nature human behaviour

ARTICLES

https://doi.org/10.1038/s41562-017-0234-y

Machine learning of neural representations of suicide and emotion concepts identifies suicidal youth

Marcel Adam Just[1]*, Lisa Pan[2], Vladimir L. Cherkassky[1], Dana L. McMakin[3], Christine Cha[4], Matthew K. Nock[5] and David Brent[2]

The clinical assessment of suicidal risk would be substantially complemented by a biologically based measure that assesses alterations in the neural representations of concepts related to death and life in people who engage in suicidal ideation. This study used machine-learning algorithms (Gaussian Naive Bayes) to identify such individuals (17 suicidal ideators versus 17 controls) with high (91%) accuracy, based on their altered functional magnetic resonance imaging neural signatures of death-related and life-related concepts. The most discriminating concepts were 'death', 'cruelty', 'trouble', 'carefree', 'good' and 'praise'. A similar classification accurately (94%) discriminated nine suicidal ideators who had made a suicide attempt from eight who had not. Moreover, a major facet of the concept alterations was the evoked emotion, whose neural signature served as an alternative basis for accurate (85%) group classification. This study establishes a biological, neurocognitive basis for altered concept representations in participants with suicidal ideation, which enables highly accurate group membership classification.

The assessment of suicide risk is among the most challenging problems facing mental health clinicians, as suicide is the second-leading cause of death among young adults[1]. Furthermore, predictions by both clinicians and patients of future suicide risk have been shown to be relatively poor predictors of future suicide attempt[2,3]. In addition, suicidal patients may disguise their suicidal intent as part of their suicidal planning or to avoid more restrictive care. Nearly 80% of patients who die by suicide deny suicidal ideation in their last contact with a mental healthcare professional[4]. This status identifies a compelling need to develop markers of suicide risk that do not rely on self-report. Biologically based markers of altered conceptual representations have the potential to complement and improve the accuracy of clinical risk assessment[5,6].

In this study, we offer an approach for the assessment of suicide risk that uses machine-learning detection of neural signatures of concepts that have been altered in suicidal individuals. This approach capitalizes on recent advances in cognitive neuroscience that use machine-learning techniques to identify individual concepts from their functional magnetic resonance imaging (fMRI) signatures[7–9]. These fMRI signatures are common and reproducible across neurotypical individuals. Moreover, the signatures can be decomposed into meaningful components. For example, the concept of 'spoon' includes a neural representation of the way it is manipulated (located in motor-related regions), as well as its role in eating (which is represented in gustatory areas, such as the insula and the inferior frontal gyrus)[7]. By contrast, 'house' is represented in regions related to shelter and physical setting or location (the parahippocampal and parietal areas)[7]. This approach has previously been used to detect altered representations in a special population, enabling the discrimination between 17 participants with high-functioning autism and 17 matched neurotypical individuals with 97% accuracy, based on their neural representations of 16 social interactions (such as to hate or hug)[10].

The current study applies this approach to determine whether the neural representations of positive, negative and suicide-related concepts are altered in a group of participants with suicidal ideation, relative to a control group. If so, are the alterations sufficiently systematic to enable an individual participant to be accurately classified as a suicidal ideator versus a neurotypical control participant? The study also investigates whether there is a classifiable difference among participants with suicidal ideation between those who have attempted suicide and those who have not. Furthermore, the neural signature of the test concepts was treated as a decomposable biomarker of thought processes that can be used to pinpoint particular components of the alteration. This decomposition attempts to specify a particular component of the neural signature that is altered, namely, the emotional component (described in more detail below).

Two lines of evidence within the suicide literature motivate the application of this approach to suicidal individuals. First, suicidal patients have demonstrated sensitivity to distinct concept alterations through their performance on behavioural measures. One of these measures is an adapted Emotional Stroop Task that assesses reaction times in response to suicide-related words relative to neutral words[11]; another measure is an adapted Implicit Association Test that assesses reaction times in response to pairing suicide-related words and self-related words[3]. These studies indicate that people with a history of suicide attempts may represent certain concepts or concept pairs differently than non-attempters. Neural markers of these behavioural patterns have never been tested.

Building on these previous studies, the current investigation uses machine-learning multivoxel analysis, which seeks a pattern of activation values (in a set of voxels distributed across a set of

[1]Department of Psychology, Carnegie Mellon University, Pittsburgh, PA, USA. [2]Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA. [3]Department of Psychology, Florida International University, Miami, FL, USA. [4]Clinical Psychology Department, Columbia University, New York, NY, USA. [5]Department of Psychology, Harvard University, Cambridge, MA, USA. *e-mail: just@cmu.edu

brain locations) that is associated with individual stimulus concepts, and can identify an individual as suicidal or not.

Beyond detecting altered neural signatures of concepts, in the present study we also aimed to detect the emotion component of the neural signatures. To detect these emotion components, we drew on an archive of previously acquired identifiable neural signatures from neurotypical participants[8]. The archive contains nine different types of emotion such as 'sadness' or 'shame'. In the analysis of the current study, we searched for the presence of four of the archived emotion signatures that have previously been detected among suicidal individuals[12–18]: 'sadness', 'shame', 'anger' and 'pride'. We hypothesized that the groups would differ in the degree of presence of these emotion signatures in the neural representations of concepts such as 'death'. We assume that the quality of the emotions is similar between neurotypical and suicidal participants (for example, 'anger', when it occurs, is similar). The ability to classify individual participants with respect to suicidal risk and to relate their altered activation patterns to altered emotional content associated with specific concepts would provide an interpretable, personalized profile for diagnosis and therapy.

