match theoretical chance levels (Jamalabadi et al., 2016). This approach revealed similar results (dACC/pre-SMA: $p = 0.006$, $p_{corr} = 0.041$, vmPFC: $p = 0.034$, $p_{corr} = 0.238$) and further demonstrated that chance level was not inflated as the mean of the group null distribution of the accuracy minus chance measure did not differ from zero for each of the two ROIs (dACC/pre-SMA: $p = 0.753$; vmPFC: $p = 0.267$). Also, as mentioned in Materials and Methods, we replicated our main finding using different smoothing parameters, or when using Harvard-Oxford atlas ROIs instead of ROIs retrieved from NeuroSynth. Moreover, when using a set of functionally (rather than anatomically) defined conflict-sensitive ROIs based on a recent meta-analysis (from Chen et al., 2018; for MNI coordinates, see their Table 2), we again observed evidence for cross-domain cross-task classification in the dACC/pre-SMA (V = 450, $p = 0.013$, BF = 3.75) but not for other conflict-sensitive ROIs (left middle occipital gyrus, right AI, left AI, left inferior frontal gyrus, left inferior parietal lobule, right inferior parietal lobule, left middle frontal gyrus), except for the left AI (V = 425, $p = 0.005$, BF = 8.61). The result again replicated when using the overall decoding approach in the dACC/pre-SMA (V = 449, $p = 0.001$, BF = 41.06), but not in the left AI (V = 335, $p = 0.260$, BF = 0.34).

Second, to further study whether the main effect was specific to the congruency identity and valence of our experimental conditions, we also tested whether task difficulty differences or differences in arousal could not explain our results. Notably, there was a small performance difference in task performance on negative versus positive affect trials. Therefore, it is possible that our cross-decoding result reflects the decoding of RT, rather than a shared conflict and affect signal. However, the RT effects for the affective domain were marginal (BF < 1) and only present for one of the two affective tasks (also, there were no effects on accuracy). Importantly, as reported above, the cross-domain decoding accuracy did not depend on which task was used for testing. Moreover, cross-domain decoding accuracy was not higher when splitting the sample for the subsample that did show the RT effect (RT effect > 0, $N = 28$, mean = 1.96, $p = 0.050$, BF = 1.66) relative to a subsample that did not show the RT effect (RT effect < 0, $N = 10$, mean = 3.50, $p = 0.017$, BF = 4.58) ($F_{(1,36)} = 0.59$, $p = 0.444$), nor did the RT effect correlate with cross-domain decoding accuracy ($R = -0.23$, $p = 0.162$). Nevertheless, to control for any potential confounding effects of RT, we also ran another control analysis regressing out RT-related effects from the neural data (via parametric modulation). This analysis led to the same conclusions as above (cross-domain cross-task dACC/pre-SMA: V = 380, $p = 0.036$, BF = 1.91, overall cross-domain dACC/pre-SMA: V = 348, $p = 0.025$, BF = 2.61). This does make the effect slightly smaller, which should not be surprising as this procedure also removes variance of interest (as the effectiveness of the congruency manipulation is often defined by RT differences).

Next, we also found that the cross-domain decoding was unlikely to be because of arousal. First, we were only able to decode arousal (high vs low arousal; based on a median-split of arousal ratings, orthogonal to valence) within-tasks in the ACC (V = 260, $p = 0.039$, BF = 1.43) and vmPFC (V = 240, $p = 0.048$, BF = 1.38), but not in the dACC/pre-SMA (V = 230, $p = 0.082$, BF = 0.83) or other ROIs ($p > 0.217$, BF < 0.39). Second, training a classifier on conflict and testing on arousal (and vice versa) did not show any evidence for cross-decoding in the dACC/pre-SMA (V = 294, $p = 0.404$, BF = 0.26), nor any of the other ROIs (p values > 0.139, BF < 0.51), except for the vmPFC (V = 401, $p = 0.015$, BF = 3.25).

