

The Neural Representations of Emotional Experiences Are More Similar Than Those of Neutral Experiences

 Martina Riberto,^{1,2} Rony Paz,^{2*}  Gorana Pobric,^{1*} and  Deborah Talmi^{3*}

¹Division of Neuroscience and Experimental Psychology, University of Manchester, Manchester M13 9PL, United Kingdom, ²Department of Neurobiology, Weizmann Institute of Science, Rehovot 7610001, Israel, and ³Department of Psychology, University of Cambridge, Cambridge CB2 3EB, United Kingdom

Stimuli that evoke the same feelings can nevertheless look different and have different semantic meanings. Although we know much about the neural representation of emotion, the neural underpinnings of emotional similarity are unknown. One possibility is that the same brain regions represent similarity between emotional and neutral stimuli, perhaps with different strengths. Alternatively, emotional similarity could be coded in separate regions, possibly those sensitive to emotional valence and arousal. In behavior, the extent to which people consider similarity along emotional dimensions when they evaluate the overall similarity between stimuli has never been investigated. Although the emotional features of stimuli may dominate explicit ratings of similarity, it is also possible that people neglect emotional dimensions as irrelevant to that judgment. We contrasted these hypotheses in (male and female) healthy controls using two measures of similarity and two picture databases of complex negative and neutral scenes, the second of which provided exquisite control over semantic and visual attributes. The similarity between emotional stimuli was greater than between neutral stimuli in the inferior temporal cortex, the fusiform face area, and the precuneus. Additionally, only the similarity between emotional stimuli was significantly represented in early visual cortex, anterior insula and dorsal anterior cingulate cortex. Intriguingly, despite the stronger neural similarity between emotional stimuli, the same participants did not rate them as more similar to each other than neutral stimuli. These results contribute to our understanding of how emotion is represented within a general conceptual workspace and of the overgeneralization bias in anxiety disorders.

Key words: emotions; multivoxels pattern analysis; neural representations; semantic knowledge; similarity judgments

Significance Statement

We tested differences in similarity between emotional and neutral scenes. Arousal and negative valence did not increase similarity ratings. When conditions were equated on semantic similarity, participants rated emotional stimuli as similar to each other as neutral ones. Despite this equivalence, the similarity among the neural representations of emotional compared with neutral stimuli was higher in regions, which also expressed similarity between neutral stimuli and in unique regions. We report a striking difference between behavioral and neural similarity; strong neural similarity between emotional pictures did not influence similarity judgments in the same participants in the behavioral rating task after the scan. These findings may have an impact on research about the neural representations of emotional categories and the overgeneralization bias in anxiety disorders.

Received July 21, 2021; revised Jan. 21, 2022; accepted Jan. 26, 2022.

Author contributions: M.R., R.P., G.P., and D.T. designed research; M.R. performed research; M.R. analyzed data; M.R., R.P., G.P., and D.T. wrote the paper.

This work was supported by the Weizmann Institute of Science PhD Studentship to M.R., G.P., and D.T. We thank Louis Renoult, Christopher Honey, Chris Frith and Carien van Reekum for commenting on a draft of this article.

*R.P., G.P., and D.T. share senior authorship.

The authors declare no competing financial interests.

Correspondence should be addressed to Martina Riberto at martina.riberito@manchester.ac.uk.

<https://doi.org/10.1523/JNEUROSCI.1490-21.2022>

Copyright © 2022 the authors

Introduction

We may judge an image of a homeless person and of a car accident as different because of their different meanings or as similar because both evoke negative feelings. Emotional similarity refers to the tendency to group stimuli because they evoke the same feelings (Riberto et al., 2019). The extent to which similarity along emotional dimensions influences perceived similarity between complex experiences is unknown. It is important to understand the effect of emotion on similarity because aberrant similarity perception influences psychological well-being (Puccetti et al., 2021) and is clinically relevant in anxiety and post-traumatic stress disorders (Laufer et al., 2016). For example, after a traumatic event,

patients may consider later experiences to be similar to the original fearful one, not because of their ostensible meaning but because of their emotional similarity.

All stimuli can be described according to their location on orthogonal dimensions, valence, and arousal, with their proximities reflecting aspects of their relationship (Russell, 1980). This perspective suggests that entirely neutral stimuli, at the origin of the axes, may be perceived to be just as similar as stimuli at the extremes. Yet, similarity inferred from single-stimulus judgements on single attributes (e.g., shape, valence) rarely explains more than half the variance in explicit ratings of similarity (Jordan et al., 2017). Indeed, highly arousing negative stimuli may be perceived less similar to each other than neutral ones if they evoke qualitatively different emotions (e.g., fear, anger). Previous comparisons reveal increased ratings of similarity among negative pictures than among randomly selected neutral pictures (Talmi, 2013) and among positive than negative stimuli (Koch et al., 2016). Unfortunately, previous rating studies used semantically related emotional stimuli, thereby confounding emotion and semantic similarity. Nevertheless, in conditioning paradigms, where semantic similarity is not a confound, wider generalization of aversively conditioned stimuli has been observed (Laufer and Paz, 2012). Therefore, we hypothesized that negative emotional stimuli will be perceived as more similar than neutral stimuli.

Neuroimaging studies observed low specificity for discrete emotions and provided evidence against a locationist perspective to the study of emotions (Hoemann et al., 2019). Instead, emotional stimuli are likely represented in distributed networks of cortical and subcortical regions, which are not functionally specific to affect (Chang et al., 2015) but carry out emotion-relevant computations, for example, the occipitotemporal regions, visual-semantic processing of emotional and neutral categories (Kragel et al., 2019); the insula and the anterior cingulate cortex (ACC), awareness of bodily sensations and visceral regulation necessary for a core affective state representation; and the ventral prefrontal cortex, positive valence (Lindquist et al., 2012). No previous work has directly compared the neural underpinnings of emotional and neutral similarity for complex, realistic stimuli, but a handful of studies used simple stimuli. Representational similarity analysis (RSA) maps similarity perception in the brain by correlating neural and behavioral data (Kriegeskorte et al., 2008a). This technique revealed increased neural similarity between conditioned stimuli in the amygdala (Visser et al., 2013), the occipitotemporal cortex (Dunsmoor et al., 2014), and the superior frontal gyrus (Visser et al., 2011), and increased similarity between stimuli that predict reward (Zeithamova et al., 2018) and pain (Wagner et al., 2020) in the hippocampus. In following with this theoretical and empirical work, we hypothesized that neural similarity will differ as a function of stimulus emotionality. Specifically, we hypothesized that the neural similarity among emotional stimuli will be greater than among neutral stimuli expressing the predicted pattern of behavioral ratings. Emotion may increase neural similarity in any regions that encode participants' self-reported similarity space but may do so more strongly in regions that serve emotion-relevant operations.

We tested these hypotheses in a series of experiments that present several strengths compared with the state of the art. We used different similarity judgement tasks and picture databases, one of which permitted for the first time control over taxonomic and thematic similarity and narrowed our search volume through innovative searchlight approaches.

Materials and Methods

Participants

A total of 90 participants were recruited from the University of Manchester in the United Kingdom and from the Weizmann Institute of Science in Israel to take part in the study (age range, 20–54 years; mean age, 30.14 years; SD, 7.17; experiment 1: 20 participants, 10 females; experiment 2: 40 participants, 20 females; experiment 3: 29 participants, 12 females; one participant was excluded for not following the instructions of the task). The sample size was selected according to previous publications in this research field (Charest et al., 2014; Chikazoe et al., 2014; Giordano et al., 2021). All participants had normal or corrected-to-normal vision and were over the age of 18. They gave informed consent before the experiment and have been reimbursed for their participation (5€ for the behavioral experiments, 22€ for the MRI experiment). The exclusion criteria were the following: a history of neurologic (e.g., head injury or concussion) or psychiatric conditions (e.g., depression, anxiety), drug or alcohol abuse, or regular medication that could influence emotional processing. The study was approved by the ethics board of the University of Manchester and of the Weizmann Institute of Science (protocol number 0287–09-TLV).

Materials

First database of complex pictures

In experiment 1, we selected 20 images taken from the Nencki Affective Picture System (NAPS) database (Marchewka et al., 2014). Picture IDs that we selected in experiment 1 are reported in Extended Data Figure 2-1. NAPS has been validated for use in emotional research (Wierzbica et al., 2015; Riegel et al., 2016) and consists of 1356 realistic, high-quality photographs divided into five categories (people, faces, animals, objects, and landscapes). To control for visual similarity, we matched the pictures for low-level visual features, which unlike subjective ratings of visual complexity, are not affected by the arousal complexity bias (Madan et al., 2018) and by the vividness bias (Todd et al., 2013). These measures included the luminance (the average pixel value of the greyscale image) and the contrast (the SD across all the pixels of the greyscale image; Bex and Makous, 2002). To quantify the colors within each image, we computed the quantity of red (R), green (G), and blue (B), according to the RGB color model. Finally, the JPEG size and the entropy of each greyscale image were used as indices of the overall visual complexity of each image (Donderi, 2006). The JPEG size was determined with a compression quality setting of 80 (on a scale from 1 to 100). Perceptually simple images are highly compressible and therefore result in smaller file sizes. The entropy, H , is computed from the histogram distribution of the 8-bit gray-level intensity values x , $H = -\sum p(x) \log p(x)$, where p represents the probability of an intensity value x . H varies with the randomness of an image. High-entropy images are noisier and have a high degree of contrast from one pixel to the next, whereas low-entropy images have rather large uniform areas with limited contrast. The sample of images included 10 emotional and 10 neutral images. The designation of images to this category was based on the NAPS ratings of valence and arousal on a 9-point scale provided by 204 European participants. We considered emotional pictures as rated <4 in the valence scale (negative valence) and >6 in the arousal scale (high arousal), whereas the neutral images ranged from 4 to 6 in both dimensions. To validate the NAPS norms, we also asked our participants to rate the valence and the arousal of the picture before the main task. Extended Data Figure 2-1 shows the picture IDs from the NAPS database (people category), divided into emotional and neutrals. Table 1 shows the mean and the SD of the different visual and emotional measures for emotional and neutral pictures, as well as the differences between them. We controlled to some extent for semantic similarity, namely, the similarity both in the theme (e.g., violence) each picture depicts, other categories it belongs to (e.g., outdoor scene), and its specific meaning. With this aim, we choose images that included more than one person in an outdoor scene from the same category—the people category. These images contain a lot of information beyond the people themselves, placing them in a rich and realistic context. The matching we achieved between emotional and neutral conditions exceeds that in most published studies and represents the current

Table 1. Differences in visual and emotional measures between emotional ($n = 10$) and neutral ($n = 10$) pictures (experiment 1)

		Categories		Statistics		
		Emotional	Neutral	t	p	d
Visual measures	Luminance	89.42 ± 27.31	110.13 ± 36.06	−1.45	0.16	−3.68
	Contrast	61.97 ± 1.16	62.19 ± 1.14	−0.49	0.96	−0.07
	R	93.24 ± 26.06	111.62 ± 26.67	−1.48	0.17	−3.48
	G	88.99 ± 27.92	110.12 ± 37.84	−1.41	0.19	−3.69
	B	82.58 ± 26.54	105.95 ± 41.24	−1.56	0.15	−4.04
	JPEG	337121.90 ± 90579.45	27707.50 ± 69785.15	1.66	0.11	3.86
Emotional measures	Entropy	7.50 ± 0.28	7.49 ± 0.24	0.11	0.91	0.02
	Valence	1.985 ± 0.743	4.840 ± 0.976	−10.22	0.000**	−3.09
	Arousal	7.170 ± 1.389	4.840 ± 1.219	8.93	0.000**	2.04

The mean and SD of each measure are shown, as well as the t and p values and Cohen's d as effect size measure for each difference; ** $p_{FWE} < 0.001$.