In summary, we test three main hypotheses:

(1) Participants with suicidal ideation will differ from non-suicidal control participants with regard to their neural representations of death-related and suicide-related concepts, to a degree that a machine-learning classifier can accurately determine whether a participant is a member of the suicidal ideator group or the control group.

(2) A similar machine-learning approach will accurately discriminate those members of the suicidal ideator group who have attempted suicide from those who have not.

(3) The neural signatures of discriminating concepts in suicidal ideators will contain different emotion component signatures (that is, have different regression weights in a linear model) than the control group, and these group differences will enable a machine-learning classifier to accurately determine whether a participant is a member of the suicidal ideator group or the control group.

Results

The main neurosemantic analyses were performed on two groups of participants: 17 suicidal ideators and 17 healthy controls. The groups were balanced on sex ratio, age, and Wechsler Abbreviated Scale of Intelligence (WASI IQ) (Table 1). The stimuli were 30 concepts (as shown in Table 2) that were each presented for 3 s, and were related to either suicide, positive affect or negative affect. The brain locations that contain the main components of the neural representations of the 30 concepts, identified by the presence of stable voxels (those whose responses to the set of stimuli were similar over multiple presentations), are shown in Fig. 1 (see Methods). Six of the concepts and five of the brain locations (Fig. 2) provided the most accurate discrimination between the two groups.

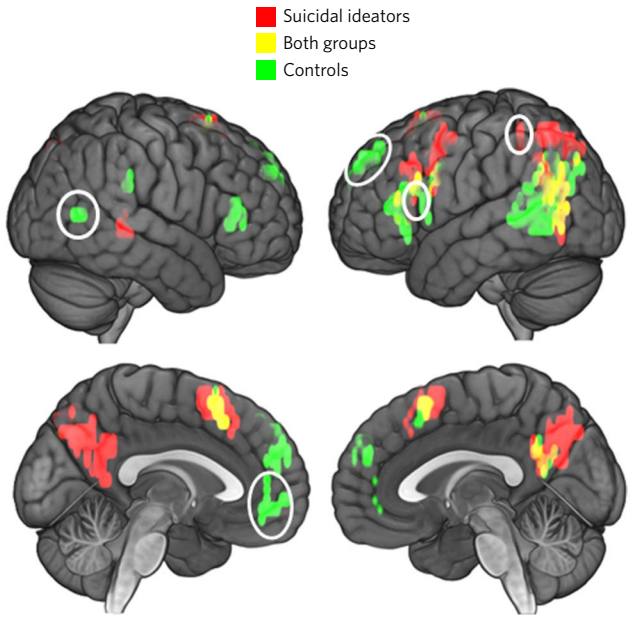

Fig. 1 | Clusters of stable voxels of the suicidal ideator group and the control group. White ellipses indicate the five discriminating locations.

Interpretable, clinically meaningful differences existed between the individuals in the suicidal ideator and control groups, and within the suicidal ideator group, there were differences between the attempters and the non-attempters. The classification procedures identified the concepts and brain locations that were most predictive of the group membership for these two sets of contrasts (that is, suicidal ideator versus control, and attempter ideator versus non-attempter ideator).

Table 1 | Demographic information and clinical variables

Measure	Participants		Test statistic (d.f.)	*P* value
	Suicidal ideators ($n=17$)	Controls ($n=17$)		
Sex ratio (male:female)	5:12	3:14	$\chi^2(1)=0.63$	0.42
Mean age	22.88 (3.57)	22.06 (2.84)	$t(32)=0.74$	0.46
WASI IQ	124.1 (10.86)	121.12 (9.70)	$t(32)=0.82$	0.420
ASIQ	57.88 (34.38)	2.76 (6.35)	$t(32)=6.5$	0.000
PHQ-9	12.24 (6.7)	0.47 (1.1)	$t(32)=7.14$	0.000
Spielberger/Anxiety State	40.12 (6.14)	46.88 (4.77)	$t(32)=3.59$	0.001
Spielberger/Anxiety Trait	47.59 (4.14)	45.88 (3.22)	$t(32)=1.34$	0.19
CTQ	41.3 (9.65)	30.24 (8.11)	$t(32)=3.62$	0.001
ASR internalizing problems	35.6 (11.9)	5.9 (5.0)	$t(32)=9.46$	0.000
ASR externalizing problems	13.9 (9.8)	4.8 (3.5)	$t(32)=3.60$	0.001
ASR total problems	83.1 (27.09)	19.65 (12.65)	$t(32)=8.74$	0.000
Number of attempts	1.41 (2.0)			
Suicide Ideation Scale	8.19 (9.06)			

Standard deviations are shown in parentheses.