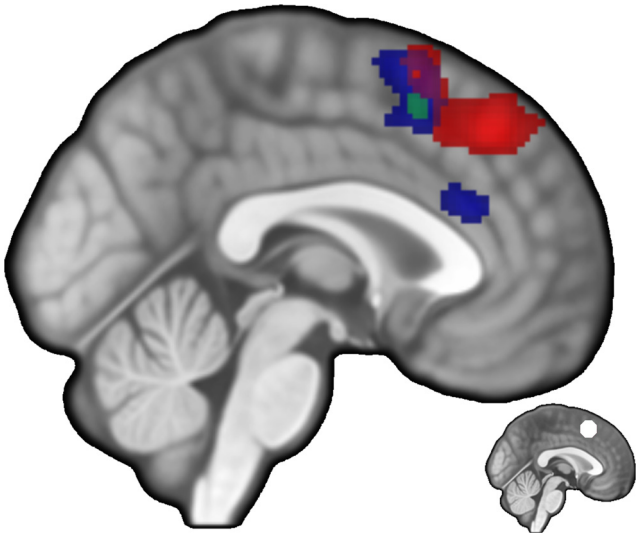

Figure 4. Cluster-corrected searchlight decoding maps restricted to the MFC after a small-volume correction (with the same mask as used by Kragel et al., 2018) (uncorrected $p < 0.005$). Within-affect within-task decoding revealed a large cluster in the dmPFC ($N = 865$; red). For the within-conflict within-task searchlight decoding, we show a similar large cluster ($N = 3156$; blue). Finally, we found cross-domain cross-task decoding in the MFC, but this cluster ($N = 87$) did not survive a cluster correction (green). Each of these clusters partially overlaps with each other as well as with our main dACC/pre-SMA ROI (bottom right).

Third, and finally, we wanted to follow-up on the unexpected result that we did not observe within and cross-task decoding of affect in the dACC/pre-SMA. One potential explanation is that the SNR was lower for our affect differences than the congruency differences. In line with this, while the univariate affect contrast (negative > positive) showed no cluster-corrected activation in the MFC (Table 2), a large cluster of dmPFC activity does become visible, when we lower the threshold ($p < 0.005$, uncorrected). Similarly, when restricting the whole-brain decoding analysis (Table 1) to the MFC using the same mask as Kragel et al. (2018), within-affect decoding did reveal a large cluster in the MFC, overlapping with our main dACC/pre-SMA ROI, as well as the within-conflict decoding and cross-domain decoding results (Fig. 4). A second reason for observing weaker affect decoding signals could be because the affect signals weakened or habituated throughout the experiment, as the same affective words and pictures were repeated in each run. Therefore, we also investigated whether affect might have been easier to decode in the beginning of the experiment, by studying cross-task affect for the first half (runs 1 and 2) and second half (runs 4 and 5), separately. Indeed, there was significant cross-task affect decoding in main dACC/pre-SMA ROI in the first half of the experiment (runs 1 and 2, V = 237, $p = 0.019$, BF = 2.38), but not in the second half of the experiment (runs 4 and 5, V = 138, $p = 0.646$, BF = 0.10). Finally, it is worth noting that our ROIs were primarily selected to detect conflict-based differences (see also our functional ROIs on conflict decoding), which again potentially lowered our chances to observe affect decoding. When using ROIs based on an affect processing meta-analysis (from Lindquist et al., 2016, their Table 2; compare with our conflict-sensitive ROIs), a pre-SMA ROI close to our original dACC/pre-SMA ROI (but a bit

more dorsal) did show significant affect ($V = 314$, $p = 0.001$, $BF = 39.70$), conflict ($V = 308$, $p = 0.002$, $BF = 27.19$), and cross-domain decoding ($V = 375$, $p = 0.042$, $BF = 1.52$). Together, these results do suggest that there was affect decoding in, or at least around, our main dACC/pre-SMA ROI.

Discussion

Together, our results reveal that the dACC/pre-SMA shows a similar voxel pattern response to conflict and negative affect. Moreover, to the best of our knowledge, our study is also the first to show decoding of conflict across conflict tasks in the dACC/pre-SMA, suggesting a shared component in the detection of conflict across the Stroop and flanker task (Jiang and Egner, 2014).