Table 2. Differences in visual and emotional measures among categories

	Categories				Statistics		
	E1	E2	N1	N2	F	p	η_p^2
Luminance	105.39 ± 21.37	95.68 ± 26.27	106.45 ± 29.79	106.22 ± 25.11	0.73	0.54	0.31
Contrast	58.63 ± 10.80	62.52 ± 6.09	64.99 ± 11.12	63.54 ± 12.69	1.22	0.31	0.05
R	115.07 ± 20.73	98.84 ± 26.71	114.15 ± 31.11	109.28 ± 26.30	1.47	0.23	0.06
G	102.18 ± 21.99	94.14 ± 26.23	104.88 ± 30.71	105.40 ± 26.07	0.70	0.55	0.03
B	96.50 ± 23.29	95.35 ± 28.25	94.32 ± 28.11	102.96 ± 28.94	0.37	0.77	0.02
JPEG	66701.89 ± 11078.33	63614.83 ± 11967.82	59643.05 ± 12220.83	67011.28 ± 28005.71	0.70	0.55	0.03
Entropy	7.58 ± 0.25	7.56 ± 0.26	7.54 ± 0.29	7.52 ± 0.22	0.20	0.90	0.00
Valence	2.91 ± 1.42	1.97 ± 1.02	4.91 ± 0.26	5.13 ± 0.30	46.93	0.00	0.84
Arousal	6.64 ± 1.40	7.74 ± 1.34	4.72 ± 1.37	4.53 ± 1.26	27.37	0.00	0.75

The mean and SD of each measure are shown, as well as the F and p values and the partial η squared for each difference. E1, emotional category 1 (poverty scenes, $n = 18$); E2, emotional category 2 (car accidents, $n = 18$); N1, neutral category 1 (laundry scenes, $n = 18$); N2, neutral category 2 (telephone scenes, $n = 18$; experiments 2–3).

state of the art in controlling emotional and neutral stimuli in research. However, the range of emotional themes was reduced compared with that in the neutral set. Therefore, emotional pictures might be rated as more similar because of the higher thematic similarity compared with neutrals.

Second database of complex pictures

In experiments 2–3, to control the emotional and neutral pictures for thematic similarity, we selected natural scenes in a way that all the categories depicted realistic events that do not co-occur in the environment. In particular, we chose 72 real-world color photographs using Google images that represented one or more people in outdoor situations. We divided them into four categories according to the scene that was depicted, resulting in 18 images per category. Two of the categories were neutral, and two were emotionally arousing and negatively valenced. These latter categories represented either poverty scenes [emotional category 1 (E1)] or car accidents [emotional category 2, (E2)]. The neutral categories portrayed either people hanging laundry to dry [neutral category 1 (N1)] or talking on the phone [neutral category 2 (N2)]. The full set of pictures can be found at in Extended Data Figure 3-1. We minimized the thematic similarity between emotional categories by selecting for each of the emotional categories action-context combinations that do not normally occur in a common theme or scenario. The same was true for the two neutral categories. To control for taxonomic similarity to some extent, all the pictures we selected shared two semantic features, they depicted people outdoors. Second, we controlled the pictures for affordance, namely the action that a scene can afford, by selecting pictures that depicted only one type of action—and therefore affordances—in each category. Specifically, in E1, people sit on the ground while begging; in E2, accident victim(s) lie either on a surface (the ground or a crashed car); in N1, people stand hanging and drying clothes; and in N2, they stand or walk in the street talking on the phone. Although these actions and affordances differed across the four categories, the design ensures that these differences did not influence comparisons across the two neutral and two emotional categories. Finally, we controlled the stimuli for visual properties, as in experiment 1. An independent sample

of 10 healthy participants rated the valence and the arousal of the stimuli, and another independent sample of 20 participants judged the similarity of the pictures. Table 2 shows the mean and SD of visual and emotional measures for each category as well as the differences among them. Table 3 shows the mean and SD of similarity measures within and between categories, as well as the differences among them.

Experimental design

A graphical representation of the general experimental design is shown in Figure 1. In all the experiments, we asked participants to judge the similarity of a set of complex pictures to test our main hypothesis for the behavioral data, that the perceived similarity between emotional compared with neutral pictures will be higher. As shown at the top of Figure 1, in the first two experiments participants performed a pairwise similarity rating task. In experiment 1, after rating the valence and arousal of each picture from the first dataset, participants rated all the possible combinations among the stimuli. In experiment 2 we focused on the ratings of interest (Fig. 1, bottom, red circles), and therefore participants only rated the similarity between emotional categories (E12) and between neutral categories (N12) of pictures from the second dataset, as well as between emotional and neutral categories (EN), with the latter pairs serving as catch trials. Experiments 1–2 ended after ~20 min. In experiment 3, after an fMRI scan, participants performed a surprise multiarrangements (MA) task to judge the similarity of the 72 pictures on a bidimensional space, as depicted at the top right of Figure 1 (duration ~1 h).

Valence and arousal rating task

The two dimensions of valence and arousal are considered to be key to the conceptual representation of semantic concepts as well as emotional stimuli. Therefore, we used these for stimulus selection so that those selected for the emotional condition differed from those selected for the neutral condition along both valence and arousal dimensions. To validate the designation of pictures from the two datasets to emotional and neutral conditions, participants completed a valence and arousal rating task, following the procedure suggested by Lang et al. (2008). Each trial

Table 3. Differences in dissimilarity among categories (validation study)

Dissimilarity within category				Dissimilarity between categories		
E1	E2	N1	N2	E12	N12	EN
0.004 ± 0.002	0.003 ± 0.003	0.004 ± 0.003	0.004 ± 0.003	0.017 ± 0.003	0.019 ± 0.002	0.024 ± 0.001
Statistics				Statistics		
Manipulation check				Main HP		
<i>F</i>	<i>p</i>	η_p^2	<i>F</i>	<i>p</i>	η_p^2	
292.56	<0.001	0.94	4.11	0.06	0.18	

The mean and SD of each condition of interest are shown, as well as the *F* and *p* values and the partial η squared for each difference. Dissimilarity within, E1, emotional category 1 (poverty scenes, *n* = 18); E2, emotional category 2 (car accidents, *n* = 18); N1, neutral category 1 (laundry scenes, *n* = 18); N2, neutral category 2 (telephone scenes, *n* = 18). Dissimilarity between, E12, emotional categories; N12, neutral categories; EN, emotional and neutral categories.

Experimental procedure

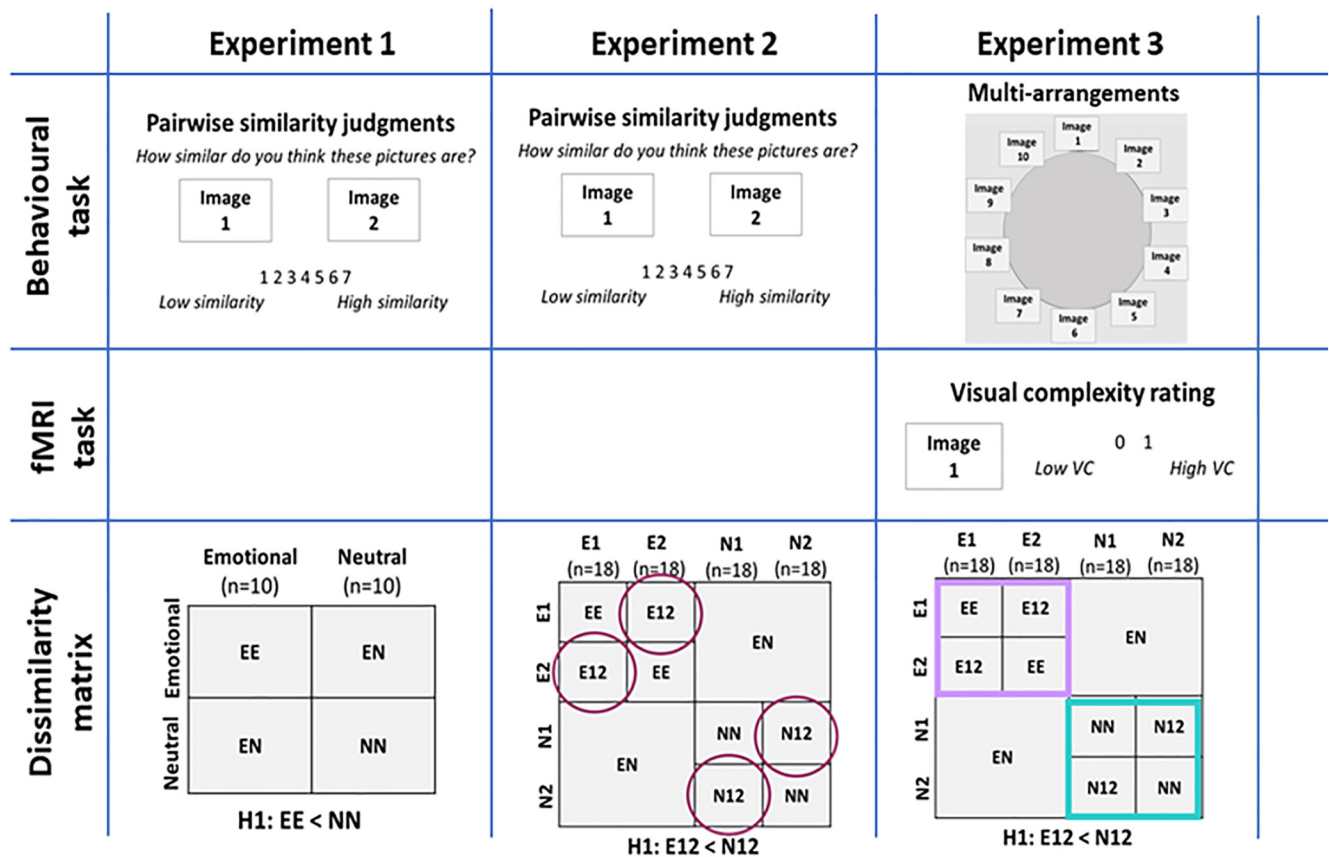


Figure 1. Graphical representation of the experimental procedure. In experiments 1 and 2, participants performed the same behavioral task. They were presented with a pair of pictures and rated their similarity on a 7-point scale (low to high similarity). In experiment 1, participants judged all the possible combinations from the first database, which consisted of 20 complex pictures (10 emotional and 10 neutral) selected from the NAPS. We expected as the main finding lower dissimilarity (higher similarity) between EE and NN pictures. In experiment 2, participants judged the similarity between emotional and neutral pictures from the second database. It consisted of 72 pictures from four semantic categories (18 pictures in each category), two emotional (E1 and E2) and two neutral (N1 and N2). Participants only rated E12, N12, and a few EN pairs only; E12 represented the similarity between E1 and E2, N12 between N1 and N2, and EN between emotional and neutral pictures. We expected lower dissimilarity (higher similarity) in the former. In both experiments 1 and 2, EN comparisons served as manipulation checks. The same database was used in experiment 3, wherein participants first judged the subjective visual complexity of each picture during an fMRI scan and then judged the similarity among all the pictures by arranging them in a circular arena. We tested the same hypothesis as in experiment 2, and extended it also to the neural data. The violet square in the dissimilarity matrix represents the emotional similarity space and the green one the neutral similarity space.

started with a central fixation cross for 500 ms. Then, participants viewed one of the images presented in the center of the screen and rated each pictures on two 9-point scales (valence scale: 1, negative emotions; 9, positive emotions; 5, neutrals; arousal scale: 1, relaxed; 9, aroused; 5 neutral). We instructed participants to respond as quickly as possible by clicking the appropriate number key and informed them that there was no right or wrong answer. Pictures from the first dataset were rated by

participants in experiment 1 before commencing that experiment, and the ratings of pictures from the second dataset were completed by a separate group of participants.

