1

2

3

4 **Shared Neural Representations of Cognitive Conflict and Negative Affect in the Dorsal**

5 **Anterior Cingulate Cortex**

6

7 Luc Vermeylen[1*], David Wisniewski[1], Carlos González-García[1], Vincent Hoofs[1], Wim

8 Notebaert[1], Senne Braem[1,2]

9

10 [1]Department of Experimental Psychology, Ghent University, Belgium

11 [2]Department of Experimental and Applied Psychology, Vrije Universiteit Brussel, Belgium

12

13 *Correspondence

14 Luc Vermeylen

15 Department of Experimental Psychology

16 Henri Dunantlaan 2

17 B – 9000 Ghent

18 BELGIUM

19 **Email:** Luc.Vermeylen@ugent.be

Abstract

Influential theories of dorsal anterior cingulate cortex (dACC) function suggest that the dACC registers cognitive conflict as an aversive signal, but no study directly tested this idea. In this pre-registered human fMRI study, we used multivariate pattern analyses to identify which regions respond similarly to conflict and aversive signals. The results show that, of all conflict- and value-related regions, only the dACC/pre-SMA showed shared representations, directly supporting recent dACC theories.

Main

The dACC has been implicated in various psychological processes such as cognitive control, somatic pain, emotion regulation, reward learning and decision making[1–3]. In the domain of cognitive control, dACC is consistently activated by cognitive conflict, that is, the simultaneous activation of mutually incompatible stimulus, task, or response representations[4]. It has been proposed that dACC generates a domain-general aversive learning signal which biases behavior away from costly information processing (e.g., conflict)[5–7]. Recent behavioral studies indeed demonstrated that humans dislike and tend to avoid conflict, and automatically evaluate conflict as aversive[8–10]. Similarly, it has been proposed that conflict and negative affect are integrated in the dACC[3,9,11]. Given these proposals and findings, one would expect conflict and negative affect to be encoded similarly in dACC ("shared representations").

One recent study tried to investigate this hypothesis using a repetition suppression procedure, and found that dACC showed an attenuated response to negative affect following cognitive conflict[12]. However, other studies failed to provide evidence for this idea. For example, a number of studies and meta-analyses demonstrated that distinct parts of the ACC are associated with
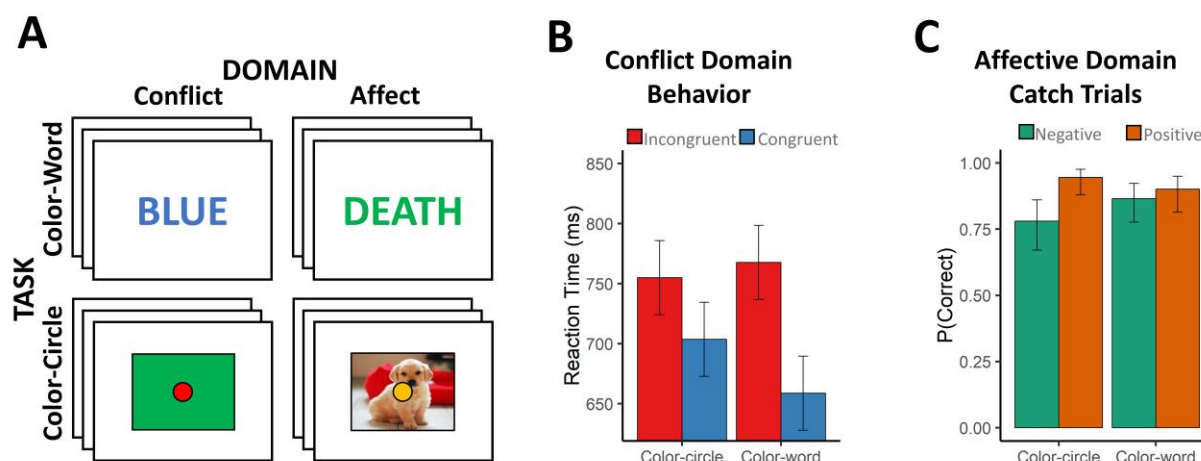
42   cognitive conflict and pain processing[13–16]. Similarly, a recent meta-analysis failed to observe

43   overlap between cognitive control, pain processing, and (negative) emotion in the medial

44   Prefrontal Cortex[17]. However, these previous studies often focus on peak activations across fMRI

45   studies that differ in experimental control, or involve intense pain responses that could mask

46   similarities with the arguably subtler affective evaluation of cognitive conflict.

47   Here, we took a different approach and developed a tightly controlled within-subjects test of

48   shared neural representations of conflict and affect in the brain. Namely, by using multivariate

49   cross-classification analyses, we assessed whether and where a classifier algorithm trained to

50   discern conflict (incongruent vs congruent events) can successfully predict affect (negative vs

51   positive events), and vice versa. Successful classification would indicate a similarity between the

52   neural pattern response, and thus a shared representational code between these two domains[18,19].