These findings fit with integrative accounts of MFC function, which propose that multiple domains (e.g., cognitive conflict, negative emotion, and pain) activate a single underlying process in the MFC (e.g., adaptive control, avoidance learning, cost-benefit value estimation), and thus predict similar activation patterns across domains (Botvinick, 2007; Shackman et al., 2011; Shenhav et al., 2013, 2016; Calhoun and Hayden, 2015; Heilbronner and Hayden, 2016; Brown and Alexander, 2017). Since our study focused on the domains of cognitive conflict and negative affect, our finding is especially relevant for theories that have tried to explain dACC's sensitivity to conflict and negative outcomes by a single unifying process (Botvinick, 2007; Shenhav, 2013, 2016). These theories have proposed the idea that conflict is in itself an aversive outcome (Botvinick, 2007; Shackman et al., 2011; Shenhav et al., 2013), and the main mechanism of the dACC is to bias behavior away from any costly, demanding, or suboptimal outcome ("domain-general avoidance learning,") (Botvinick, 2007) or to allocate control based on the expected benefits discounted by the expected costs of controlling behavior ("cost-benefit value estimation") (Shenhav et al., 2013).

Importantly, our study was not set up to disentangle the multiple candidate processes as to why conflict and negative affect would elicit a shared neural pattern response in the MFC; so while our main finding (i.e., similar voxel pattern response to conflict and negative affect) follows naturally from above-mentioned models that predict similar patterns of brain activity for different domains, it does not favor a specific mechanism (e.g., domain-general avoidance learning). Still, our finding does lend credence to the idea that that cognitive control can be understood as an emotional process (Inzlicht et al., 2015), and that conflict can be registered as an aversive event (i.e., Botvinick, 2007; Dreisbach and Fischer, 2015; Dignath et al., 2020), at least in the dACC/pre-SMA. Alternatively, it could be interpreted as evidence for the demanding or controlling nature of negative affect rather than the affective nature of conflict. It has been show that negative affect evokes cognitive control through the need for distraction suppression (e.g., Okon-Singer et al., 2013). Therefore, the abstract similarity between conflict and affect can be framed from either the perspective of shared valence or a shared signaling of control demands. Our data do not permit conclusive evidence for either interpretation. Nevertheless, additional analyses further show that our effect is unlikely to be driven by general differences in RT or task difficulty, or differences in arousal. Future research will be necessary to disentangle the exact underlying nature of the observed similarity.

Other recent studies failed to find similarities, and suggest that theories of MFC function should not look for a unitary neural implementation, but rather focus on a unified computational mechanism (Kragel et al., 2018). Our study does not argue against this idea, as shared neural "representations" in fMRI could still be driven by different local neighboring neural populations, a hypothesis that is difficult to address with fMRI. Still, it could be possible that the observed pattern similarity is because our dACC/pre-SMA ROI is situated at the boundary of two functionally distinct regions (coding for conflict and affect, respectively) or networks (frontoparietal/control vs cingulo-opercular/salience) and thus reflects the mixed selectivity of these voxels. The (uncorrected) searchlight decodings restricted to the MFC presented in Figure 4 can be informative here. These decoding maps show that conflict and affect are mostly represented by distinct regions within the MFC, with affect being represented adjacent to, but more anterior than conflict (also note that conflict was represented along the cingulate gyrus in the ACC, which we preregistered as one of the main candidate regions for a shared representation). Closer inspection of the shared pattern response to conflict and affect shows that it is near, but not on the boundary of regions coding for conflict and affect, and is mostly encapsulated within the conflict region. Therefore, while it is possible that the shared pattern response is because of mixed selectivity of voxels in our main ROI, we believe our findings do license further consideration of a unitary neural implementation.