Behavioral measures of similarity

The data from the behavioral experiments were used as measures of perceived similarity, that is, similarity ratings in experiments 1–2 and

Euclidean distance in experiment 3. To make sure that the behavioral findings were independent of the specific instructions participants were given we used two separate task instructions (pairwise ratings in experiment 2, multiarrangement in experiment 3).

Pairwise similarity rating task

In experiments 1–2, participants rated the similarity of paired pictures on a 7-point scale (1 = low similarity, 7 = high similarity). In experiment 1, they rated all possible pairwise combinations (190 pairs), resulting from the database of 20 complex pictures. In experiment 2, because of the time constraint, we divided the 72 pictures into two subsets (even and odd, $n = 36$ within each subset); in addition, we focused on pairs in E12 and N12 as well as some in EN as catch trials (total pairs = 170; 81 in both E12 and N12, and 8 in EN). We chose the pairwise presentation because each pair is independently rated, and also small differences in similarity judgements can be detected, compared with a triad forced-choice similarity task, wherein only binary responses are provided (Miller, 1994; Goldstone et al., 1997). We instructed participants to base their judgment on the overall meaning of the picture, without considering any visual details (e.g., the background color, the number of people). We also informed them that there was no right or wrong answer. We purposefully did not bias them by instructing them to emphasize any dimension because we wanted our laboratory measure of behavioral similarity perception to quantify, as closely as possible, natural, holistic similarity perception outside the lab. Finally, to make sure that the behavioral findings were independent of the specific instructions, participants followed two separate task instructions (pairwise ratings in experiment 2, multiarrangement in experiment 3).

Multiarrangements task

In the validation study of the second database and in experiment 3, participants judged the similarity among all the pictures by using the MA task. We chose it because it is a quick and efficient task for acquiring similarity judgements in experiments with a relatively large number of stimuli. Kriegeskorte and Mur (2012) established the MA test-retest reliability ($r = 0.81$) as well as external validity. The task comprised different trials. In each trial, a subset of 16 stimuli was presented along the perimeter of a circle, or arena, on a computer screen. Participants had unlimited time to drag and drop the stimuli in the arena according to their similarity, so similar stimuli were placed close to each other and dissimilar stimuli apart. In other words, the distance among stimuli in the arena reflected their dissimilarity. We instructed participants to focus on the content of the pictures and to ignore visual details (e.g., the color of the background, the number of people in the scene). A trial ended when participants arranged all the stimuli in the arena. Subsequent trials started with another subset of stimuli to be arranged, selected by using the lift-the-weakest algorithm for adaptive design of item subsets. This method optimizes trial efficiency by adaptively selecting item subsets whose dissimilarity estimates presented the weakest evidence. The task ended after ~1 h, when participants judged all the possible combinations among stimuli.

MRI procedure

In experiment 3, images were acquired on a whole-body MRI scanner (Magnetom Trio, TIM, Siemens) with a 12-channel head coil. Functional images were acquired with a susceptibility weighted EPI sequence (TR/TE = 2000/30 ms, flip angle = 75°, voxel dimensions = 3 × 3 × 3.5 mm, 192 slices) in four separate scanning sessions (up to 2 min between sessions). Anatomical T1-weighted images were acquired after the functional scans (MPRAGE, TR/TI/TE = 2500/900/2.32 ms, flip angle = 8°, voxel dimensions = 1 mm isotropic, 32 slices).

As shown in Figure 1, during the fMRI scan, participants viewed the 72 complex pictures on a blank screen (size 800 × 800 pixels); we asked them to rate the visual complexity of the pictures to make them focus on the stimuli by pressing the right or the left button of the response box. Images were presented in a random order for 3 s, during which participants had to make their ratings, interleaved with a black fixation cross (mean jitter 3 s). The task was divided into four runs, during which every picture was presented once, thus resulting in four repetitions for each

picture with a total duration of ~50 min. We instructed participants that there was no right or wrong answer in the task; rather, they had to focus on their subjective perception during the ratings. To guide participants in the ratings, we suggested to them that a picture of few objects, colors, or structures would be less complex than a very colorful picture of many objects composed of several components according to Madan et al. (2018). Behavioral and fMRI tasks instructions differed, as it is not possible to measure both neural representational similarity and behavioral similarity using the same instructions. Similar procedures were also adopted in previously published papers in this research field (Kriegeskorte et al., 2008b; Chikazoe et al., 2014; Chavez and Heatherton, 2015). This is because to compute the neural representation of each picture (and then feed it into the RSA), in the MRI session we need participants to focus on one picture at a time; but behavioral measures of similarity perception requires participants to consider picture pairs.

Statistical data analysis

In the similarity judgements tasks, we expected higher similarity (lower dissimilarity) within category than between categories. We also expected higher similarity (lower dissimilarity) between emotional than between neutral conditions, as showed at the bottom of Figure 1. The first prediction serves as manipulation check because a good category boundary simultaneously maximizes the within-category similarity and minimizes the between categories similarity; the second prediction represents our main hypothesis and applies also for the neural data. In experiment 1, EN was calculated by averaging the dissimilarity between emotional and neutral pictures and the dissimilarity within emotional (EE) and within neutral (NN) categories by averaging the dissimilarity between emotional and between neutral pictures, respectively, for each participant. In experiment 3, EE represented the averaged dissimilarity within E1 and within E2, NN the averaged dissimilarity within N1 and within N2, and EN across both E1 and E2, and N1 and N2 for each participant. Finally, in experiments 2–3, E12 was measured by averaging the dissimilarity between the two emotional categories and N12 between the two neutral categories. The conditions of experiment 3 are the same in the validation study. Additional details about the statistical analyses are reported in the following sections.

Behavioral data analysis

We analyzed these data by using RSA. Specifically, in experiment 1, the similarity ratings were entered as input in a 20 × 20 similarity matrix for each participant. The rows and the columns represented the experimental stimuli, and each cell reflected the similarity rating for each pair. Then for each subject, an RDM was computed. We first standardized the similarity ratings by subtracting 1 (the lowest similarity rating) from each rating x , and then dividing by 6 (highest similarity rating minus lowest similarity rating). Second, we transformed them into correlational distances by subtracting the ratings from 1. The correlational distance ranges from 0 to 2 (0 for perfect correlation, and thus high similarity; 1 for no correlation; 2 for perfect anticorrelation) and was entered as input in each cell of the RDM. As a consequence, the RDM is symmetric about a diagonal of zeros. Next, we extracted from the single-subject RDM the mean dissimilarity and the SD of the conditions of interest as mentioned in the key hypotheses. These were entered as dependent variables in a repeated-measures ANOVA, with the conditions as grouping factor (experiment 1: EE, NN, and EN; experiment 2: E12, N12, EN). In the validation study and in experiment 3 (MA task), similarity was measured as Euclidean distance between stimuli in the arena. Specifically, at the end of each trial, a partial RDM is estimated showing the Euclidean distance between stimuli within each trial. At the end of the task, a global 72 × 72 RDM is estimated by averaging the partial RDMs with an iterative rescaling. This scaling procedure takes into account that in each trial, participants focused on a specific subset, and, therefore, there is not a permanent relationship between screen distance and dissimilarities across trials (Kriegeskorte and Mur, 2012). Then, we extracted from each participant's global RDM the mean and the SD of the conditions of interest as mentioned in the key hypotheses. These were entered as dependent variables in a repeated-measures ANOVA, which served to test

Table 4. Differences in visual complexity ratings among categories

Categories				Statistics		
E1	E2	N1	N2	<i>F</i>	<i>p</i>	η_p^2
0.308 ± 0.163	0.728 ± 0.211	0.297 ± 0.187	0.267 ± 0.170	5.34	<0.001**	0.63
<i>Post hoc</i>						
E1 vs E2	E1 vs N1	E1 vs N2	E2 vs N1	E2 vs N2	N1 vs N2	
−0.42, 0.000**	0.01, 1.00	0.04, 1.00	0.43, 0.000**	0.46, 0.000**	0.03, 1.00	

The proportion of high complexity ratings within each category (total number of high complexity responses divided by 18) was averaged across sessions. Mean and SD of each category and the statistics of the difference among them are reported at the top of the table. Bonferroni *post hoc* corrections for multiple comparisons ($p < 0.05$) are summarized at the bottom; * $p_{FWE} < 0.05$, ** $p_{FWE} < 0.001$.

lower dissimilarity in EE and NN than in E12, N12, and EN and the main hypothesis (lower dissimilarity in E12 than N12). Bonferroni *post hoc* corrections for multiple comparisons ($p < 0.05$) were used to explore the nature of the effect. The results of the validation study are shown in Table 3.

We conducted additional analyses to test differences in the variance across participants in the judgments of similarity between emotional than neutral stimuli. With this aim, we conducted two-sample *F* tests for variance, one for each contrast of interest (experiment 1: EE vs NN; experiment 2: E12 vs N12; experiment 3: EE vs NN and E12 vs N12).

Multidimensional scaling. We performed the multidimensional scaling (MDS) to visualize the structure of the similarity space, wherein proximities reflect similarities among stimuli and are measured on an ordinal scale. The rank order of proximities determines the dimensionality of the space and the metric configuration of the points representing the stimuli (Shinkareva et al., 2013). As reported in previous studies in this research field, we assumed this space to be bidimensional, with valence and arousal as orthogonal dimensions (Russell and Bullock, 1985). The goodness of fit of the MDS representation is estimated with the Stress measure.

Analysis of emotional (valence and arousal) and visual complexity ratings. Valence and arousal ratings were entered as dependent variables in two repeated-measures ANOVAs, with picture type (emotional vs neutral) as a within-group factor in experiment 1, and category as a within-group factor in experiments 2–3. Moreover, we analyzed the visual complexity ratings from the fMRI task by transforming them into a continuous variable. Specifically, for each subject we calculated the proportion of high complexity responses by dividing the number of high complexity responses within each category by 18 (the number of pictures within each category) and then averaged them across sessions. These were entered as dependent variables in a repeated-measures ANOVA, with category (i.e., E1, E2, N1, and N2) as grouping factor. The results from the valence and arousal ratings of experiment 1 are reported in Table 1, those from experiments 2–3 in Table 2. The results of the visual complexity rating task are shown in Table 4. Data analyses were conducted in MATLAB R2018a (MathWorks), and IBM SPSS Statistics for Windows, version 25.0.

Neuroimaging data analysis

Preprocessing. Neuroimaging data were preprocessed and analyzed using Statistical Parametric Mapping (SPM12; <http://store.elsevier.com/product.jsp?isbn=9780123725608>) and MATLAB R2018a (MathWorks). Functional images were slice-time corrected to reduce the mismatching between acquisition timing of different slices and realigned to a reference (mean) image to minimize the variance because of head movements. These were then coregistered to the high-resolution T1-weighted structural image, which was coregistered and normalized to Montreal Neurological Institute (MNI) space. Finally, functional images were normalized to a standard template volume based on the MNI reference brain to achieve a more precise comparison across individuals. Spatial smoothing was performed only on functional data analyzed with a conventional univariate approach using a 6 mm full-width at half-maximum isotropic Gaussian kernel. No spatial smoothing was carried over on the multivariate functional data, according to the standard practices for multivariate pattern analysis studies (Haxby et al., 2001; Kriegeskorte et al., 2008b). The preprocessing for the univariate tests was identical to the

one for the RSA with the exception of using a 6 mm FWHM Gaussian smoothing kernel (Kriegeskorte et al., 2006).