53   Specifically, 38 human subjects performed a color Stroop[20] and flanker task[21] in the conflict

54   domain, and two closely matched tasks in the affective domain (Fig. 1A). Importantly, we used

55   two tasks in each domain in order to demonstrate an abstract representation of conflict (and

56   affect), that is independent of conflict type (and affect source)[22]. Conflict and affect-related brain

57   signals were used to perform a leave-one-run-out cross-classification analysis using a linear

58   Support Vector Machine (see Methods). We performed preregistered Region of Interest (ROI)

59   and whole brain searchlight analyses (Supplementary Table 1), and report accuracy-minus-

60   chance values for each ROI and searchlight sphere (ROIs: Amygdala, Anterior Cingulate Cortex

61   [ACC], dACC/pre-SMA, Anterior Insula [AI], Posterior Cingulate Cortex [PCC], Ventral

62   Striatum [VS], and the ventromedial Prefrontal Cortex [vmPFC]).

63

64

65   **Figure 1.** Task Design and Behavioral Data. **(A)** Task design. Subjects either judged the color of

66   words or the color of circles. In the conflict domain, the color either matched or mismatched with

67   word meaning or background color creating congruent or incongruent conditions, respectively.

68   In the affective domain, positive or negative words and pictures were used to create the

69   respective conditions. These four task contexts were presented block-wise. **(B)** In the conflict

70   domain, typical congruency effects were found ($F_{(1,37)}=148.81$, $p<.001$, $BF>100$), which were

71   larger in the color-word task ($F_{(1,37)}=35.55$, $p<.001$, $BF>100$). **(C)** On catch trials in the affective

72   domain, subjects had to make a valence judgement (positive or negative) on the affective

73   background stimuli.

74

75   The behavioral data (Fig. 1B and Supplementary Table 3) and univariate brain results

76   (Supplementary Table 2) from the conflict tasks showed the typical differences between

77   congruent and incongruent trials. In the affective tasks, catch trials (where subjects had to make a

78   valence judgement instead of a color judgement) and a post-experiment incidental memory test

79   were used to inform processing of the (task-irrelevant) affective stimuli (see Supplementary

80   Table 4 for behavioral results). We observed above-chance catch trial performance (chance level

81   = 50%; see Fig. 1C and Methods) and successful post-experiment incidental recognition of the

82   affective stimuli (Supplementary Figure 5), ensuring that subjects processed the affective

83   pictures.

84   In a first set of multivariate pattern analyses, we trained and tested a classifier within-task (within

85   the Stroop or flanker task; Fig. 2A, left panels; which regions respond to conflict within tasks?),

86   as well as cross-task (train and test on different tasks; which regions respond similarly to conflict

87   independent of low-level task features?), in each of our preregistered ROIs (for analysis details,

88   see Method and Fig. 2B, left panels). Within-task ROI analyses in the conflict domain

89   (congruent vs. incongruent) revealed evidence for above chance-level decoding in the dACC/pre-

90   SMA (*Wilcoxon V*=327, *P*=.009, *BF10*=8.48), but not in in any of the other regions (all *P*>.060,

91   *BF*<0.60) (Fig. 2A, right panel). This decoding accuracy in the dACC/pre-SMA did not differ by

92   task ($F_{(1,37)}$=0.72, *P*=.400, *BF*=0.34). Second, the results show for the first time a conflict

93   representation independent of conflict task as within-conflict cross-task ROI analyses revealed

94   above-chance level conflict decoding in the dACC/pre-SMA (*V*=283, *P*=.012, *BF*=5.57). Again,

95   decoding accuracy did not differ between cross-task combination (i.e., from flanker to Stroop or

96   Stroop to flanker) ($F_{(1,37)}$=0.89, *P*=.352, *BF*=0.35) (Fig. 2B, right panel). These results were also

97   replicated in an overall decoding approach where the classifier was trained and tested in the

98   whole domain regardless of task (resulting in more samples to train the classifier; Supplementary

99   Fig. 1A). Within the affective domain (positive vs. negative), we also performed these within-

100   and cross-task decoding analyses. However, while these analyses showed evidence for affect

101   information in the insula, they did not show evidence for decoding in the dACC/pre-SMA

102   (Supplementary Fig. 2).

103   Finally, we evaluated our main hypothesis by training a classifier on discerning conflict

104   (incongruent vs congruent) and testing its performance on discerning affect (negative vs

105   positive), and vice versa. For this analysis, we focussed on the cross-domain cross-task decoding

106   (train and test in different domains on different tasks) as this analysis also controls for more low-

107   level shared features between the two tasks (Fig. 2C, right panel). The cross-domain cross-task

108   ROI decoding revealed evidence for cross-classification in the dACC/pre-SMA (V=330, P=.007,

109   BF=8.43; Fig. 2C, right panel), which did not differ by cross-task combination ($F_{(1,37)}$=0.36,

110   P=.551, BF=0.29). None of the other ROIs reached significance (all Ps>.101). These results

111   were replicated with the overall decoding approach in the main dACC/pre-SMA ROI (V=449,