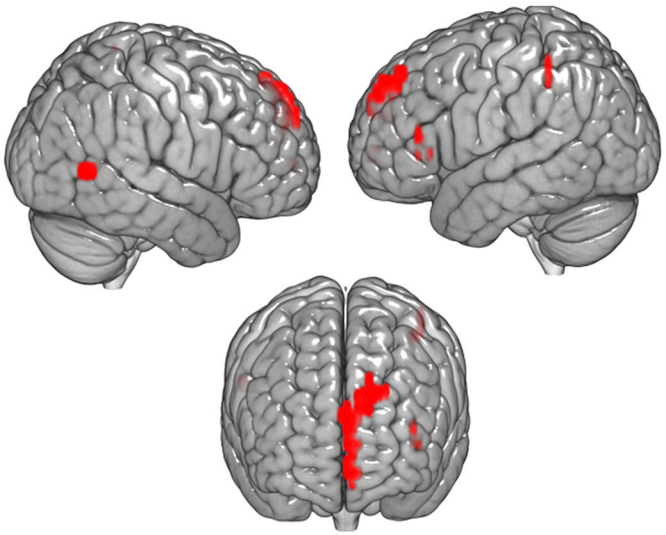

Fig. 2 | Discriminating brain locations for distinguishing between suicidal ideator and control group membership.

Table 2 | Stimulus concepts

Suicide	Positive	Negative
Apathy	Bliss	Boredom
Death	Carefree	Criticism
Desperate	Comfort	Cruelty
Distressed	Excellent	Evil
Fatal	Good	Gloom
Funeral	Innocent	Guilty
Hopeless	Kindness	Inferior
Lifeless	Praise	Terrible
Overdose	Superior	Trouble
Suicide	Vitality	Worried

Neurosemantic classification of suicidal ideator versus control group. A Gaussian Naive Bayes (GNB) classifier trained on the data of 33 out of 34 participants predicted the group membership of the remaining participant with a high accuracy of 0.91 ($P<0.000001$), correctly identifying 15 of the 17 suicidal participants and 16 of the 17 controls (sensitivity = 0.88, specificity = 0.94, positive predictive value (PPV) = 0.94, negative predictive value (NPV) = 0.89).

The features of the classifier were the neural representations of the six most discriminating concepts (as described in more detail in Methods). The neural representation of each concept, as used by the classifier, consisted of the mean activation level of the five most stable voxels in each of the five most discriminating locations.

The concepts that most strongly discriminated between the groups were 'death', 'cruelty', 'trouble', 'carefree,' 'good' and 'praise'. The most discriminating brain regions included the left superior medial frontal area, medial frontal/anterior cingulate, right middle temporal area, left inferior parietal area and the left inferior frontal area (Fig. 2 and Table 3). All of these regions, especially the left superior medial frontal area and medial frontal/anterior cingulate, have repeatedly been strongly associated with self-referential thought (which is consistent with the behavioural findings in suicidal patients reported in[3]). The separation between the ideator and control groups in the multidimensional scaling of the activation features used by the classifier is shown in Fig. 3. The distributions of the activation levels in two locations for the 17 ideator participants and 17 controls for the concepts 'death' and 'good' are shown in Supplementary Fig. 1.

To determine how many and which concepts were most discriminating between ideators and controls, a reiterative procedure analogous to stepwise regression was used, which found the next most discriminating concept at each step. The procedure is further described in Supplementary Information.

This procedure identified 'death' as the most discriminating single concept. The concepts that followed in descending order of discriminating ability were 'carefree', 'good' and 'cruelty', followed by 'praise' and 'trouble'. To determine how many and which brain locations were most discriminating between the ideators and controls, a similar stepwise procedure was performed.

Because the ideator and control groups differed with respect to other measures besides suicidal ideation, it is useful to demonstrate that the high classification accuracy remains intact after statistically controlling for such differences (namely, differences in Spielberger Anxiety/State, Patient Health Questionnaire (PHQ-9), Childhood Trauma Questionnaire (CTQ), and Adult Self-Report (ASR)). When these differences were statistically controlled for (using methods described in the literature[19,20] — see Supplementary Information for details), the classification accuracy slightly increased (from 0.91 to 0.94) (sensitivity = 0.88, specificity = 1, PPV = 1, NPV = 0.94), indicating the applicability of the model to groups that differ with respect to these clinical variables beyond suicidal ideation.

An additional quantitative assessment of the generalizability of the model applied a more conservative cross-validation technique. Instead of training the model on data from all but one participant, this additional assessment left out the data of half of the participants (8 of 17) from each group for testing, and the model was trained on the data of the remaining 9 participants. (Because there are a huge number of ways to leave out half of the participants from each group, 1,000 random selections of such partitionings were performed and the outcomes were averaged.) The classification accuracy remained at a highly reliable level of 0.76, showing that a model based on a much smaller sample of the participants generalizes to the remaining sample, which establishes an added test of the generalizability of the model.

Neurosemantic classification of suicidal ideators who have made an attempt versus ideators who have not. Another classifier was able to distinguish, within the group of 17 suicidal ideator participants, those who had previously made an attempt (9 participants) from those who had not (8 participants). This classification resulted in a high accuracy of 0.94 (16 out of 17 correct, 1 non-attempter misclassified, $P<0.0002$, sensitivity = 1, specificity = 0.88, PPV = 0.90, NPV = 1). The concepts that best discriminated between attempters and non-attempters were 'death', 'lifeless' and 'carefree'. The most discriminating brain regions for this classification were a subset of the regions that discriminated ideators from controls, namely, the left superior medial frontal area, medial frontal/anterior cingulate and the right middle temporal area. The most discriminating concepts and locations were obtained using the same stepwise reiterative procedure (described in Supplementary Information) that was used in the ideator–control classification. The separation between the attempter and non-attempter groups in the multidimensional scaling of the activation features used by the classifier is shown in Fig. 4. The distributions of the activation levels in two locations for the nine ideators with a suicide attempt and the eight ideators without such an attempt for the concepts 'death' and 'lifeless' are shown in Supplementary Fig. 2.