Individual-level model for RSA analysis

After preprocessing, functional data from each voxel were analyzed using the general linear model (GLM). Each stimulus was modeled as a separate event beginning with picture presentation onset, using the canonical function in SPM12 and included in the model as regressor of interest (72 regressors per session). Six motion correction parameters were also modeled within each session and included in the model as regressor of no interest. From this GLM analysis, we obtained a single β image for each stimulus. Contrast images for each stimulus against the implicit baseline were generated based on the fitted responses and averaged across sessions. The resulting 72 T-contrast images were used as inputs for RSA.

Individual-level models for univariate analyses

Although our hypotheses were specific to the multivariate representations, we also performed three conventional univariate analyses, referred to as GLM1, 2, and 3. GLM1 was performed as a manipulation check to evaluate the probability that any differences in the RSA analysis were because of differences in the average univariate activations among conditions. For this reason, GLM1 used individual-level models that were almost identical to those used for the RSA, the only difference being that instead of modeling 72 stimuli, here each category (i.e., E1, E2, N1, N2) was modeled as a separate condition (four regressors per session) beginning with each picture presentation onset, using the canonical function in SPM12.

GLMs 2–3 were performed as a second manipulation check to test whether our study replicated previous findings showing higher recruitment of emotional regions during the processing of emotional than neutral stimuli across the 4 sessions (GLM 2) and within session 1 only (GLM 3). For this reason, individual-level models were altered to be maximally sensitive to the difference between emotional and neutral stimuli. Specifically, in GLM 2–3 we included the temporal derivative to take into account temporal differences in the BOLD signal between emotional and neutral conditions (Friston et al., 1998; Calhoun et al., 2004; Heinzl et al., 2005).

Region of Interests definition

We defined the regions of interest (ROIs) by using the Automated Anatomical Labeling (AAL) template in the Wake Forest University (WFU) PickAtlas toolbox (https://www.nitrc.org/projects/wfu_pickatlas) and Anatomy toolbox (<https://www.fil.ion.ucl.ac.uk/spm/ext/#AAL>), and constructed with MarsBaR 0.43 (<http://marsbar.sourceforge.net>). We used the WFU_PickAtlas toolbox to define the bilateral early visual cortex (EVC) as Broadmann (Ba) 17, the dorsomedial prefrontal cortex corresponded to the Ba 8 and 9, the ventromedial prefrontal cortex to the Ba 10, and the dorsal and ventral ACCs to the Ba 32 and 24. The retrosplenial cortex (RSC), the occipital place area (OPA), and the parahippocampal place area (PPA) were respectively defined as follows: the bilateral RSC as Ba 29 and Ba 30, the OPA as an 8 mm sphere around the coordinates reported by Julian et al. (2016; left OPA: −34, −77, 21; right OPA: 34, −77, 21), and the PPA as an 8 mm sphere around the coordinates reported by Henson and Mouchlianitis (2007; left PPA: −27, −45, −12; right PPA 30, −42, −9). The face fusiform area (FFA) was defined as an 8 mm sphere around the coordinates reported by Henson and Mouchlianitis (2007; left FFA: −42,

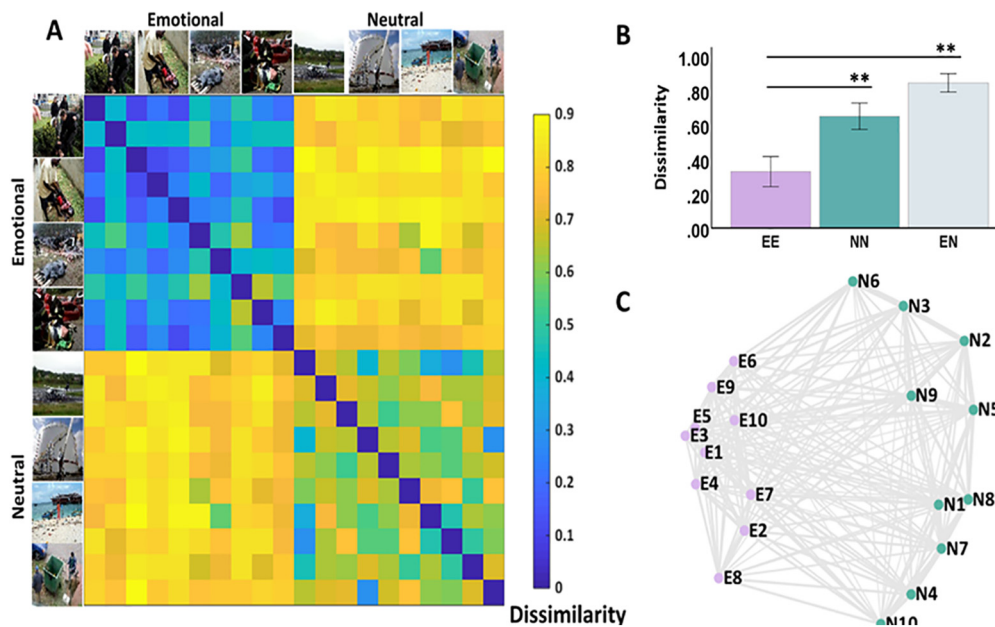


Figure 2. *A*, The RDM of 20 complex pictures (10 emotional, 10 neutral), averaged across participants. It is symmetric about a diagonal of zeros, the rows and the columns represent the stimuli, and each cell the dissimilarity, measured as 1 standardized similarity rating between stimuli within each specific pair. Yellow colors denote high dissimilarity, blue colors low dissimilarity. *B*, The average dissimilarity within emotional pictures (EE), within neutral pictures (NN), and between emotional and neutral pictures (EN, gray). Error bars represent ± 2 SEM; $**p < 0.001$. *C*, The MDS plot of the 20 pictures in a bidimensional space. Additional information supporting Figure 2 can be found in Extended Data Figure 2-1.

–51, –18; right FFA: 42, –45, –21). The medial temporal lobe comprised the entorhinal cortex defined with the Anatomy toolbox, and the bilateral hippocampus, the perirhinal cortex, and the parahippocampal cortex defined with AAL. The same toolbox was used for the bilateral inferior temporal cortex (ITC), the anterior temporal lobe, the amygdala, the thalamus, the insula, the precuneus (Prec), and the bilateral orbitofrontal cortex (OFC; superior, middle, inferior, and medial). We combined these ROIs into one ROIs mask, which was used in the searchlight RSA.

Univariate group analyses

From each individual-level GLM, we obtained a single β image for each condition. We then compared emotional and neutral conditions (emotional greater than neutral), thereby producing one contrasted image for each subject. The contrasted image from each subject was then entered as a dependent variable in a one-sample t test. Both the univariate and the multivariate results were inclusively masked to only include our ROIs involved in the visual, semantic, and emotional processing of complex pictures, as defined above in the ROIs definition.

RSA group analyses: quantifying neural similarity

Brain-behavior correlations. To test our main hypothesis (i.e., higher neural similarity between the two emotional than the two neutral categories), we first conducted a very precise localization technique, the searchlight RSA, to investigate which brain regions (within the ROIs mask) represented the participants' similarity space. This was conducted by computing the Spearman's correlation between brain activation-pattern RDMs and behavioral RDMs (second order isomorphism). The behavioral RDM represented the participants' similarity space resulted from the MA task, created as explained in the paragraph about the behavioral data analysis. Three separate analyses were conducted. The first used the entire RDM (with all the 72 stimuli, all RDM); the second focused exclusively on the emotional stimuli (36 stimuli, emotional RDM), and the third on the neutral stimuli (36 stimuli, neutral RDM), depicted as violet and green squares at the bottom of Figure 1, respectively. We conducted these latter two analyses to explore whether any brain region was involved in the representation of either the emotional or the neutral categories. For the purpose of these three analyses, three brain activation-pattern RDMs were constructed for each participant in the same way. The participant's brain activation-pattern RDMs were computed by entering the T-contrast images into a matrix with all the voxels in the

rows and the experimental stimuli in the columns. Then, for each subject and each of the three analyses, a $3 \times 3 \times 3$ voxels spherical cluster was moved throughout the brain, and at each location in the ROIs mask a correlational distance (among t values) was assigned to the center voxel of the sphere, resulting in a $(x, y, z, \text{number of pairs})$ brain activation-pattern RDM for each subject. This measure quantified the dissimilarity across voxels in a given searchlight sphere for each specific pair. The number of pairs represented all the possible combinations between experimental stimuli (2556 pairs with 72 stimuli, 630 with 36 stimuli). Next, for each stimulus, the similarity between brain and behavioral RDMs was estimated using a pairwise Spearman's correlation. This provides a correlational map between the behavioral and the brain RDMs for each subject, which reveals where the similarity space is best represented in the brain (highest correlation), and an n map, wherein the number of voxels that contributes to each correlational value is reported in each entry. The correlational coefficients were Fisher's z transformed, and inference was performed at each voxel by performing a one-side signed-rank test across subjects, testing the null hypothesis of no correlation between brain and behavior RDMs. The resulting p values (uncorrected) were thresholded to control the false-discovery rate (FDR). We performed two different FDR correction procedures to yield a more conservative as well as a more lenient set of results. In the conservative procedure, we divided the p values by the total number of voxels in the ROIs mask. In the more lenient procedure, we divided the p values by the number of voxels that contribute to each correlational value (between brain and model RDM). The number of voxels was extracted from the n map associated with the correlational map.