112   P=.021, BF=4.65; Supplementary Fig. 1C).

113

114
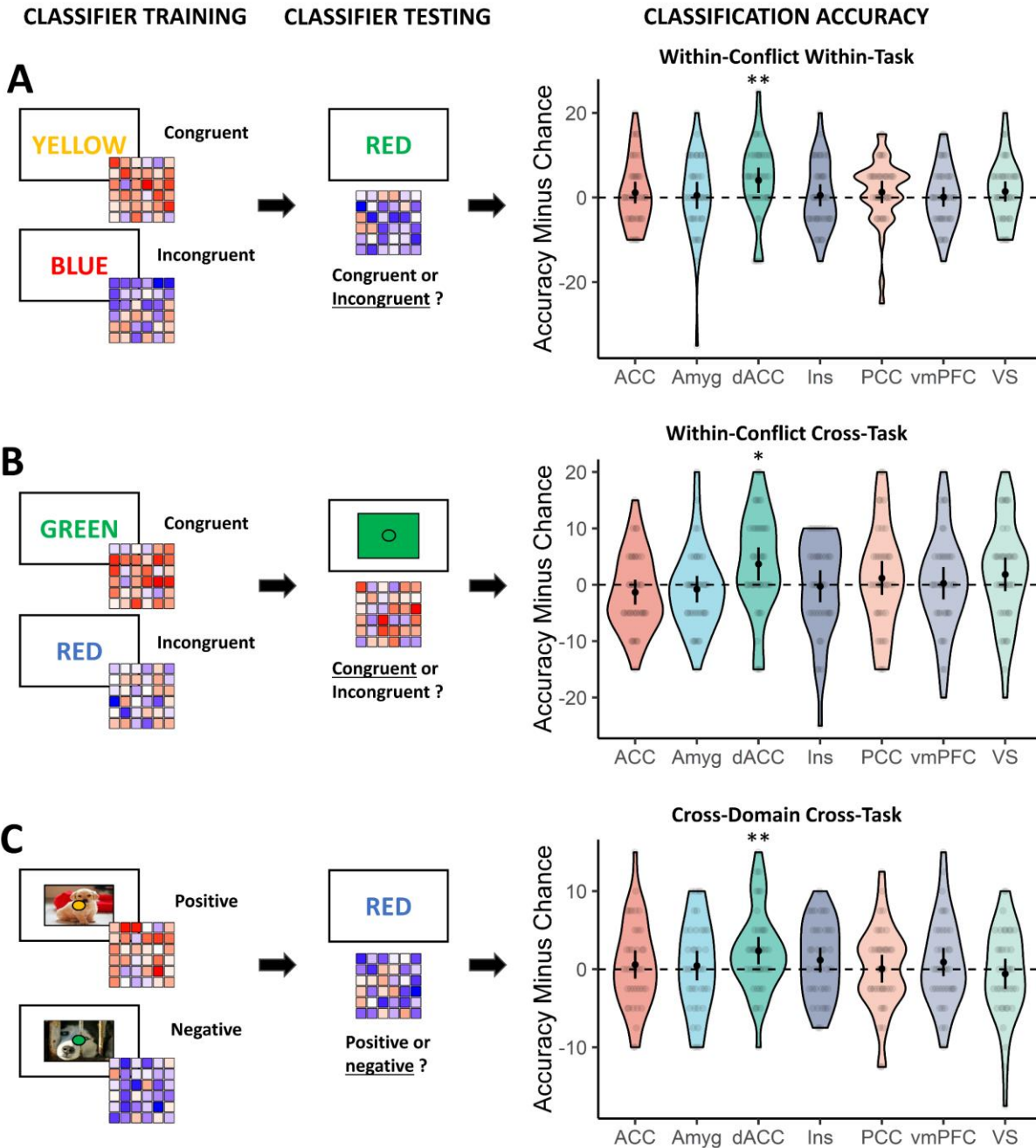
Figure 2. Main Results. **(A)** Training and testing the classifier within the conflict domain, within

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 affect and testing its performance on

classifying conflict across-tasks (and vice versa). *P<.05; **P<.01; black dots and error bars

119   represent mean and $\pm$ 95 CI respectively; transparent dots represent individual data points; the

120   shape of the violin shows the distribution of the data.

121

122   A number of control analyses further confirmed our main finding. First, we replicated this result

123   using different smoothing parameters (Supplementary Fig. 3), or when using spherical ROIs

124   instead of the Harvard-Oxford atlas ROIs (Supplementary Fig. 4). Second, also when using a set

125   of functionally (rather than anatomically) defined conflict-sensitive ROIs based on a recent meta-

126   analysis[23] (Supplementary Fig. 1, panel D), we again observed evidence for cross-domain cross-

127   task classification in the dACC/pre-SMA (V=450, P=.013, BF=3.75) but not for other conflict-

128   sensitive ROIs (left MOG, right AI, left AI, left IFG, left IPL, right IPL, left MFG), except for

129   the left AI (V=425, P=.005, BF=8.61). The result again replicated when using the overall

130   decoding approach in the dACC/pre-SMA (V=449, p=.001, BF=41.06), but not in the left AI

131   (V=335, P=.260, BF=0.34).