Still, one might wonder why we did, and others did not, find shared neural representations of conflict and affect. First, we used a multivariate approach rather than a univariate approach which is more sensitive and more informative to assess similarity in brain activation (without necessarily being orthogonal to a univariate approach) (Davis et al., 2014). Second, our design was specifically set up to evaluate similarity between two specific domains (conflict and negative affect) rather than very general domains (cognitive control, negative emotion, pain). Third, related to this, we used a within-subjects design (rather than the between-subject nature of meta- or mega-analyses), which allowed for more sensitive analyses and to make the different task contexts as similar as possible (high level of experimental control). Interestingly, there is one recent study that also used a within-subject MVPA approach but focused on the domains of conflict and pain, rather than conflict and negative affect (Silvestrini et al., 2020). In this study, the authors assessed the degree to which a literature-based pattern predictive for conflict was reactivated during pain following a low- or high-demand Stroop task. They found that a pattern predictive of Stroop conflict (according to Neurosynth) was reactivated during pain in the anterior mid-cingulate cortex following the high-demand Stroop task, proposedly providing neural support for the observation that exerting control leads to a subsequent increase in pain response. Therefore, although this study did not test the presence of a shared pattern response, it does suggest that conflict processing can interact with pain processing through distinct activation patterns within the anterior mid-cingulate cortex. However, it is unclear whether this aftereffect of conflict processing was specific to their pain condition, or would also have been observed in a no-pain control condition.

One unexpected finding relates to the fact that we did not observe a similar (significant) above-chance decoding of affect in the dACC/pre-SMA, but did observe cross-domain decoding. Although this finding was surprising to us at first, it most likely suggests differences in SNR between the two domains and does not invalidate the cross-domain decoding result (for a review and discussion of similar findings, see van den Hurk and Op de Beeck, 2019). This idea was also further corroborated by follow-up analyses, which showed affect decoding in the dACC/pre-

SMA when using other analysis techniques, or focusing on the first half of the experiment only. Moreover, a lower SNR in the affect domain can also be explained by the fact that affect was not relevant for the main task (which allowed us to keep the affective tasks as similar as possible to the conflict tasks).

Finally, while our analyses were based on theories and a NeuroSynth ROI of the dACC, investigating the (uncorrected) whole-brain searchlight decoding maps did reveal that the conflict, affect, and cross-domain decoding all trigger a dorsomedial frontal area, which might be more accurately referred to as pre-SMA rather than dACC. This area roughly corresponds to our main dACC/pre-SMA ROI, which was retrieved by searching "dACC" in NeuroSynth (Yarkoni et al., 2011); therefore, we decided to refer to this ROI as dACC/pre-SMA.

In conclusion, by using a well-controlled within-subject design and multivariate analysis techniques, we show that conflict and negative affect evoke a similar voxel pattern response in the MFC. Our finding brings important nuance to other recent studies suggesting different neural activations or representations, and helps to further constrain theories of integrative MFC function.

References

Bhandari A, Gagne C, Badre D (2018) Just above chance: is it harder to decode information from prefrontal cortex hemodynamic activity patterns? J Cogn Neurosci 30:1473–1498.

Botvinick MM (2007) Conflict monitoring and decision making: reconciling two perspectives on anterior cingulate function. Cogn Affect Behav Neurosci 7:356–366.

Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD (2001) Conflict monitoring and cognitive control. Psychol Rev 108:624–652.

Botvinik-Nezer R, Holzmeister F, Camerer CF, Dreber A, Huber J, Johannesson M, Kirchler M, Iwanir R, Mumford JA, Adcock RA, Avesani P, Baczkowski BM, Bajracharya A, Bakst L, Ball S, Barilari M, Bault N, Beaton D, Beitner J, Benoit RG, et al. (2020) Variability in the analysis of a single neuroimaging dataset by many teams. Nature 582:84–88.

Braem S, King JA, Korb FM, Krebs RM, Notebaert W, Egner T (2017) The role of anterior cingulate cortex in the affective evaluation of conflict. J Cogn Neurosci 29:137–149.

Braem S, Bugg JM, Schmidt JR, Crump MJ, Weissman DH, Notebaert W, Egner T (2019) Measuring adaptive control in conflict tasks. Trends Cogn Sci 23:769–783.

Brown JW, Alexander WH (2017) Foraging value, risk avoidance, and multiple control signals: how the anterior cingulate cortex controls value-based decision-making. J Cogn Neurosci 29:1656–1673.

Bush G, Luu P, Posner MI (2000) Cognitive and emotional influences in anterior cingulate cortex. Trends Cogn Sci 4:215–222.