Differences in neural dissimilarity between emotional and neutral categories. We conducted a second set of analyses to test our main hypothesis, that is, a higher neural similarity (lower dissimilarity) between the two emotional compared with the two neutrals categories (similarity E12 > N12). We tested this effect in the brain clusters that we observed to be involved in representing both the entire (72 stimuli) and partial (36 stimuli) participants' similarity space. With this aim, we created different masks, one for each significant cluster. In the case of ROIs that were significantly correlated with both the emotional and the neutral similarity space, we selected the clusters correlated with the neutral similarity space. Then, for each subject and each mask, we computed a brain activation-pattern RDM, where each entry represented the correlational distance (1 minus Spearman's correlation) between brain activations across

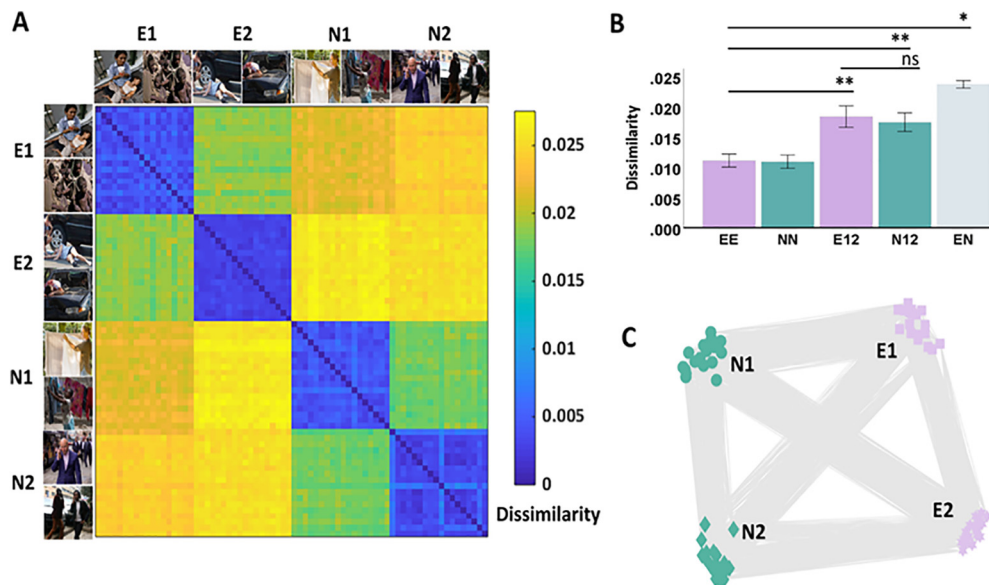


Figure 3. **A**, The RDM of 72 complex pictures (emotional categories: E1, poverty (1–18); E2, car accidents (19–36); neutral categories: N1, laundry (37–54); N2, phone call (55–72), averaged across participants. It is symmetric about a diagonal of zeros, the rows and the columns represent the stimuli, and each cell the dissimilarity (measured as Euclidean distance) between stimuli within each specific pair. Yellow colors denote high dissimilarity, blue colors low dissimilarity. **B**, The average dissimilarity within emotional pictures (averaged across E1 and E2), within neutral pictures (averaged across N1 and N2), between emotional pictures (E12), between neutral pictures (N12), and between emotional and neutral pictures (EN). Error bars indicate ± 2 SEM; *, $p_{FWE} < 0.05$; **, $p_{FWE} < 0.001$. **C**, The MDS plot of the 72 pictures in a bidimensional space. Additional information supporting Figure 3 can be found in Extended Data Figure 3-1.

voxels within that mask, and the rows and the columns represented the experimental stimuli. We refer to this as ROI RDM. It is symmetric about a diagonal of zeros and resulted in 2556 cells in the lower triangular part, which reflected the pairwise dissimilarity of the response patterns associated with the stimuli for each ROI. Then, within each participant ROI RDM, we calculated the mean of the conditions of interest (E12 and N12) and entered them as dependent variables in paired *t* tests, one for each cluster ($p < 0.05$). The RSA was performed using the MRC-CBU (Medical Research Center-Cognition and Brain Sciences) RSA toolbox for MATLAB (<http://www.mrc-cbu.cam.ac.uk/methods-and-resources/toolboxes>).

As in the behavioral experiments, we tested any differences in the variance across participants in the neural dissimilarity between E12 and N12. We explored this effect in brain clusters wherein we observed significant differences in neural dissimilarity between E12 and N12. We conducted different two-sample *F* tests for variance, one for each cluster.

Results

In a series of experiments with different datasets of real-world pictures, we explored whether emotions are associated with increased perceived similarity, both subjective (ratings) and objective (neural) similarity. We hypothesized that for both dependent measures, perceived similarity will be higher (dissimilarity lower), that is, (1) within category compared with between categories and (2) between the emotional categories compared with the neutral categories.

Behavioral evidence for increased similarity between emotional stimuli

Experiment 1 confirmed our hypotheses. We observed a significant main effect of our conditions ($F_{(2,18)} = 91.00, p < 0.001, \eta_p^2 = 0.83$), with lower dissimilarity within (i.e., EE and NN) than between (i.e., EN) categories ($p < 0.001$) and in EE than in NN ($p < 0.001$). When represented in the similarity space, emotional pictures were displaced closer to each other than neutral pictures. This resulted in a Stress value of 0.05, indicating a good fit of this model. These findings are shown in Figure 2.

Table 5. Differences in the variance in similarity judgements between emotional and neutral stimuli

Behavioral experiments	Conditions	<i>F</i> value	<i>F</i> critical	<i>p</i> value
Experiment 1	EE NN	1.28	2.53	0.36
	0.04 0.03			
Experiment 2	E12 N12	0.72	1.89	0.16
	0.05 0.07			
Experiment 3	EE NN	0.79	2.13	1.00
	0.000 0.000			
	E12 N12	1.34	2.13	0.44
	0.000 0.000			

The variance averaged across participants for each conditions, and the statistics of each difference between conditions are reported. In experiment 1, EE and NN represent the variance within emotional, and within neutral stimuli, respectively, averaged across participants. In Experiments 2–3, E12 and N12 signify the variance between the two emotional and the two neutral categories, respectively, averaged across participants. Finally, in experiment 3, EE and NN represent the variance within E1 and E2 and within N1 and N2, averaged across participants.

In experiment 2, with the second dataset, which controlled for the higher thematic similarity between emotional pictures, we observed different results. Specifically, we found lower dissimilarity in E12 and N12 compared with EN ($F_{(2,38)} = 27.40, p < 0.001, \eta_p^2 = 0.41$) but no differences in similarity ratings between the two emotional and the two neutral categories. The same results were replicated using the MA task in experiment 3. Our manipulation check revealed lower dissimilarity within category (i.e., EE and NN) than between categories (i.e., E12, N12, EN; $F_{(4,26)} = 214.76, p < 0.001, \eta_p^2 = 0.88$) but no difference because of emotion in the critical comparison between E12 and N12. In the bidimensional space, the proximities between the two emotional and between the two neutral categories do not differ. The Stress value was 0.10, indicating a fair fit of this model. This reduction in the goodness of fit compared with experiment 1 might suggest that the weight of the semantic dimension in subjective similarity may have been higher in experiments 2–3, where four categories were included, compared with experiment 1, where stimuli were not grouped by semantic category. These findings are shown in Figure 3.

Table 6. Differences in BOLD signal change between emotional and neutral categories

Analysis	p_{FWE}	K	p uncorrected	t	x	y	z	Label
GLM 2	0.001	197	<0.001	8.33	36	−46	−19	FFA R
Emotional > Neutral, all sessions	0.023	114	0.002	8.16	−39	−46	−19	FFA L
	0.039	100	0.003	6.72	−39	−82	11	OPA L
	0.003	178	<0.001	5.49	33	26	−4	Insula R
	<0.001	107	<0.001	4.89	−27	23	−4	Insula L
GLM 3	0.014	104	0.001	6.92	6	−13	−1	Thalamus R
Emotional > Neutral Session 1	0.029	88	0.002	6.24	−39	−46	−19	FFA L
	0.003	138	<0.001	5.65	39	−64	−10	Temporal inferior R/FFA R
	0.002	146	<0.001	5.58	−36	20	2	Insula L
	0.038	82	0.002	5.26	−39	−82	11	OPA L
	0.026	90	0.002	4.59	48	35	−10	OFC R

Only regions that survive correction for multiple comparisons using $p_{FWE} < 0.05$ are reported. Small volume correction using the ROI mask was applied in both analyses. R, Right; L, left.

Finally, in all the experiments, we did not observe any significant differences in the variance across participants between emotional and neutral conditions. This allows us to exclude an alternative explanation of the behavioral results; that is, that the similarity between emotional pictures can be affected by individuals' emotional granularity (Barrett et al., 2001). High-granular individuals would be more aware of the differences of their emotional experiences when viewing pictures from the two emotional categories and may rate them as less similar, whereas low-granular individuals may rate them as more similar, ultimately masking the difference between emotional and neutral categories. However, if this explanation was correct, we would expect increased variance in ratings of emotional pictures. Instead, there were no significant differences in rating variance between emotional and neutral categories. These results are reported in Table 5.

Manipulation check: Univariate differences between the emotional and neutral conditions

In GLM 1, no clusters (number of voxels > 10) survived the correction for multiple comparisons, suggesting that RSA results are unlikely to be contaminated by mean signal differences. Conversely, we replicated previous findings in GLM 2–3; these results are reported in Table 6 and Figure 4.

Brain–behavior correlations

We conducted a searchlight RSA to investigate the brain regions within the ROIs mask that represented the participants' self-reported similarity space. First, we tested whether the neural-pattern similarity within the ROIs mask was significantly correlated with the entire (72×72) similarity space, composed of neutral and emotional categories. These data only survived our more lenient correction for multiple comparisons ($p_{FDR} < 0.05$; see above, Materials and Methods). We observed that clusters in the bilateral ITC, the right FFA, and the right precuneus represented the participants' similarity space. These findings are reported in Table 7 and Figure 5A.

Second, we performed the same analysis separately for the emotional and neutral pictures to explore whether any brain region was involved in the representation of either the emotional or the neutral categories (Fig. 1, violet and green squares). The results from these analyses survived the more conservative correction for multiple comparisons ($p_{FDR} < 0.05$; see above, Materials and Methods). We found that participants' emotional similarity space was significantly correlated with clusters in lower and higher level visual processing regions, as well as regions involved in emotional processing. These included the bilateral EVC, bilateral OPA, bilateral PPA, bilateral FFA, bilateral precuneus, bilateral dorsal ACC, and in the left anterior insula (aIns).

By contrast, participants' neutral similarity space was significantly correlated with clusters in higher level visual regions only, including the bilateral OPA, the bilateral PPA, and left FFA. These findings are reported in Table 7 and Figures 6A and 7A.

Neural evidence for increased similarity between emotional stimuli

We performed the ROIs RSA to explore whether the neural representations of emotional categories are more similar than those associated with neutral categories. This analysis was conducted in brain clusters from the above analysis, namely, those that significantly correlated with the whole participants' similarity space (Fig. 5B), as well as with its emotional (Fig. 6B), and neutral (Fig. 7B) similarity spaces. As predicted, the neural pattern dissimilarity of emotional categories was lower than the one of neutral stimuli in all the previously reported clusters ($p < 0.05$), apart for the right PPA. In addition, we observed trends toward significance in support of our hypothesis in the right EVC ($p = 0.11$) and in one cluster in the left PPA ($p = 0.06$). These findings are reported in Table 8 and in Figures 5B, 6B, and 7B.

Finally, we did not observe any significant differences in the variance across participants between E12 and N12 in any brain clusters. These results are reported in Table 9.

Discussion

We investigated behavioral and neural similarity measures between complex emotional and neutral stimuli using two similarity judgment tasks and two stimulus databases, the second of which was very tightly controlled. We report two novel findings. First, the similarity between neural representations of stimuli from two negatively valenced, emotionally arousing categories was greater than the neural similarity between stimuli from two neutral categories. This increase was observed while participants were processing individual stimuli rather than interstimulus relationships. Some, but not all, of the clusters expressing similarity among emotional stimuli preferentially also expressed similarity among neutral stimuli. Second, once semantic similarity was controlled, participants rated the similarity of stimuli from two emotional categories to be equivalent to that of stimuli from two neutral categories. Thus, the greater neural similarity between emotional pictures did not influence perceived similarity in the same participants. We discuss the implications of these results below.