132   Together, our results are the first to show that the dACC/pre-SMA shows a similar voxel pattern

133   response to conflict and negative affect, and thereby offer important support for the popular

134   proposal that the dACC registers conflict as an aversive signal[3,5,6], thought to bias behavior away

135   from costly, demanding or suboptimal outcomes (as evidenced by behavioral avoidance and

136   negative evaluation of conflict[8,9]).

137   Moreover, our study is also the first to show decoding of conflict across conflict tasks in the

138   dACC, suggesting a shared component in the detection of conflict across the Stroop and flanker

139   task[22]. The fact that we did not observe a similar (significant) above-chance decoding of affect in

140   the dACC, but did observe cross-domain decoding, might seem surprising. However, this most

8

141   likely suggests differences in signal to noise ratio (SNR) between the two domains and does not

142   invalidate the cross-domain decoding result[24]. A lower SNR in the affect domain can be

143   explained by the fact that affect was not relevant for the main task.

144   The present findings also contradict the idea that cognitive control and affect are processed in

145   distinct subdivisions of the ACC (e.g., dorsal-cognitive vs. ventral-emotional[14]). While the

146   integration of cognitive control and affect in the dACC gained traction over the last two

147   decades[3,25], direct evidence for this idea was lacking, and recent (meta-analytical) studies were

148   more in line with the idea that both are processed in different subregions[13,17]. These studies were

149   problematic for many theories of dACC functioning as these theories often hold the (implicit)

150   assumption that dACC's response to suboptimal outcomes (e.g., conflict) has an evaluative

151   component (e.g., signaling avoidance learning[3,5], expected value of control[6,7], value of the non-

152   default option[26], evaluating action-outcome expectancies[27]). By using a tightly controlled within-

153   subject design and multivariate analysis techniques, we now show that conflict and negative

154   affect are indeed integrated in the dACC/pre-SMA, thereby providing important support for a

155   more integrative view and current theories of dACC functioning.

156   **Acknowledgements**

163   in the present experiment were carried out with adequate understanding and written consent of

164   the subjects and are in accordance with the Declaration of Helsinki.

165   **Author Contributions**

166   S.B. and W.N. developed the study concept. S.B., W.N. and L.V. contributed to the study design.

167   Data collection was performed by L.V. and V.H.. Data analysis was performed by L.V. under the

168   supervision of S.B., D.W. and C.G.C.. The manuscript was drafted by L.V. in cooperation with

169   S.B., W.N., D.W. and C.G.C.. All authors approved the final version of the manuscript for

170   submission.

171   **Competing Interests**

172   The authors have no competing interests to declare.

173   **References**

174   1. Ebitz, R. B. & Hayden, B. Y. *Nat. Neurosci.* **19**, 1278 (2016).

175   2. Heilbronner, S. R. & Hayden, B. Y. *Annu. Rev. Neurosci.* **39**, 149–170 (2016).

176   3. Shackman, A. J. *et al. Nat. Rev. Neurosci.* **12**, 154–167 (2011).

177   4. Botvinick, M. M. *et al. Psychol. Rev.* **108**, 624–652 (2001).

178   5. Botvinick, M. M. *Cogn. Affect. Behav. Neurosci.* **7**, 356–366 (2007).

179   6. Shenhav, A., Botvinick, M. M. & Cohen, J. D. *Neuron* **79**, 217–240 (2013).

180   7. Shenhav, A., Cohen, J. D. & Botvinick, M. M. *Nat. Neurosci.* **19**, 1286 (2016).

181   8. Dreisbach, G. & Fischer, R. **24**, 255–260 (2015).

182   9. Inzlicht, M., Bartholow, B. D. & Hirsh, J. B. *Trends Cogn. Sci.* **19**, 126–132 (2015).

183   10. Dignath, D., Eder, A. B., Steinhauser, M. & Kiesel, A. *Psychon Bull Rev.* In press.

184   11. Lieberman, M. D. & Eisenberger, N. I. *Proc. Natl. Acad. Sci.* **112**, 15250–15255 (2015).

185   12. Braem, S. *et al. J. Cogn. Neurosci.* **29**, 137–149 (2017).

186   13. Jahn, A., Nee, D. E., Alexander, W. H. & Brown, J. W. *J. Neurosci.* **36**, 12385–12392

187   (2016).

188   14. Bush, G., Luu, P. & Posner, M. I. *Trends Cogn. Sci.* **4**, 215–222 (2000).

189   15. Lieberman, M. D., Burns, S. M., Torre, J. B. & Eisenberger, N. I. *Proc. Natl. Acad. Sci.* **113**,

190   E2476–E2479 (2016).

191   16. De La Vega, A., Chang, L. J., Banich, M. T., Wager, T. D. & Yarkoni, T. *J. Neurosci.* **36**,

192   6553–6562 (2016).

193   17. Kragel, P. A. *et al. Nat. Neurosci.* **21**, 283 (2018).

194   18. Kaplan, J. T., Man, K. & Greening, S. G. *Front. Hum. Neurosci.* **9**, 151 (2015).

195   19. Wisniewski, D. *Front. Psychol.* **9**, (2018).

196   20. Stroop, J. R. *J. Exp. Psychol.* **18**, 643 (1935).