Alterations in the emotional content of the neural representations of the discriminating concepts. Neurosemantic signature measures are interpretable activation patterns that contain information

about the thought processes to which they correspond. This makes it possible to analyse the psychological nature of an alteration of a given concept in a clinical population. In the case of suicidal ideation, we postulated that the emotional content of the neural representations of the discriminating words would differentiate between the suicidal ideator and control groups, consistent with previous behavioural findings[11].

In the analysis of the current results, we searched for the presence of four previously acquired emotion signatures ('sadness', 'shame', 'anger' and 'pride')[8] within the neural representations of the six concepts that best discriminated between the ideator and control groups. Only four of the nine emotions for which signatures existed were used because a model with all nine emotions (that is, 'sadness', 'shame', 'anger', 'pride', 'disgust', 'envy', 'fear', 'lust' and 'happiness') would overfit the data (activation levels in five of the most discriminating locations). This particular set of four emotions (that is, 'sadness', 'shame', 'anger' and 'pride') were chosen as it resulted in the highest classification accuracy of the two groups. Furthermore, most of these four emotions have been implicated as precursors and motives for suicidal behaviour. Interpersonal discord (that is, 'anger') and embarrassment are two prominent motivations for adolescent suicide attempts[21]. 'Shame' is prominent in studies of male suicide attempters[22]. In a content analysis of more than 1,200 suicide notes, 'sadness' (for example, 'hopelessness' and 'sorrow'), 'anger' (for example, 'anger' and 'blame') and 'guilt' were particularly prominent; although, positive emotions that expressed relief either on the part of the suicide victim or on the intended recipient of the note were common[23]. However, note that, here, our neurosemantic tests probe for the emotional content in the representation of particular concepts (such as 'death'), not for an enduring emotional trait.

The neurosemantic signature of each of the six discriminating concepts was modelled as a linear combination of 'sadness', 'shame', 'anger' and 'pride', with the expectation that there would be group differences in the regression weights of the emotions. Consistent with this expectation, in the suicidal ideator group, the concept of 'death' reliably ($t(32)=2.67$, $P<0.012$) evoked more (that is, had a higher regression weight for) shame, whereas the concept of 'trouble' evoked reliably more 'sadness' in this group compared with the control group ($t(32)=2.24$, $P<0.032$). (These t tests are uncorrected for multiple comparisons, to provide an initial overview.) 'Trouble' also evoked reliably less 'anger' ($t(32)=2.78$, $P<0.01$) and 'carefree' evoked less 'pride' ($t(32)=2.96$, $P<0.006$) in the suicide ideator group. In general, the negatively valenced discriminating concepts evoked more 'sadness' and 'shame' but less 'anger' in the suicidal ideator group than in the control group.

In ideators who had made an attempt, the suicide-related concept 'death' evoked reliably less 'sadness' ($t(15)=2.91$, $P<0.01$) than in those who had not made an attempt, and the other suicide-related concept 'lifeless' evoked reliably more 'anger' ($t(15)=3.58$, $P<0.003$) than in those ideators who had not made an attempt. Furthermore, in the ideators who had made an attempt, the positive concept 'carefree' evoked reliably less 'anger' ($t(15)=2.34$, $P<0.03$) than in non-attempters.

These results are generally consistent with previous fMRI findings of altered emotion processing at the neural level (in response to face stimuli) in suicidal participants[24]. To further systematically assess the emotion signature group differences, the emotion signature weights were used as features of a classifier that attempted to identify group membership.

Identification of group membership on the basis of emotion signature differences in the distinguishing concepts. We investigated whether the emotional content of the neural signature of a concept could indicate whether a given participant was an ideator or a control participant, or, within ideators, whether they had made an attempt. The features that were used in this classification were the regression coefficients in the model previously discussed, indicating the degree of presence of each of the emotion signatures in their neural representation of each discriminating concept (for example, how much 'shame' was present in a participant's neural representation of 'death').

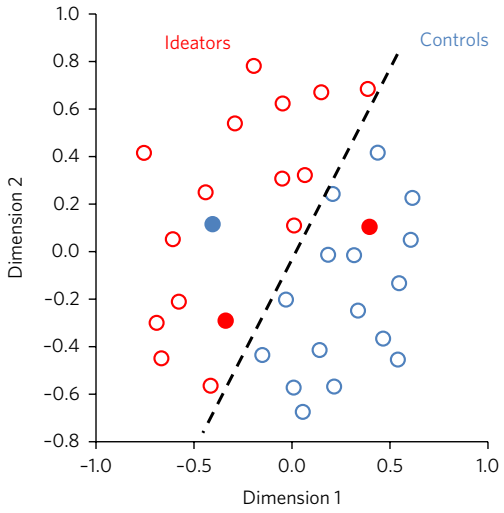

Fig. 3 | Group separation in the multidimensional scaling of the activation features of the participants used by the classifier. Ideators ($n=17$) are indicated by red circles and controls ($n=17$) by blue circles. Filled circles indicate misclassifications. The scaled features (activation levels in five brain locations for six discriminating words) were computed in 32 cross-validation folds, averaged across the folds. The dashed line shows the separability of the two groups in this two-dimensional space.