Increased neural similarity between emotional than neutral realistic events

In experiment 3, we observed increased neural similarity between emotional than neutral categories, in that in brain

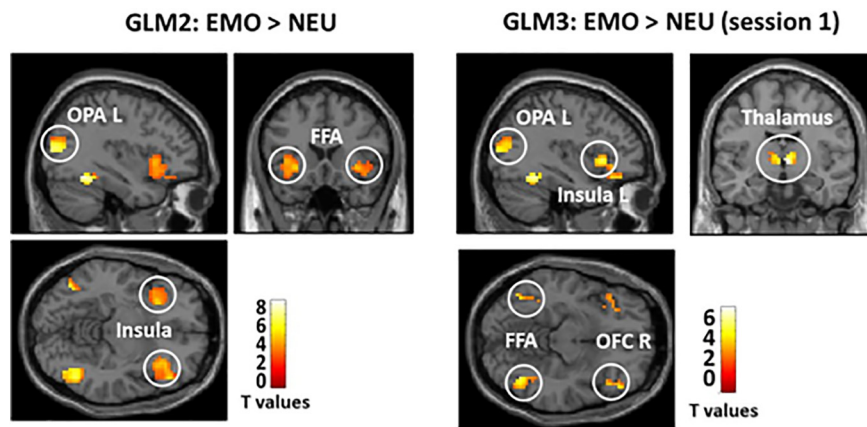


Figure 4. Differences in BOLD signal change between emotional and neutral categories, across four sessions (GLM 2, left) and in session 1 only (GLM 3, right). Only regions that survive correction for multiple comparisons using $p_{FWE} < 0.05$ are reported. Small volume correction using the ROI mask was applied in both analyses.

Table 7. Brain–behavior correlations

Analysis	Regions	x	y	z	n voxels	t	p_{FDR}	d
All RDM (72 × 72)	ITC L	11	−61	−9	154	38.81	<0.001**	3
	ITC R	46	−63	−9	21	40.5	<0.001**	10
	FFA R	42	−47	−19	69	29.64	<0.001**	4
	Prec R	3	−57	20	32	13.11	<0.001**	3
Emotional RDM (36 × 36)	EVC L	−13	−86	4	72	15.1	<0.001**	1.82
	EVC R	6	−93	−1	21	16.17	<0.001**	4.28
	OPA L	−34	−80	18	103	117.25	<0.001**	11.43
	OPA R	34	−80	18	103	74.97	<0.001**	7
	PPA L	−21	−43	−10	32	58.57	<0.001**	11.25
	PPA R	24	−43	−16	29	89.19	<0.001**	16
	FFA L	−36	−49	−22	41	25.38	<0.001**	5
	FFA R	39	−49	−22	34	26.18	<0.001**	4
	Prec	−2	−60	40	1249	109.58	<0.001**	4
	dACC L	−3	27	1	18	27.79	<0.001**	2.5
	dACC R	3	15	42	20	29.18	<0.001**	7.5
Neutral RDM (36 × 36)	aINS L	−35	27	1	18	27.79	<0.001**	7.5
	OPA L	−29	−78	20	11	20.44	<0.001**	6.67
	OPA R	36	−79	16	43	57.2	<0.001**	10
	PPA L	−30	−49	−10	17	18.7	<0.001**	4
	PPA R	25	−46	−13	10	7.19	<0.001**	4.28
	FFA L	−42	−52	−13	10	7.19	<0.001**	2.25

Top, correlations between the entire (72 × 72) stimulus space (i.e., all RDM) and the brain. Significant correlations were observed in the bilateral ITC, right FFA, and the right prec. Middle, correlations between the emotional (36 × 36) similarity space (i.e., emotional RDM) and the brain. Significant correlations were observed in the bilateral EVC, OPA, PPA, FFA, Prec, dACC, and left aINS. Bottom, correlations between the neutral (36 × 36) similarity space (i.e., neutral RDM) and the brain. Significant correlations were observed in the bilateral OPA, PPA, and left FFA. In all these analyses, correlational coefficients were Fisher’s z transformed and entered as dependent variables in a one-side t test (separately for each brain region), testing the null hypothesis of no correlation between the participants’ similarity space and the neural activation patterns. The resulting p values were thresholded to control for the FDR; ** $p_{FDR} < 0.001$. R, Right; L, left.

clusters involved in encoding participants’ entire similarity space, the neural similarity between emotional categories was stronger than between neutral categories. These clusters were located in the ventral visual stream, which underpins semantic categorization (Clarke and Tyler, 2014), and in regions involved in affect representations (e.g., precuneus; Kim et al., 2017) and modulation [e.g., dorsal anterior cingulate cortex (dACC); Saarimäki et al., 2018]. To our knowledge, this is the first report of the neural underpinnings of perceived similarity between complex emotional stimuli, while using a pictures set controlled for visual and semantic attributes.

This finding has implications for research about the neurobiological correlates of categorization and generalization.

Previous studies (Visser et al., 2011; Dunsmoor et al., 2014) observed increased neural similarity among exemplars that predicted threat. They proposed that this mechanism was adaptive, enabling individuals to differentiate emotionally salient stimuli from others, and supporting broad generalization between items that predict fitness-relevant outcomes. Although our work differed from these studies, where the emotional response was induced through pavlovian conditioning, we found the same effect here. This converging evidence suggests that it is evolutionarily more important to integrate emotional information into neural representations to increase the relevance and generalizability of stimuli that predict a negative outcome. These findings concur with the conclusions that emotion serves as a fundamental

feature of cognition, in that any representation of the world is an integrated product among emotion, perception, and thought (e.g., “That is a good thing.”) rather than discrete and isolated psychological events (e.g., “That is a thing. I feel good.”; Todd et al., 2020).

We extended previous findings about brain regions involved in representing emotional categories and dimensions by exploring, for the first time, differences in the neural representations of the relationships between emotional and neutral stimuli. The bilateral ITC, right FFA, and right precuneus represented the entire similarity space and exhibited greater neural similarity between the two emotional than the two neutral categories. As part of the hierarchical network in the ventral visual stream, the ITC integrates relevant low- and high-level features, resulting in an emergent category structure (Prince and Konkle, 2020). Accumulating research agrees on the inferior occipitotemporal regions as the potential neurobiological underpinnings of semantic categorization of objects (Jordan et al., 2015), faces (Guntupalli et al., 2017), and places (Epstein and Baker, 2019). Other regions in the ITC involved in action observation and in representing acting bodies, including FFA, take part to scenes encoding (Groen et al., 2018). Accordingly, Brooks et al. (2019) demonstrated that subjects’ conceptual space predicts the neural pattern activation in the right FFA (Brooks et al., 2019). We may have observed stronger neural similarity between emotional categories in these regions because of the influence of the precuneus, which is involved in valence representation and structurally connected with the ITC (Lin et al., 2020).

When we investigated the emotional and the neutral parts of participants’ similarity space, we observed higher emotional similarity in the EVC, OPA, and PPA, as well as in the dACC and anterior insula. OPA and PPA relate low-level visual features encoded in the EVC with the high-level aspects of the scene (Epstein and Baker, 2019) and may be modulated by regions that are sensitive to salience (anterior insula, dACC; Lindquist et al., 2012), resulting in higher similarity. Interestingly, our finding that the insula represented the emotional, but not neutral, similarity space replicate those of Levine et al. (2018), who reported that it represented similarity ratings among emotional stimuli, although they were not controlled for semantic similarity, and the ratings were of emotional rather than overall similarity. It would be worth exploring whether we would replicate the same

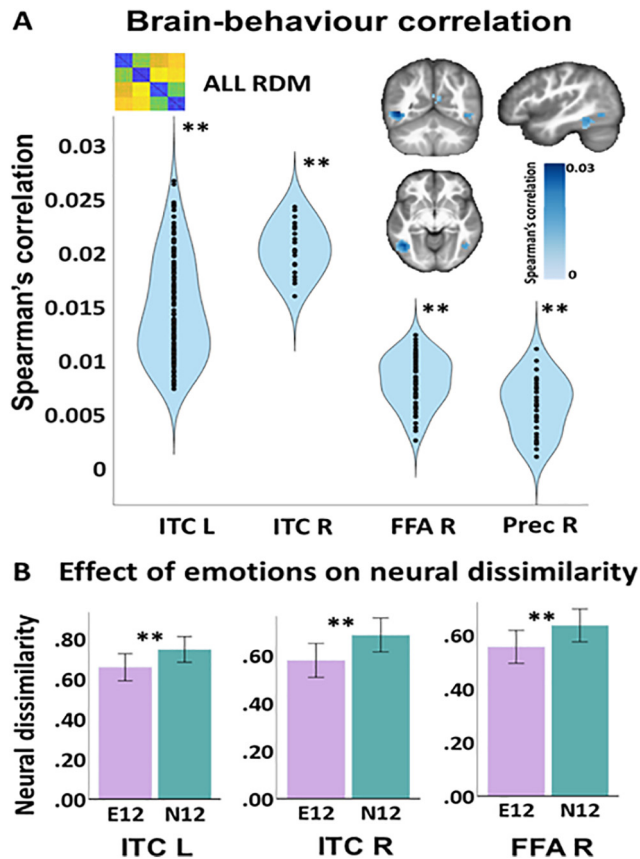


Figure 5. *A*, Correlation between the entire (72×72) stimulus space (i.e., all RDM) and the brain. Significant correlations were observed between the behavioral all RDM and clusters in the bilateral ITC, right FFA, and the right Prec. Correlational coefficients were Fisher's z transformed and entered as dependent variables in a one-side t test (separately for each brain region), testing the null hypothesis of no correlation between the participants' similarity space and the neural activation patterns. The resulting p values were thresholded to control for the FDR; $**p_{\text{FDR}} < 0.001$. *B*, Differences in neural dissimilarity (measured as correlational distance) between emotional and neutral stimuli in different brain clusters, including the bilateral ITC and the right FFA. The dissimilarity between emotional categories (E12) was calculated by averaging the dissimilarity between E1 and E2 and the dissimilarity between neutral categories (N12) by averaging the dissimilarity between N1 and N2 for each participant. These were entered as dependent variables in paired t tests, one for each brain cluster ($p < 0.05$); $**p < 0.001$. L, Left; R, right.

results using Levine et al. (2018) instructions (Riberto et al., 2020).

Finally, the same effect was observed in the EVC, which relies on more fine-grained representations of the stimuli (Coutanche et al., 2016) and encodes low-level visual features of the stimuli that afford the decoding of a broad range of emotion categories (Barrett and Bar, 2009). Specific combinations of low-level features (e.g., luminance) along with high-level information (e.g., presence of faces or scenes) can act as cues and afford specific categories of emotional response (Kragel et al., 2019). This might be paralleled by neural synchronization, which connects the different neuronal populations involved in the processing of each feature (e.g., low- and high-level visual, emotional features) with the distant brain networks involved in each feature during the emotional experience (Sander et al., 2018).