197   21. Eriksen, B. A. & Eriksen, C. W. *Percept. Psychophys.* **16**, 143–149 (1974).

198   22. Jiang, J. & Egner, T. *Cereb. Cortex* **24**, 1793–1805 (2013).

199   23. Chen, T. *et al.* A *Brain Struct. Funct.* **223**, 3813–3840 (2018).

200   24. van den Hurk, J. & de Beeck, H. P. O. *bioRxiv* 592410 (2019).

201   25. Okon-Singer, H., Hendler, T., Pessoa, L. & Shackman, A. J. *Front. Hum. Neurosci.* **9**, 58

202   (2015).

203   26. Calhoun, A. J. & Hayden, B. Y. *Curr. Opin. Behav. Sci.* **5**, 24–31 (2015).

204   27. Brown, J. W. & Alexander, W. H. *J. Cogn. Neurosci.* **29**, 1656–1673 (2017).

205

206

207

208

209   **Methods**

210   *Participants*

211   The study was pre-registered with the pre-registration template from AsPredicted.org on the

212   Open Science Framework (https://osf.io/p5frq/). As pre-registered, 40 participants participated in

213   our study. Two participants were excluded (one due to excessive head motion [>2.5mm

214   translation] and one aborted the scanning session). The average age of the remaining 38

215   participants (13 male) was 23.71 years (*SD*=3.53, min=18, max=33). Thirty-six participants

216   were right-handed, one was left-handed and one was ambidextrous (as assessed by the Edinburgh

217   Handedness Inventory[28]). Every participant had normal or corrected to normal vision and

218   reported no current or history of neurological, psychiatric or major medical disorder. Every

219   participant gave their informed written consent before the experiment, and was paid 35 euros for

220   participating afterwards. The study was approved by the local ethics committee (University

221   Hospital Ghent University, Belgium).

222   *Experimental Paradigm*

223   The experiment was implemented using Psychopy 2 version 1.85.2[29]. On each trial, participants

224   had to judge the color of a target stimulus in the center of the screen, using two MR-compatible

225   response boxes (each box had two buttons) to indicate one out of four possible response options

226   (red, blue, green and yellow). The key-to-color mapping was counterbalanced between

227   participants. The exact features of the target stimulus varied block-wise, depending on one of

228   four different task-contexts. Specifically, participants either had to respond to the color of words

229   ("color-word naming task") or respond to the color of circles ("color-circle naming task"), which

230   both had a conflict and affective version.

231　The conflict-version of the color-word naming task was a Stroop task[20], where the meaning of

232　the words could either be congruent or incongruent with the actual color of the word. For

233　example, participants could see the words "BLUE", "RED", "GREEN" or "YELLOW" (Dutch:

234　"ROOD", "BLAUW", "GROEN" or "GEEL") presented in a blue, red, green or yellow font. The

235　conflict version of the color-circle naming task was essentially a color-based variant on the

236　Eriksen flanker task[21], where the irrelevant feature consisted of a colored background square

237　which could either be congruent or incongruent with the color of the circle. Here, participants

238　could see blue, red, green or yellow circles presented on a blue, red, green or yellow background

239　square. In both tasks, half of the trials were congruent (e.g., "RED" in a red font; a red circle

240　presented on a red square background) while the other half of the trials were incongruent (e.g.,

241　"RED" in a blue font; a red circle on a blue square background).

242　The affect-versions of the color-word naming and color-circle naming tasks made use of

243　irrelevant affective words or pictures, respectively. In the color-word naming task, 16 positive

244　and 16 negative words were presented[30] that were matched on arousal, power, age of acquisition,

245　Dutch word frequency[31], word length and grammatical category (Noun, Adjective and Verbs).

246　The affective picture distractors in the background of the color-circle naming task were retrieved

247　from the OASIS database[32]. Sixteen positive and 16 negative pictures were presented that were

248　matched on semantic category (Animals, Objects, People, Scenery) and arousal. This resulted in

249　a total of eight conditions: congruent, incongruent, positive or negative trials, that either involved

250　words or pictures/colored backgrounds. While our stimuli were matched on arousal, we also

251　performed a control analysis where we trained a classifier to distinguish low versus high

252　arousing stimuli (matched on valence) and tested its performance on distinguishing congruent

253　versus incongruent stimuli (and vice versa). In contrast to our affect decoding results, this cross-

254　domain cross-task decoding was not significant in the dACC/pre-SMA (V=294, P=403,

255　*BF10*=0.26).

256　Each trial started with a fixation sign ("+") that was presented for 3 to 6.5 seconds (in steps of

257　0.5 s; M=3.5 s; drawn from an exponential distribution). Next, the target stimulus was presented

258　for 1.5 seconds (fixed presentation time regardless of RT). In order to increase the saliency of the

259　irrelevant dimension (conflict and affect), the onset of the affective word or picture preceded the

260　presentation of the target feature by 200 ms during which the color of the target feature (word or

261　circle) was white.