Calhoun AJ, Hayden BY (2015) The foraging brain. Curr Opin Behav Sci 5:24–31.

Chen T, Becker B, Camilleri J, Wang L, Yu S, Eickhoff SB, Feng C (2018) A domain-general brain network underlying emotional and cognitive interference processing: evidence from coordinate-based and functional connectivity meta-analyses. Brain Struct Funct 223:3813–3840.

Cox DD, Savoy RL (2003) Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. Neuroimage 19:261–270.

Davis T, LaRocque KF, Mumford JA, Norman KA, Wagner AD, Poldrack RA (2014) What do differences between multi-voxel and univariate analysis mean? How subject-, voxel-, and trial-level variance impact fMRI analysis. Neuroimage 97:271–283.

De La Vega A, Chang LJ, Banich MT, Wager TD, Yarkoni T (2016) Large-scale meta-analysis of human medial frontal cortex reveals tripartite functional organization. J Neurosci 36:6553–6562.

Devinsky O, Morrell MJ, Vogt BA (1995) Contributions of anterior cingulate cortex to behaviour. Brain 118:279–306.

Dignath D, Eder AB, Steinhauser M, Kiesel A (2020) Conflict monitoring and the affective-signaling hypothesis: an integrative review. Psychon Bull Rev 27:193–216.

Dreisbach G, Fischer R (2015) Conflicts as aversive signals for control adaptation. Curr Dir Psychol Sci 24:255–260.

Ebitz RB, Hayden BY (2016) Dorsal anterior cingulate: a Rorschach test for cognitive neuroscience. Nat Neurosci 19:1278–1279.

Eriksen BA, Eriksen CW (1974) Effects of noise letters upon the identification of a target letter in a nonsearch task. Percept Psychophys 16:143–149.

Gehring WJ, Willoughby AR (2002) The medial frontal cortex and the rapid processing of monetary gains and losses. Science 295:2279–2282.

Görgen K, Hebart MN, Allefeld C, Haynes JD (2018) The same analysis approach: Practical protection against the pitfalls of novel neuroimaging analysis methods. Neuroimage 180:19–30.

Haxby JV (2012) Multivariate pattern analysis of fMRI: the early beginnings. Neuroimage 62:852–855.

Haynes JD (2015) A primer on pattern-based approaches to fMRI: principles, pitfalls, and perspectives. Neuron 87:257–270.

Hebart MN, Baker CI (2018) Deconstructing multivariate decoding for the study of brain function. Neuroimage 180:4–18.

Hebart MN, Görgen K, Haynes JD (2015) The Decoding Toolbox (TDT): a versatile software package for multivariate analyses of functional imaging data. Front Neuroinform 8:88.

Heilbronner SR, Hayden BY (2016) Dorsal anterior cingulate cortex: a bottom-up view. Annu Rev Neurosci 39:149–170.

Hendriks MH, Daniels N, Pegado F, Op de Beeck HP (2017) The effect of spatial smoothing on representational similarity in a simple motor paradigm. Front Neurol 8:222.

Inzlicht M, Bartholow BD, Hirsh JB (2015) Emotional foundations of cognitive control. Trends Cogn Sci 19:126–132.

Jahn A, Nee DE, Alexander WH, Brown JW (2016) Distinct regions within medial prefrontal cortex process pain and cognition. J Neurosci 36:12385–12392.

Jamalabadi H, Alizadeh S, Schönauer M, Leibold C, Gais S (2016) Classification based hypothesis testing in neuroscience: below-chance level classification rates and overlooked statistical properties of linear parametric classifiers. Hum Brain Mapp 37:1842–1855.

Jiang J, Egner T (2014) Using neural pattern classifiers to quantify the modularity of conflict–control mechanisms in the human brain. Cereb Cortex 24:1793–1805.

Kamitani Y, Sawahata Y (2010) Spatial smoothing hurts localization but not information: pitfalls for brain mappers. Neuroimage 49:1949–1952.

Kaplan JT, Man K, Greening SG (2015) Multivariate cross-classification: applying machine learning techniques to characterize abstraction in neural representations. Front Hum Neurosci 9:151.