We also expected that the orbitofrontal, ventral, and dorso-medial prefrontal cortices were involved in representing the similarity space or just the emotional part. However, we did not find significant correlations with the behavioral data there, perhaps because of the implicit processing of affect in experiment 3. Nor

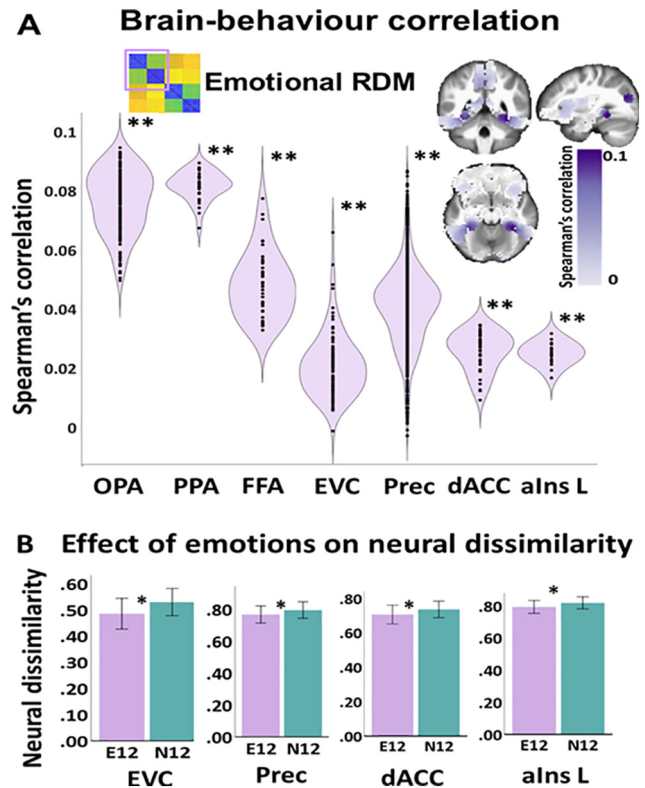


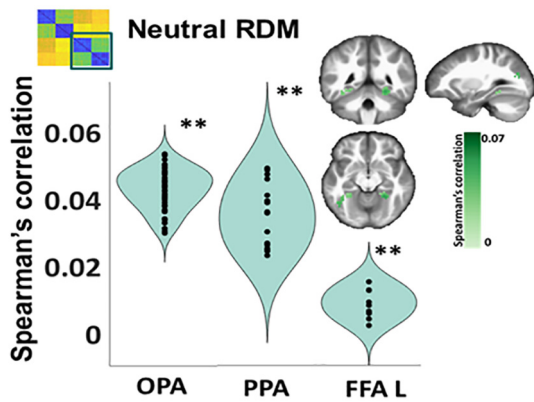
Figure 6. *A*, Correlation between the emotional (36×36) similarity space (i.e., emotional RDM) and the brain. Significant correlations were observed between the behavioral emotional RDM and clusters in the bilateral OPA, PPA, FFA, EVC, Prec, dACC, and left alns. Correlational coefficients were Fisher's z transformed and entered as dependent variables in a one-side t test (separately for each brain region). For simplicity, we averaged the left and the right sides of the clusters wherein both sides were significant. The resulting p values were thresholded to control the FDR; $**p_{\text{FDR}} < 0.001$. *B*, Differences in neural dissimilarity (measured as correlational distance) between emotional and neutral stimuli in different brain clusters, including the bilateral EVC, Prec, dACC, and left alns. The dissimilarity between emotional categories (E12) was calculated by averaging the dissimilarity between E1 and E2 and the dissimilarity between neutral categories (N12) by averaging the dissimilarity between N1 and N2 for each participant. These were entered as dependent variables in paired t tests, one for each brain cluster ($p < 0.05$); $*p < 0.05$. L, left.

did we observe correlations with the amygdala, perhaps because it habituates quickly to repeated stimuli (Plichta et al., 2014).

No differences in perceived similarity between emotional and neutral pictures of realistic events

In experiment 1, when thematic similarity was not controlled, we found higher similarity between emotional than neutral stimuli. In a valence-arousal space, emotional stimuli were placed closer to each other than neutral ones. The goodness of fit suggested that affective features were the most salient in similarity judgments. This result is in keeping with dimensional perspectives on emotions (Barrett and Russell, 1999) and recent empirical data (Cowen and Keltner, 2017), although our data cannot distinguish between effects based on valence or arousal dimensions. Strikingly, when we controlled for the higher thematic relatedness between emotional stimuli by selecting stimuli from separate semantic categories, the rated similarity between stimuli from the two emotional categories was equivalent to that between those from the two neutral categories. Ratings clustered according to the four categories and the goodness of fit dropped to fair, suggesting that the semantic meaning of each picture—not negative emotion—was the most relevant feature.

A Brain-behaviour correlation



B Effect of emotions on neural dissimilarity

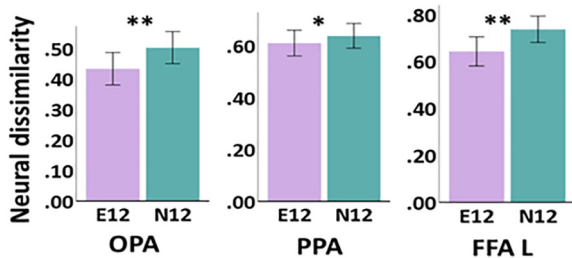


Figure 7. *A*, Correlation between the neutral (36 × 36) similarity space (i.e., neutral RDM) and the brain. Significant correlations were observed between the behavioral neutral RDM and clusters in the bilateral OPA, PPA, and left FFA. Correlational coefficients were Fisher’s z transformed and entered as dependent variables in a one-side *t* test (separately for each brain region). For simplicity, we averaged the left and the right sides of the clusters when both sides were significant. The resulting *p* values were thresholded to control the FDR; **p*_{FDR} < 0.05, ***p*_{FDR} < 0.001. *B*, Differences in neural dissimilarity (measured as correlational distance) between emotional and neutral stimuli in different brain clusters, including the bilateral OPA, PPA, and left FFA. The dissimilarity between emotional categories (E12) was calculated by averaging the dissimilarity between E1 and E2, and the dissimilarity between neutral categories (N12) by averaging the dissimilarity between N1 and N2 for each participant. These were entered as dependent variables in paired *t* tests, one for each brain cluster (*p* < 0.05); **p* < 0.05, ***p* < 0.001. L, Left.

These findings accord with claims that participants’ conceptual workspace comprises integrated perceptual, affective, and semantic dimensions (Prince and Konkle, 2020). The evolved sensitivity to emotion, evident in the neural data, may be dampened when the context suggests it is less relevant, which is in keeping with previous literature attesting to the strong context effects on similarity (Goldstone et al., 1997). The relative contribution of semantic and affective features to overall similarity could be tested in future by collecting separate ratings of semantic or emotional similarity or by manipulating the weight of the semantic and emotional dimensions. This opens up a new direction in semantic cognition research, which so far has not considered affective dimensions as key to semantic categorization (Lambon Ralph, 2014).

Limitations

Our study presents several limitations that can be addressed in future works. First, we studied only negative emotions and only two categories within each level of affect. Second, stimuli were presented during a rapid-event-related design. Although this is a common approach, it might have influenced our results by

Table 8. Effect of emotions on neural dissimilarity

Analysis	ROIs	E12	N12	<i>t</i>	<i>p</i>	<i>d</i>
All RDM (72 × 72)	ITC L	0.66 ± 0.18	0.75 ± 0.17	−7.90	<0.001**	−0.51
	ITC R	0.58 ± 0.19	0.69 ± 0.19	−10.88	<0.001**	−0.58
	FFA R	0.56 ± 0.17	0.64 ± 0.17	−8.89	<0.001**	−0.47
	Prec R	0.80 ± 0.10	0.83 ± 0.09	−3.60	0.001*	−0.32
	EVC L	0.57 ± 0.19	0.64 ± 0.17	−6.20	<0.001**	−0.39
Emotional RDM (36 × 36)	EVC R	0.49 ± 0.16	0.53 ± 0.14	−1.65	0.110	−0.27
	OPA L	0.49 ± 0.18	0.55 ± 0.17	−4.69	<0.001**	−0.34
	OPA R	0.42 ± 0.17	0.48 ± 0.17	−5.36	<0.001**	−0.35
	PPA L	0.83 ± 0.09	0.85 ± 0.04	−1.93	0.064	−0.31
	PPA R	0.64 ± 0.14	0.64 ± 0.14	−0.64	0.523	0.00
	FFA L	0.64 ± 0.17	0.74 ± 0.15	−11.34	<0.001**	−0.63
	FFA R	0.55 ± 0.17	0.63 ± 0.16	−8.88	<0.001**	−0.48
	Prec	0.77 ± 0.15	0.80 ± 0.14	−3.56	0.001*	−0.21
	dACC L	0.68 ± 0.15	0.71 ± 0.14	−3.19	0.003*	−0.21
	dACC R	0.73 ± 0.16	0.76 ± 0.14	−2.40	0.023*	−0.20
	aINS L	0.79 ± 0.11	0.82 ± 0.10	−2.85	0.008*	−0.29
Neutral RDM (36 × 36)	OPA L	0.47 ± 0.17	0.54 ± 0.16	−7.25	<0.001**	−0.42
	OPA R	0.40 ± 0.15	0.46 ± 0.16	−5.86	<0.001**	−0.39
	PPA L	0.55 ± 0.13	0.58 ± 0.14	−2.79	0.009*	−0.22
	PPA R	0.47 ± 0.13	0.47 ± 0.14	−0.73	0.473	0.00
	FFA L	0.64 ± 0.17	0.77 ± 0.15	−11.34	<0.001**	−0.81

Difference in neural dissimilarity (measured as correlational distance) among conditions. The dissimilarity between emotional categories (E12) was calculated by averaging the dissimilarity between E1 and E2, and the dissimilarity between neutral categories (N12) by averaging the dissimilarity between N1 and N2 for each participant. These measures were first computed in brain clusters significantly involved in the representation of the whole (72 stimuli) participants’ similarity space (top of the table). Then, we computed E12 and N12 in brain clusters significantly involved in the representation of the emotional (middle of the table) and neutral (bottom of the table) participants’ similarity space. We entered them as dependent variables in paired *t* tests, one for each brain cluster. Bonferroni *post hoc* corrections for multiple comparisons (*p* < 0.05) are summarized at the bottom; **p*_{FWE} < 0.05, ***p*_{FWE} < 0.001. AI, Anterior insula; R, Right; L, left.

Table 9. Differences across participants in the variance in neural dissimilarity between emotional and neutral stimuli

ROIs	E12	N12	<i>F</i> value	<i>F</i> critical	<i>p</i> value
ITC	0.03	0.03	1.04	2.13	0.45
Precuneus	0.02	0.02	1.09	2.13	0.82
EVC	0.02	0.02	1.27	2.13	0.52
OPA	0.02	0.02	1.02	2.13	0.97
PPA	0.02	0.02	0.99	2.13	1.00
FFA	0.02	0.02	1.07	2.13	0.85
dACC	0.02	0.02	1.28	2.13	0.52
aIns L	0.01	0.01	1.15	2.13	0.71

The variance averaged across participants for each condition within each cluster and the statistics of each difference between conditions are shown. E12 and N12 represent the variance between the two emotional and the two neutral categories, respectively, averaged across participants. For simplicity, we averaged the left and the right sides of the clusters. L, left.

increasing across-trial correlations (Visser et al., 2016), decreasing our statistical power. Finally, we cannot infer any causal role of emotions on neural similarity. Future studies could use transcranial magnetic stimulation to further explore this aspect of the findings.

Conclusion

Stimuli that evoke negative feelings are perceived as more similar to each other unless care is taken to eliminate their taxonomic and thematic links. Once such semantic links are controlled, negative emotional and neutral stimuli may be judged as equally similar. A set of brain regions beyond those that are functionally specific to affect expressed emotional similarity preferentially. The stronger neural similarity between emotional pictures did not influence explicitly perceived similarity in the same

participants in the immediately proceeding behavioral rating task, perhaps because the weights of the multiple dimensions of participants' conceptual workspace can change dynamically. Our findings may illuminate the clinically relevant overgeneralization bias in anxiety disorders. People with anxiety may have increased propensity to consider later, emotionally similar experiences as globally similar to the original fearful one and thereby make maladaptive choices.