262　Participants performed five scanning runs and during each run the subjects performed each of the

263　four task contexts in separate blocks. The order of the four blocked task contexts was fixed

264　within participant but counterbalanced between participants. Each block hosted 32 trials (16

265　congruent/positive and 16 incongruent/negative) which were presented in a pseudo-random

266　fashion with the following restriction: neither relevant nor irrelevant features of the target

267　stimulus could be repeated. This restriction was used to investigate confound-free congruency

268　sequence effects (see [33]; but this was not the aim of the current study and will not be discussed

269　further). In total, each participant made 640 trials (i.e., five runs of four blocks of 32 trials).

270　In each task context (block), we also included one catch trial (at random, but not in the first two

271　or last two trials of each block). In these catch trials, the presentation of the task-irrelevant word,

272　picture, or colored square would not be followed by the presentation of the target color, and

273　remain on screen for three seconds. Participants were instructed that during these catch trials,

274　when no color information was present in the relevant dimension, their goal was to judge the

275　irrelevant dimension depending on the cognitive domain. In the conflict domain, participants had

276　to respond to the meaning of the word ("RED", "BLUE", "GREEN" or "YELLOW") or to the

277   color of the background square (red, blue, green or yellow) by using the respective key that

278   would be used to judge the relevant dimension. In the affective domain, participants had to judge

279   the affective word or background picture as either positive or negative by pressing all keys once

280   or twice (response mapping for positive and negative stimuli counterbalanced between

281   participants). The purpose of these catch trials was to increase the saliency of the irrelevant

282   dimension.

283   Before the scanning session, participants were welcomed and instructed to read the informed

284   consent after which they started practicing the experimental paradigm. After the scanning

285   sessions, participants performed an unannounced recognition memory test on old and new

286   affective words and pictures. Here, participants had to indicate whether they had previously seen

287   the word or picture in the experiment (old/new judgement). The new words were matched with

288   the old words in terms of valence, arousal, power, age of acquisition, word length, frequency,

289   grammatical category. The new pictures were matched on valence, arousal and semantic

290   category. In both a behavioral (n = 20) and fMRI pilot (n = 20), we already established that

291   participants showed adequate performance on both the main task and the recognition memory

292   task. Finally, participants completed four questionnaires (Need for Cognition, Behavioral

293   Inhibition/Activation Scale, Positive and Negative Affect Schedule, Barret Impulsivity Scale)

294   and were thanked for their participation. No significant correlations between these questionnaire

295   scales and cross-classification accuracies were found, so we do not report these results.

296   *Behavioral Data Analysis*

297   Behavioral analyses were performed in R (RStudio version 1.1.463, www.rstudio.com). For the

298   reaction time (RT) analyses, we removed incorrect, premature (< 150 ms), and extreme

299   responses (RTs outside 3 SD from each condition mean for each participant). This resulted in an

15

300   average of 94.42 % of the trials left for the RT analyses (SD=3.18, min=84.22, max=98.28). We

301   conducted a repeated measures ANOVA on the reaction time and accuracy measure with the

302   within-subject factors Condition (conflict domain: congruent vs. incongruent, affective domain:

303   positive vs. negative) and Task (color-word naming vs. color-circle naming). We also assessed

304   post-scanning recognition memory of affective stimuli with a probit generalized linear mixed

305   effects model on the probability to say that the stimulus was 'old' with fixed effects for

306   Experience (old vs. new), Valence (positive vs. negative) and Task Type (word vs. picture) and

307   crossed random effects for Participant and Item. We also pre-registered some exclusion criteria

308   based on behavioral performance. Participants with a mean RT outside 3 SD from the sample

309   mean or a hit rate below 3 SD or 60 % (chance level=25 %) from the sample mean were

310   excluded. Participants that performed poorly on the post-scanning recognition memory test, i.e.,

311   hit rate or false alarm rate outside 3 SD of the sample mean were also excluded. In the end, no

312   exclusions based on task performance had to be made. While performance on catch trials was not

313   a pre-registered exclusion criterion, we found that two participants responded on chance level in

314   the catch trials of the affective domain (chance level=50 %, positive vs. negative judgement).

315   Excluding these participants did not change our conclusions.

316   *fMRI data acquisition*

317   fMRI data was collected using a 3T Magnetom Trio MRI scanner system (Siemens Medical

318   Systems, Erlangen, Germany), with a sixty-four-channel radio-frequency head coil. A 3D high-

319   resolution anatomical image of the whole brain was acquired for co-registration and

320   normalization of the functional images, using a T1-weighted MPRAGE sequence (TR=2250 ms,

321   TE=4.18 ms, TI=900 ms, acquisition matrix=256 × 256, FOV=256 mm, flip angle=9°, voxel

322   size=1 × 1 × 1 mm). Furthermore, a field map was acquired for each participant, in order to

16

323   correct for magnetic field inhomogeneities (TR=520 ms, TE1=4.92 ms, TE2=7.38 ms, image

324   matrix=70 x 70, FOV=210 mm, flip angle=60°, slice thickness=3 mm, voxel size=3 x 3 x 2.5

325   mm, distance factor=0%, 50 slices). Whole brain functional images were collected using a T2*-

326   weighted EPI sequence (TR=1730 ms, TE=30 ms, image matrix=84×84, FOV=210 mm, flip

327   angle=66°, slice thickness=2.5 mm, voxel size=2.5 x 2.5 x 2.5 mm, distance factor=0%, 50

328   slices) with slice acceleration factor 2 (Simultaneous Multi-Slice acquisition). Slices were

329   orientated along the AC-PC line for each subject.