Keuleers E, Brysbaert M, New B (2010) SUBTLEX-NL: a new measure for Dutch word frequency based on film subtitles. Behav Res Methods 42:643–650.

Kragel PA, Kano M, Van Oudenhove L, Ly HG, Dupont P, Rubio A, Delon-Martin C, Bonaz BL, Manuck SB, Gianaros PJ, Ceko M, Reynolds Losin EA, Woo CW, Nichols TE, Wager TD (2018) Generalizable representations of pain, cognitive control, and negative emotion in medial frontal cortex. Nat Neurosci 21:283–289.

Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. Proc Natl Acad Sci USA 103:3863–3868.

Kurdi B, Lozano S, Banaji MR (2017) Introducing the Open Affective Standardized Image Set (OASIS). Behav Res Methods 49:457–470.

Lieberman MD, Eisenberger NI (2015) The dorsal anterior cingulate cortex is selective for pain: results from large-scale reverse inference. Proc Natl Acad Sci USA 112:15250–15255.

Lieberman MD, Burns SM, Torre JB, Eisenberger NI (2016) Reply to Wager et al.: Pain and the dACC: the importance of hit rate-adjusted effects and posterior probabilities with fair priors. Proc Natl Acad Sci USA 113:E2476–E2479.

Lindquist KA, Satpute AB, Wager TD, Weber J, Barrett LF (2016) The brain basis of positive and negative affect: evidence from a meta-analysis of the human neuroimaging literature. Cereb Cortex 26:1910–1922.

Moors A, De Houwer J, Hermans D, Wanmaker S, van Schie K, Van Harmelen AL, De Schryver M, De Winne J, Brysbaert M (2013) Norms of valence, arousal, dominance, and age of acquisition for 4,300 Dutch words. Behav Res Methods 45:169–177.

Nieuwenhuis S, Holroyd CB, Mol N, Coles MG (2004) Reinforcement-related brain potentials from medial frontal cortex: origins and functional significance. Neurosci Biobehav Rev 28:441–448.

Okon-Singer H, Lichtenstein-Vidne L, Cohen N (2013) Dynamic modulation of emotional processing. Biol Psychol 92:480–491.

Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh Inventory. Neuropsychologia 9:97–113.

Op de Beeck HP (2010) Against hyperacuity in brain reading: spatial smoothing does not hurt multivariate fMRI analyses? Neuroimage 49:1943–1948.

Peirce JW (2007) PsychoPy: psychophysics software in Python. J Neurosci Methods 162:8–13.

Rainville P (2002) Brain mechanisms of pain affect and pain modulation. Curr Opin Neurobiol 12:195–204.

Schmidt JR (2019) Evidence against conflict monitoring and adaptation: an updated review. Psychon Bull Rev 26:753–771.

Shackman AJ, Salomons TV, Slagter HA, Fox AS, Winter JJ, Davidson RJ (2011) The integration of negative affect, pain and cognitive control in the cingulate cortex. Nat Rev Neurosci 12:154–167.

Shenhav A, Botvinick MM, Cohen JD (2013) The expected value of control: an integrative theory of anterior cingulate cortex function. Neuron 79:217–240.

Shenhav A, Cohen JD, Botvinick MM (2016) Dorsal anterior cingulate cortex and the value of control. Nat Neurosci 19:1286–1291.

Silvestrini N, Chen JI, Piché M, Roy M, Vachon-Presseau E, Woo CW, Wager TD, Rainville P (2020) Distinct fMRI patterns colocalized in the cingulate cortex underlie the after-effects of cognitive control on pain. Neuroimage 217:116898.

Stroop JR (1935) Studies of interference in serial verbal reactions. J Exp Psychol 18:643–662.

van den Hurk J, Op de Beeck HP (2019) Generalization asymmetry in multivariate cross-classification: when representation A generalizes better to representation B than B to A. BioRxiv 592410.

Wisniewski D (2018) Context-dependence and context-invariance in the neural coding of intentional action. Front Psychol 9:2310.

Yarkoni T, Poldrack RA, Nichols TE, Essen DC, Wager TD (2011) Large-scale automated synthesis of human functional neuroimaging data. Nat Methods 8:665–670.