References

- Barrett LF, Bar M (2009) See it with feeling: affective predictions during object perception. *Philos Trans R Soc Lond B Biol Sci* 364:1325–1334.
- Barrett LF, Russell JA (1999) The structure of current affect: controversies and emerging consensus. *Curr Dir Psychol Sci* 8:10–14.
- Barrett LF, Gross J, Christensen TC, Benvenuto M (2001) Knowing what you're feeling and knowing what to do about it: mapping the relation between emotion differentiation and emotion regulation. *Cognition* 15:713–724.
- Bex PJ, Makous W (2002) Spatial frequency, phase, and the contrast of natural images. *J Opt Soc Am A Opt Image Sci Vis* 19:1096–1106.
- Brooks JA, Chikazoe J, Sadato N, Freeman JB (2019) The neural representation of facial-emotion categories reflects conceptual structure. *Proc Natl Acad Sci U S A* 116:15861–15870.
- Calhoun VD, Stevens MC, Pearlson GD, Kiehl KA (2004) fMRI analysis with the general linear model: removal of latency-induced amplitude bias by incorporation of hemodynamic derivative terms. *Neuroimage* 22:252–257.
- Chang LJ, Gianaros PJ, Manuck SB, Krishnan A, Wager TD (2015) A sensitive and specific neural signature for picture-induced negative affect. *PLoS Biol* 13:e1002180.
- Charest I, Kievit RA, Schmitz TW, Deca D, Kriegeskorte N (2014) Unique semantic space in the brain of each beholder predicts perceived similarity. *Proc Natl Acad Sci U S A* 111:14565–14570.
- Chavez RS, Heatherton TF (2015) Representational similarity of social and valence information in the medial pFC. *J Cogn Neurosci* 27:73–82.
- Chikazoe J, Lee DH, Kriegeskorte N, Anderson AK (2014) Population coding of affect across stimuli, modalities and individuals. *Nat Neurosci* 17:1114–1122.
- Clarke A, Tyler LK (2014) Object-specific semantic coding in human perirhinal cortex. *J Neurosci* 34:4766–4775.
- Coutanche MN, Solomon SH, Thompson-Schill SL (2016) A meta-analysis of fMRI decoding: quantifying influences on human visual population codes. *Neuropsychologia* 82:134–141.
- Cowen AS, Keltner D (2017) Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proc Natl Acad Sci U S A* 114: E7900–E7909.
- Donderi DC (2006) Visual complexity: a review. *Psychol Bull* 132:73–97.
- Dunsmoor JE, Kragel PA, Martin A, LaBar KS (2014) Aversive learning modulates cortical representations of object categories. *Cereb Cortex* 24:2859–2872.
- Epstein R. A. and C. I. Baker (2019) Scene perception in the human brain. *Annu Rev Vis Sci* 5:373–397.
- Friston KJ, Fletcher P, Josephs O, Holmes A, Rugg M, Turner R (1998) Event-related fMRI: characterizing differential responses. *Neuroimage* 7:30–40.
- Giordano BL, Whiting C, Kriegeskorte N, Kotz SA, Gross J, Belin P (2021) The representational dynamics of perceived voice emotions evolve from categories to dimensions. *Nat Hum Behav* 5:1203–1211.
- Goldstone RL, Medin DL, Halberstadt J (1997) Similarity in context. *Mem Cognit* 25:237–255.
- Groen II, Greene MR, Baldassano C, Fei-Fei L, Beck DM, Baker CI (2018) Distinct contributions of functional and deep neural network features to representational similarity of scenes in human brain and behavior. *Elife* 7:e32962.
- Guntupalli JS, Wheeler KG, Gobbini MI (2017) Disentangling the representation of identity from head view along the human face processing pathway. *Cereb Cortex* 27:46–53.
- Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293:2425–2430.
- Heinzel A, Bermpohl F, Niese R, Pfennig A, Pascual-Leone A, Schlaug G, Northoff G (2005) How do we modulate our emotions? Parametric fMRI reveals cortical midline structures as regions specifically involved in the processing of emotional valences. *Brain Res Cogn Brain Res* 25:348–358.
- Henson RN, Mouchlianitis E (2007) Effect of spatial attention on stimulus-specific haemodynamic repetition effects. *Neuroimage* 35:1317–1329.
- Hoemann K, Xu F, Barrett LF (2019) Emotion words, emotion concepts, and emotional development in children: a constructionist hypothesis. *Dev Psychol* 55:1830–1849.
- Jordan MC, Greene MR, Beck DM, Fei-Fei L (2015) Basic level category structure emerges gradually across human ventral visual cortex. *J Cogn Neurosci* 27:1427–1446.
- Jordan MC, Ellis C, Osherson D, Cohen J (2017) The relative contribution of features and dimensions to semantic similarity. *J Vis* 17:1245.
- Julian JB, Ryan J, Hamilton RH, Epstein RA (2016) The occipital place area is causally involved in representing environmental boundaries during navigation. *Curr Biol* 26:1104–1109.
- Kim J, Shinkareva SV, Wedell DH (2017) Representations of modality-general valence for videos and music derived from fMRI data. *Neuroimage* 148:42–54.
- Koch A, Alves H, Krüger T, Unkelbach C (2016) A general valence asymmetry in similarity: good is more alike than bad. *J Exp Psychol Learn Mem Cogn* 42:1171–1192.
- Kragel PA, Reddan MC, LaBar KS, Wager TD (2019) Emotion schemas are embedded in the human visual system. *Sci Adv* 5:eaaw4358.
- Kriegeskorte N, Mur M (2012) Inverse MDS: inferring dissimilarity structure from multiple item arrangements. *Front Psychol* 3:245.
- Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. *Proc Natl Acad Sci U S A* 103:3863–3868.
- Kriegeskorte N, Mur M, Bandettini PA (2008a) Representational similarity analysis-connecting the branches of systems neuroscience. *Front Syst Neurosci* 2:4.
- Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini PA (2008b) Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* 60:1126–1141.
- Lambon Ralph MA (2014) Neurocognitive insights on conceptual knowledge and its breakdown. *Phil Trans R Soc B* 369:20120392.
- Lang PJ, Bradley MM, Cuthbert BN (2008) International affective picture system (IAPS): affective ratings of pictures and instruction manual. Gainesville, FL: NIMH, Center for the Study of Emotion & Attention.
- Laufer O, Paz R (2012) Monetary loss alters perceptual thresholds and compromises future decisions via amygdala and prefrontal networks. *J Neurosci* 32:6304–6311.
- Laufer O, Israeli D, Paz R (2016) Behavioral and neural mechanisms of overgeneralization in anxiety. *Curr Biol* 26:713–722.
- Levine SM, Wackerle A, Rupprecht R, Schwarzbach JV (2018) The neural representation of an individualized relational affective space. *Neuropsychologia* 120:35–42.
- Lin Y-H, Young IM, Conner AK, Glenn CA, Chakraborty AR, Nix CE, Bai MY, Dhanaraj V, Fonseka RD, Hormovas J, Tanglay O, Briggs RG, Sughrue ME (2020) Anatomy and white matter connections of the inferior temporal gyrus. *World Neurosurg* 143:e656–e666.
- Lindquist KA, Wager TD, Kober H, Bliss-Moreau E, Barrett LF (2012) The brain basis of emotion: a meta-analytic review. *Behav Brain Sci* 35:121–143.
- Madan CR, Bayer J, Gamer M, Lonsdorf TB, Sommer T (2018) Visual complexity and affect: ratings reflect more than meets the eye. *Front Psychol* 8:2368.
- Marchewka A, Zurawski Ł, Jednoróg K, Grabowska A (2014) The Nencki Affective Picture System (NAPS): introduction to a novel, standardized, wide-range, high-quality, realistic picture database. *Behav Res Methods* 46:596–610.
- Miller GA (1994) The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol Rev* 101:343–352.
- Plichta MM, Grimm O, Morgen K, Mier D, Sauer C, Haddad L, Tost H, Esslinger C, Kirsch P, Schwarz AJ, Meyer-Lindenberg A (2014) Amygdala habituation: a reliable fMRI phenotype. *Neuroimage* 103:383–390.
- Prince JS, Konkle T (2020) Computational evidence for integrated rather than specialized feature tuning in category-selective regions. *J Vis* 20:1577.

- Puccetti NA, Schaefer SM, Van Reekum CM, Ong AD, Almeida DM, Ryff CD, Davidson RJ, Heller AS (2021) Linking amygdala persistence to real-world emotional experience and psychological well-being. *J Neurosci* 41:3721–3730.
- Riberto M, Pobric G, Talmi D (2019) The emotional facet of subjective and neural indices of similarity. *Brain Topogr* 32:956–964.
- Riberto M, Pobric G, Talmi D (2020) A response to “Investigating emotional similarity: a comment on Riberto, Pobric and Talmi (2019).” *Brain Topogr* 33:288–288.
- Riegel M, Żurawski Ł, Wierzba M, Moslehi A, Klocek Ł, Horvat M, Grabowska A, Michałowski J, Jednoróg K, Marchewka A (2016) Characterization of the Nencki Affective Picture System by discrete emotional categories (NAPS BE). *Behav Res Methods* 48:600–612.
- Russell JA (1980) A circumplex model of affect. *J of personality and social psychology* 39:1161–1178.
- Russell JA, Bullock M (1985) Multidimensional scaling of emotional facial expressions: similarity from preschoolers to adults. *J Pers Soc Psychol* 48:1290–1298.
- Saarimäki H, Ejtehadian LF, Glerean E, Jääskeläinen IP, Vuilleumier P, Sams M, Nummenmaa L (2018) Distributed affective space represents multiple emotion categories across the human brain. *Soc Cogn Affect Neurosci* 13:471–482.
- Sander D, Grandjean D, Scherer KR (2018) An appraisal-driven component approach to the emotional brain. *Emot Rev* 10:219–231.
- Shinkareva SV, Wang J, Wedell DH (2013) Examining similarity structure: multidimensional scaling and related approaches in neuroimaging. *Comput Math Methods Med* 2013:796183.
- Talmi D (2013) Enhanced emotional memory: cognitive and neural mechanisms. *Curr Dir Psychol Sci* 22:430–436.
- Todd RM, Schmitz TW, Susskind J, Anderson AK (2013) Shared neural substrates of emotionally enhanced perceptual and mnemonic vividness. *Front Behav Neurosci* 7:40.
- Todd RM, Miskovic V, Chikazoe J, Anderson AK (2020) Emotional objectivity: neural representations of emotions and their interaction with cognition. *Annu Rev Psychol* 71:25–48.
- Visser RM, Scholte HS, Kindt M (2011) Associative learning increases trial-by-trial similarity of BOLD-MRI patterns. *J Neurosci* 31:12021–12028.
- Visser RM, Scholte HS, Beemsterboer T, Kindt M (2013) Neural pattern similarity predicts long-term fear memory. *Nat Neurosci* 16:388–390.
- Visser RM, de Haan MI, Beemsterboer T, Haver P, Kindt M, Scholte HS (2016) Quantifying learning-dependent changes in the brain: single-trial multivoxel pattern analysis requires slow event-related fMRI. *Psychophysiology* 53:1117–1127.
- Wagner IC, Rütgen M, Lamm C (2020) Pattern similarity and connectivity of hippocampal-neocortical regions support empathy for pain. *Soc Cogn Affect Neurosci* 15:273–284.
- Wierzba M, Riegel M, Pucz A, Leśniewska Z, Dragan WŁ, Gola M, Jednoróg K, Marchewka A (2015) Erotic subset for the Nencki Affective Picture System (NAPS ERO): cross-sexual comparison study. *Front Psychol* 6:1336.
- Zeithamova D, Gelman BD, Frank L, Preston AR (2018) Abstract representation of prospective reward in the hippocampus. *J Neurosci* 38:10093–10101.