330   *fMRI data analysis*

331   fMRI data analysis was performed using Matlab (version R2016b 9.1.0, MathWorks) and

332   SPM12 (www.fil.ion.ucl.ac.uk/spm/software/spm12/). Raw data was imported according to

333   BIDS standards (http://bids.neuroimaging.io/) and functional data was subsequently realigned,

334   slice-time corrected, normalized (resampled voxel size 2 mm^3) and smoothed (full-width at half

335   maximum of 8 mm). The preprocessed data was then entered into a first-level general linear

336   model analysis (GLM), and subsequently into a multivariate pattern analysis (MVPA[34–37]).

337   Results were analyzed using a mass-univariate approach. Although we pre-registered that we

338   would not normalize and smooth the data for our classification analyses, we found that Signal-to-

339   Noise Ratio (SNR) was significantly improved with these additional preprocessing steps

340   (Supplementary Fig. 3A). In addition, an independent classification analysis (classifying left vs.

341   right responses in primary motor cortex) showed that decoding accuracies were significantly

342   higher with these additional preprocessing steps (Supplementary Fig. 3B). Knowing that

343   decoding information in the PFC is notoriously difficult as decoding accuracies are close to

344   chance (relative to decoding in occipitotemporal cortex[38]), and the finding that smoothing can

345   and does often improve SNR and decoding performance[39–41], we decided to optimize our MVPA

17

346 analyses by decoding on normalized and smoothed data. For completeness, however, we also

347 depict the results from our main cross-classification analysis for different levels of smoothing

348 (FWHM 0, 4 and 8 mm; see Supplementary Fig. 3C).

349 First-level GLM analyses consisted of 5 identically modeled sessions (i.e., the five runs). Each

350 session consists of eight regressors of interest (for the eight conditions, see above), four block

351 regressors (to account for the blocked presentation of each combination of word versus picture

352 versions of the conflict versus affect tasks), two nuisance regressors (that model performance

353 errors and catch trials) and six movement regressors. The regressors were convolved with the

354 canonical HRF. The modeled duration of the regressors of interest (the eight conditions) and

355 nuisiance regressors (errors, catch trials) was zero, while the modeled duration of the block

356 regressors was equal to the length of the blocks.

357 Next, the beta images from the first-level GLM were submitted to leave-one-run-out decoding

358 scheme with 'The Decoding Toolbox'[42]using a linear support-vector classification algorithm

359 (C=1). We performed whole-brain searchlight decoding (sphere radius: 3 voxels; Supplementary

360 Table 1) as well as ROI decoding (see below for ROI methods). Cross-validation decoding was

361 conducted within the affective (positive vs. negative) and conflict (congruent vs. incongruent)

362 domain for each task separately ("within-domain within-task classification"). To assess the

363 generalizability of the classifier within the domain, we also conducted cross-classification

364 analyses where we trained the classifier on one task and tested its performance on the other task

365 for each task type combination (from color-circle naming to color-word naming and vice versa)

366 separately ("within-domain cross-task classification"). To investigate the generalizability of

367 these classifiers across the domain (our main hypothesis), we trained the classifier in the conflict

368 domain and tested its performance in the affective domain, and vice versa. We conducted these

369   analyses cross task type combinations (i.e., from color-circle naming to color-word naming, or

370   from color-word naming to color-circle naming) to further control for low-level task features,

371   following the same reasoning as the within-domain cross-task classification analyses. The results

372   from these classification analyses were then averaged to return the cross-domain cross-task

373   decoding results. For each of these three decoding analyses, we also ran ANOVAs to evaluate

374   whether the result differed depending on the task (e.g., color-circle naming versus color-word

375   naming) or task-to-task direction (i.e., from color-circle naming to color-word naming, or from

376   color-word naming to color-circle naming). Finally, we also report an "overall decoding"

377   analysis, where the classifier was trained across the two task types at once, thereby ignoring

378   whether the event featured words or pictures/colored backgrounds.

379   Each classification analysis resulted in 'accuracy-minus-chance' decoding maps for each subject.

380   These maps were then entered into a group second-level GLM analysis in SPM12. Here, a one-

381   sample t-test determined which voxels show significant accuracy above chance level.

382   Next to MVPA, we also conducted classic univariate analyses. Here, we constructed a set of

383   contrasts subtracting (A) positive from negative conditions and (B) congruent from incongruent

384   conditions for (1) each task separately as well as across both tasks. These contrast images were

385   then entered into a second-level analysis in which a one-sample t-test determined which voxels

386   show significant activation for each contrast. We applied a statistical threshold of $p < 0.001$

387   (uncorrected) at the voxel level, and $p < 0.05$ (family-wise error corrected) at the cluster level on

388   all analyses (Supplementary Table 2).