The GNB classifier correctly identified the group membership (ideator or control) of the 34 participants with 0.85 accuracy (14 ideators and 15 controls correctly identified, sensitivity = 0.82, specificity = 0.88, PPV = 0.88, NPV = 0.83). (Using the regression weights of only two of the emotions ('pride' and 'shame') resulted in the same classification accuracy (0.85) as using all four emotions.) The distributions of emotion regression weights of 'sadness' and 'shame' in the representations of 'death' and 'good' for the 17 ideator participants and the 17 controls are shown in Supplementary Fig. 3.

The same approach of using emotion regression coefficients as features was applied to distinguish the nine ideators who had made an attempts versus the eight ideators who had not made an attempt in the set of 17 ideators. Using the regression coefficients of the emotions of the three concepts that best discriminated attempters from non-attempters ('death', 'lifeless' and 'carefree') as classifier features, it was possible to identify the group membership of the 17 participants as attempters or non-attempters with 0.88 accuracy (eight attempters and seven non-attempters were correctly identified, sensitivity = 0.89, specificity = 0.88, PPV = 0.89, NPV = 0.88). As in the classification above, it was possible to achieve comparable accuracy using only a subset of the predictor variables.

Thus, the alterations of the neural signatures of the discriminating concepts in the ideator group and within the group (the attempter subgroup) can be meaningfully attributed in large part to their evoking of a different profile of specific emotions than in the comparison group. These two classification accuracies based on the emotion signature weights (0.85 and 0.88) were only slightly lower than the classification accuracies directly based on the activation data (0.91 and 0.94). This result indicates that the emotional content is an important way in which concepts are altered in suicidality and in suicidality after attempt, and therefore provides potential targets for therapy.

Table 3 | Cluster locations that are predictive for suicidal ideator and control group membership classification

Brain region	MNI centroid coordinates			Radius (mm)
	x	*y*	*z*	
Suicidal ideator group				
Left inferior parietal	−42	−43	50	5.0
Left inferior frontal gyrus — pars triangularis	−42	29	8	5.1
Control group				
Left superior medial frontal	−11	52	33	10.5
Medial frontal/anterior cingulate	−6	50	−3	8.3
Right middle temporal	56	−62	10	2.5

Correlations between neural alterations of concept representations and self-report measures of suicidal ideation. The degree of neural alteration of concepts in individual suicidal ideators can be quantitatively assessed and related to the self-reported measure of suicidal ideation. Here, the neural representation for each suicidal ideator participant was the vector of activation levels for the six most distinguishing concepts in the three most distinguishing brain regions (namely, the control group locations shown in Table 3). The neurotypical norm to which this measure was compared was the mean of the corresponding vectors averaged across the control participants. The measure of alteration for each suicidal ideator was the distance from this norm (computed as one minus the correlation between the control group mean vector and the vector of the suicidal ideator participant). There was a marginally reliable correlation ($r=0.48$, $P<0.051$) between the degree of concept alteration and the log-transformed self-reported Adult Suicidal Ideation Questionnaire (ASIQ) measure of suicidality, as shown in Supplementary Fig. 4.

Locations of the neural representations (clusters of stable voxels) for the two groups. There was a substantial similarity in neural representation of 30 concepts between the two groups in terms of the involved brain locations, with one large exception. Only the control group had clusters of stable voxels (that is, voxels that have a similar semantic tuning curve across the 30 stimulus concepts in each of the multiple presentations of the stimulus set) in the anterior frontal regions, namely, the superior medial frontal and anterior cingulate areas, whereas the ideator group showed negligible stable activation in these frontal regions, as shown in Fig. 1. By contrast, the ideator group had more clusters of stable voxels in the left inferior parietal region. These distinguishing brain locations have a substantial role in discriminating between the ideator and control participants based on the neural activation evoked by the discriminating concepts. Notably, the accuracy of identifying which of the 30 stimulus items that the participant was thinking about based on its fMRI signature was similar for the two groups: 0.71 and 0.75 for the suicidal ideator and control groups, respectively.

General linear modelling (GLM) univariate analyses of the same groups of participants (17 ideators and 17 controls) as in the main classification failed to show false-discovery rate-corrected or family-wise-corrected significance between groups in the activation patterns for all 30 concepts considered together, nor for various subsets of the concepts, such as the six discriminating concepts, nor for any of the three categories of concepts. By contrast, the multivoxel analyses of the patterns that correspond to individual concepts as described above provided excellent group separability.

Testing the classification algorithm on another sample. The data of 21 additional ideator participants, although excluded from the main analyses because of the lower technical quality of their data, were nevertheless available to use as a test of the generalization of the classifier to another sample. The data quality was measured in terms of the low accuracy of classification of the 30 stimulus items (rank accuracy < 0.60) and the generally greater head-motion parameters (mean maximum = 1.81 mm) than the 17 participants in the main study (mean = 1.27 mm, $t(77)=2.73$, $P<0.01$). Nevertheless, the classifier developed from the first set of 17 ideators and 17 controls was used, without any modifications, to try to distinguish these 21 suicidal ideators from the 17 control participants with good data quality. As in the main classification, the features of the classifier were the neural representations of the six most discriminating concepts. The neural representation of each concept comprised the mean activation level of the five most stable voxels in each of the five most discriminating locations. The resulting classification accuracy was 0.87 ($P<0.000002$, sensitivity = 0.81, specificity = 0.94, PPV = 0.94, NPV = 0.8), replicating the findings from the main analysis. Although high-quality data from both the ideator group and the control group may be necessary for model development, once a model is developed, it can accurately classify suicidal participants with lower data quality. Thus, the findings were replicated on a second sample of ideators, supporting the generalizability of the model.