389   *ROI analyses*

390   As part of our pre-registered main analysis plan, we conducted ROI decoding analyses. We set

391   out to study the Amygdala, Anterior Cingulate Cortex (ACC), dorsal Anterior Cingulate

392   Cortex/pre-SMA (dACC/pre-SMA), Anterior Insula (AI), Parietal Cingulate Cortex (PCC),

393   Ventral Striatum (VS), and the ventromedial PFC (vmPFC). All ROIs were obtained from the

394   Harvard-Oxford cortical and subcortical structural atlases, thresholded at 25%. As the dACC

395   ROI was not defined in the Harvard-Oxford atlas, we decided to retrieve this ROI from

396   Neurosynth[43] by entering "dacc" as search term (returning 162 studies reporting 4547

397   activations). Although this ROI was based on the "dacc" search term, the peak effect of studies

398   reporting dACC activity actually lies more dorsally than the cingulate gyrus, overlapping with

399   the pre-SMA[11]*. Therefore, we refer to this ROI as the dACC/pre-SMA. Next, we built a 10 mm

400   sphere around the peak activation point in this activation map (association map). Because the

401   dACC ROI was spherical (in contrast to the other six atlas ROIs), we also re-analyzed our results

402   from the atlas ROIs with 10 mm spherical alternatives retrieved from Neurosynth, which

403   returned highly similar results and did not change our statistical conclusions.

404   In addition to the pre-registered ROI analyses which were based on anatomically determined

405   ROIs, we also ran a second set of ROI analyses with functionally informed ROIs. Namely, we

406   created 10 mm sphere ROIs for all conflict-sensitive regions based on the most recent and

407   inclusive meta-analysis we could find on cognitive conflict[23].

408   Each ROI decoding analysis returned one accuracy-minus-chance value per ROI and participant.

409   We tested whether these values were significantly higher than zero (one-tailed) with the non-

410   parametric Wilcoxon signed-rank test and a Bayesian t-test (using the default priors from the

411   BayesFactor package in R; Cauchy prior width: $r=.707$). We report the Bayes Factor (BF) that

412   quantifies the evidence for the alternative hypothesis (i.e., decoding accuracy is higher than

413   zero). Our pre-registered stopping criterion was if the main finding was $BF>6$ (i.e., or if we had

414   reached 40 subjects, for financial reasons), but we would like to note that, if so, this result was

415   typically also $p<.00714$, which is the Bonferroni-corrected alpha for the main set of 7 ROIs.

416   Finally, we investigated whether the significant cross-task cross-domain classification accuracy

417   correlated with the following behavioral indices: post-scanning affective recognition memory (d-

418   prime), congruency sequence effects in reaction time and error rate and congruency sequence

419   effects in reaction time and error rates (p-values of reported correlations are Holm-corrected for

420   five tests) (see Supplementary Figure 5).

421   **Data Availability**

422   The minimal data necessary to replicate the reported findings can be found on the Open Science

423   Framework (https://osf.io/p5frq/). Raw fMRI data and preprocessing scripts will be uploaded to a

424   repository in the near future.

425   **Code Availability**

426   The custom code used for the analyses of this study can be found on the Open Science

427   Framework (https://osf.io/p5frq/).

428   **References**

429   28. Oldfield, R. C. *Neuropsychologia* **9**, 97–113 (1971).

430   29. Peirce, J. W. *J. Neurosci. Methods* **162**, 8–13 (2007).

431   30. Moors, A. *et al. Behav. Res. Methods* **45**, 169–177 (2013).

432   31. Keuleers, E., Brysbaert, M. & New, B. *Behav. Res. Methods* **42**, 643–650 (2010).

433   32. Kurdi, B., Lozano, S. & Banaji, M. R. *Behav. Res. Methods* **49**, 457–470 (2017).

434   33. Braem, S. *et al. Trends Cogn. Sci.* **23**, 769–783 (2019).

435   34. Cox, D. D. & Savoy, R. L. *NeuroImage* **19**, 261–270 (2003).

436   35. Kriegeskorte, N., Goebel, R. & Bandettini, P. *Proc. Natl. Acad. Sci.* **103**, 3863–3868 (2006).

437   36. Haxby, J. V. *NeuroImage* **62**, 852–855 (2012).

438   37. Haynes, J.-D. *Neuron* **87**, 257–270 (2015).

439   38. Bhandari, A., Gagne, C. & Badre, D. *J. Cogn. Neurosci.* **30**, 1473–1498 (2018).

440   39. Kamitani, Y. & Sawahata, Y. *NeuroImage* **49**, 1949–1952 (2010).

441   40. Hendriks, M. H. A., Daniels, N., Pegado, F. & Op de Beeck, H. P. *Front. Neurol.* **8**, (2017).

442   41. Op de Beeck, H. P. *NeuroImage* **49**, 1943–1948 (2010).

443   42. Hebart, M. N., Görgen, K. & Haynes, J.-D. *Front. Neuroinformatics* **8**, (2015).

444   43. Yarkoni, T., Poldrack, R. A., Nichols, T. E., Essen, D. C. V. & Wager, T. D. *Nat. Methods* **8**,

445   665–670 (2011).

446