The model also did reasonably well in identifying concept alterations that were associated with having made an attempt within the excluded 21 suicidal ideators. Those participants who had made an attempt versus those who had not were correctly classified with an accuracy of 0.61 ($P<0.04$, 13 out of 21 participants were correctly classified).

These results indicate that the models developed on the basis of the data of participants with less noise in their data can be successfully applied to participants with more-noisy data. However, a model that is developed from the data of either ideator or control participants with noisy data does not discriminate groups well. We attribute such noise to an inability to rigorously sustain attention to the task and to maintain head position in the scanner. The implication is that high-quality data from both the ideator group and the control group are necessary for model development, but once a model is developed, it can achieve accurate identification of suicidal ideator participants with lower data quality.

By testing the performance of the neurosemantic classifier on the additional larger sample of independent ideator participants beyond those who provided the data for the classification algorithm, we provide a replication within this study, thus strengthening support for the generalizability of the model, which applies to all of the recruited participants.

Discussion

The findings from this study provide a biological foundation for altered concept representations in those with suicidal thoughts and recent suicidal behaviour. The differences in the neural representations of concepts enable accurate classification of suicidal ideator versus control group membership, as well as suicidal ideator versus suicide attempter — the latter distinction being one that few risk factors are able to make[17]. These two findings show that suicidal ideation and attempt are associated with measurable alterations in the way a person thinks about death, suicide, and other positive and negative concepts. The recently developed fMRI methods for measuring the neural representation of a concept makes it possible to compare neurotypical to clinical representations of concepts, and within a clinical population, to compare suicidal ideation with and without suicidal behaviour.

The specific concepts that were altered in people with suicidal ideation — 'death', 'cruelty', 'trouble', 'carefree', 'good' and 'praise' — include items from all three stimulus categories: one that is

suicide-related, two that are negative, and three positive concepts. The valuation of what is important and good in life and what is not seems to be altered in ideators. Our results provide a neurally based, quantitative measure of this alteration.

Most of the ideators showed high levels of self-reported depression that is characterized by the 'cognitive triad', which includes a negative view of self, the world and the future[25]. Pessimism about the future, or hopelessness, has been shown to be correlated with and predictive of future suicidal behaviour above and beyond depression[26,27]. The observed alterations of specific concepts may be reflecting more general cognitive changes of this type.

The differences in the emotion signature components of the altered concepts provide additional information about the nature of the perspective change. As described above, the concept of 'death' evoked more shame, whereas the concept of 'trouble' evoked more sadness in the suicidal ideator group. 'Trouble' also evoked less anger in the suicidal ideator group than in the control group. The positive concept 'carefree' evoked less pride in the suicidal ideator group. This pattern of differences in emotional response suggests that the altered perspective in suicidal ideation may reflect a resigned acceptance of a current or future negative state of affairs, manifested by listlessness, defeat and a degree of anhedonia (less pride evoked in the concept of 'carefree'). This type of neurally acquired information helps to characterize the disorder as well as provide specific targets for intervention.

The altered perspective seems to be even more clear in the contrast between suicidal ideators who had made an attempt and those who had not, where the most altered concepts were 'death', 'lifeless' and 'carefree', which includes two suicide-related concepts and one positive concept. The finding of a meaningful difference between ideators with and without a history of a suicide attempt is consistent with previous findings that show differential reaction times in response to suicide-related words relative to neutral words[11], and in response to the paired concepts of 'death' and 'self' versus 'life' and 'self'[3]. Furthermore, the emotion signature differences show an interpretable pattern. For example, the suicide-related concept 'death' evoked less 'sadness' in the ideators who had made an attempt than in those who had not. The two subgroups of ideators differ in their emotional response to particular concepts.

Those ideators who had made an attempt may have thought of death with less sadness than those ideators who had not, whereas the overall group of ideators experienced more shame than controls when thinking about death. It has been shown that many suicidal ideators vacillate between an attraction to life and attraction to death[28], and that having moral objections to suicide is protective against engaging in a suicidal act even with suicidal ideation[29].

We speculate that for those who are conflicted about engaging in a suicidal act, the thought of facilitating death is shameful, whereas those ideators who have made an attempt show greater attraction to and acceptance of death, and hence less sadness in thinking about it. This perspective is also consistent with decreased 'anger' associated with the concept of 'lifeless' in ideators with a history of an attempt.

Neuroimaging studies also provide evidence of emotion alteration associated with suicide risk. fMRI studies have found altered processing of angry faces in suicide attempters, and anger and hostility are strongly related to suicidal behaviour[24,30], as well as hostility being strongly predictive of suicidal behaviour[31,32].

More generally, the ability of a machine-learning classifier to make discriminations within the suicidal ideator group indicates the specificity of the neurosemantic assessment approach. The classifier is not simply detecting an abnormality that is likely to be present in many disorders, such as depression. It makes accurate discriminations within the ideator group, distinguishing

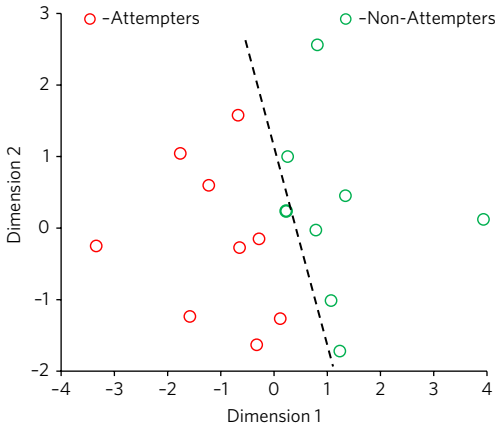

Fig. 4 | Group separation in the multidimensional scaling of the activation features of the nine ideators who have attempted suicide and the eight ideators without attempts used by the classifier. Attempters ($n=9$) are indicated by red circles and non-attempters ($n=8$) by green circles. The features (activation levels in three brain locations for three discriminating words) were scaled in two dimensions. The dashed line shows the separability of the two groups in this two-dimensional space.

those who had a previous history of a suicide attempt, and therefore are at higher risk for future suicidal behaviour. Although it is possible that these findings were due to the greater severity of suicidal ideation and depression in past attempters, the specificity of the discriminating concepts 'death' and 'suicide' suggest a possible application of the approach in the assessment of imminent suicidal risk. Moreover, we have identified differences in the emotions experienced by those ideators with and without a history of suicide attempt, such as differences in anger in thinking about death that are not likely to be explained merely by differences in depressive symptoms.

There are several types of evidence indicating that the activation pattern (neural signature) for a given emotion truly indexes that emotion. A study[8] found that, first, the emotion signatures are sufficiently specific to accurately identify which emotion was being experienced. Second, in a validation check of the emotion manipulation (an instruction to drama student participants to evoke a particular verbally named emotion such as shame), a separate condition presented images from the International Affective Picture System that depicted disgust. The classifier trained on the instruction-evoked activation patterns of the emotions correctly identified the emotion evoked by the disgust pictures with 0.91 rank accuracy, indicating the strong similarity of the disgust activation patterns evoked in two very different ways, which provides support for the construct validity of the measure. Third, the neural signatures of the emotions were similar across participants, such that a classifier trained on the emotion signatures of all but one participant could identify the emotions of the remaining participant with 0.71 rank accuracy. This finding of the commonality of emotion signatures across participants indicates the convergent validity of these neural signatures. The current study provides additional evidence for reliability and usefulness of the approach by finding that the emotion signature weights in a concept representation are features that can identify membership in the ideator group. Given the limited previous use of this potentially powerful approach to analysing emotional content from neural signatures, there should be caution concerning the inferences that can be made.

Thus the findings also enable progress beyond stating that one group is measurably different from another. They enable at least part of the difference to be attributed to the emotional component

of a concept representation. Unlike a dictionary definition of a concept, a neural representation includes the emotional response to the concept. Some concepts, such as 'snake', have long been known to entail an emotional response. The findings here show that certain concepts evoke different emotions in people with suicidal ideation compared with controls, and also evoke different emotions in suicidal ideators dependent on whether they have ever made a suicide attempt. When used as the features of a classifier, these differences in the emotional component in the neural signature of a concept can be used to provide accurate classification of group membership (in both the ideator–control classification and the attempter–non-attempter classification).

fMRI capabilities have made it possible to characterize the altered brain activity of a clinical population as having a higher or lower level of activation in a brain region (for example, the anterior cingulate) than a control group during the performance of a task. By contrast, our approach attempts to characterize a network of altered neural activity that constitutes the representation of a concept and the emotion it evokes. At a given brain location, for some concepts, the activation level is higher in the ideator group and for other concepts it is lower. The current study makes an early attempt at relating a pattern of activation values across multiple brain locations to neurotypical and altered representations of particular concepts and their emotional component in a manner that seeks consilience between brain activity and psychological states. At the same time, it remains possible to determine which brain structures are the sites of a clinical alteration.

This study is distinctive in neuroimaging research on suicidal ideation and behaviour because it directly focuses on how suicidal individuals think about various concepts, rather than on responses to tasks that, however salient, do not directly mirror the experience of the suicidal person. This neurosemantic assessment has face validity because those suicide attempters at highest risk and with the highest suicide intent engaged in suicidal ideation because they wanted to die (and therefore thought about suicide as being more attractive) or wanted to escape an impossible situation or feeling state, which might lead to altered responses to various death-related and life-related concepts.

There are several potential benefits of this neurosemantic approach. The identification of differential patterns of regional activation could suggest brain regions to target using brain stimulation techniques, such as transcranial magnetic stimulation or transcranial direct current stimulation[33]. The identification of altered emotional responses to suicide-related concepts could prove very useful to a psychotherapist in trying to improve the patient's attraction to life and decrease the attraction to suicide and death. If these findings have predictive value, then they would also be useful in guiding a clinician's decisions about psychotherapeutic targets and in monitoring overall suicidal risk. The neurosemantic approach can also guide the development of less costly and more easily disseminable methods that can potentially yield similar information, such as electroencephalogram assessment of neural concept representations, as demonstrated for neurotypical participants[34]. Despite its greater cost, this approach might also be effective in highly suicidal individuals who are repeatedly hospitalized for suicidal crises or those who require a higher level of care, such as an intensive outpatient programme.

An unexplored prospective benefit of the approach is its potential to predict imminent suicidal risk. A longitudinal investigation of a larger cohort of individuals with suicidal ideation could repeatedly assess the altered neural representations to determine whether there is a neural signature of an imminent attempt. Such information would be invaluable with respect to the small percentage(that is, 5%) of patients in psychiatric inpatient care who make up almost half of suicides subsequent to discharge from a hospital[35]. In future prospective studies, it would be of great interest to learn whether our neurosemantic assessments are useful in monitoring for current suicidal risk and in predicting future suicide attempts. If so, this approach could be useful for monitoring ongoing suicidal risk and response to treatment.

Study limitations. Performance of the task requires highly cooperative and focused participants (not everyone can keep their attention intensely focused for 30 min). However, we also showed that the models developed on the less-noisy participants' data can be successfully applied to more noisy data from other participants, which substantially improves the chances for potential clinical applications. Moreover, it may be possible in the future to develop shorter batteries that focus on concepts that are most likely to identify altered responses associated with suicidal risk and that would require sustained attention over a shorter period.

Another limitation is that the current study does not provide a contrast between suicidal ideator and psychiatric control participants who are affected by psychopathology in general. However, the ability to distinguish within the suicidal ideator group between attempters and non-attempters suggests that our classification is more specific and not just related to psychopathology in general. Within its limitations, the current study provides a first step in assessing a psychiatric disorder of the brain and mind that takes both of these facets into account.

Methods

Participants. Participants were 79 young adults who either had current suicidal ideation ($n=38$) or were healthy controls with no personal or family history of psychiatric disorder or suicide attempt ($n=41$). Exclusion criteria included neurological disorders, anoxia history, head injuries, a Wechsler verbal score of <80 (ref. [36]), current use of sedative medication, pregnancy, ineligibility for MRI, psychosis, substance misuse or positive urine drug/saliva alcohol screen.

Assessment. History of suicide attempt (defined as potentially self-injurious behaviour with some non-zero intention of dying) was assessed with the Suicide History Form and Suicide Intent Scale[37,38]. The severity of suicidal ideation was assessed using the interview-rated Columbia-Suicide Severity Rating Scale (C-SSRS)[39] and the self-reported ASIQ[40]. General psychopathology, depression, anxiety and history of child maltreatment were assessed using the ASR[41,42], the PHQ-9 (ref. [43]), the Adult Spielberger State Trait Anxiety Inventory (STAI-T)[44] and the CTQ[45], respectively.

Participants in neurosemantic analyses. The neurosemantic analyses below are based on 34 participants, 17 participants per group whose fMRI data quality was sufficient for accurate (normalized rank accuracy >0.6) identification of the 30 individual concepts from their fMRI signatures. The selection of participants included in the primary analyses was based only on the technical quality of the fMRI data. The data quality was assessed in terms of the ability of a classifier to identify which of the 30 individual concepts they were thinking about with a rank accuracy of at least 0.6, based on the neural signatures evoked by the concepts. The participants who met this criterion also showed less head motion ($t(77)=2.73$, $P<0.01$). The criterion was not based on group discriminability. The 17 participants selected for the primary data analysis and the 21 remaining suicidal participants did not differ on demographic data, diagnoses, clinical severity of depression, anxiety or suicidal ideation, or history of suicide attempt. The data of the participants with poor data quality were also analysed, as reported in the Results section.

A previous study of autism spectrum disorder using a similar approach[10] used 17 participants with good data quality per group, hence, the target of a similar sample size. Three additional control participants who had also satisfied this criterion were selected at random and excluded to equate the group sizes. The final groups were balanced on sex ratio, age, and WASI IQ. Participants in the suicidal ideator group were significantly more symptomatic than the control group on almost all other measures, as shown in Table 1. There were no systematic differences between the 17 ideators whose data were used in the neurosemantic analysis and the 21 participants whose data were excluded, other than the poor classification accuracy on the 30 concepts. We attribute the suboptimal fMRI data quality (inaccurate concept identification from its neural signature) of the excluded participants to some combination of excessive head motion and an inability to sustain attention to the task of repeatedly thinking about each stimulus concept for 3 s over a 30-min testing period. Despite their exclusion from the main neurosemantic analysis, we show that there is valuable information in the fMRI data of the excluded suicidal ideator participants. The comparison of self-report data between the 34 participants included in the neurosemantic analyses and the remaining (excluded) participants is reported in Supplementary Information.

The study protocol was approved by the University of Pittsburgh and Carnegie Mellon University Institutional Review Boards. All participants gave their informed written consent.

Stimuli. The stimuli were three groups of ten words each, half of which were nouns and half were adjectives related to: (1) suicide (for example, 'death' and 'overdose'); (2) negative affect (for example, 'sad' and 'gloom'); and (3) positive affect (for example, 'happy' and 'carefree') as shown in Table 2. The set of 30 stimulus items was presented 6 times, in different random orders. Each item was displayed for 3 s followed by a 4-second blank interval to allow for the delay in haemodynamic response. Long fixation intervals of 17 s were included periodically to provide an activation baseline. The stimuli were displayed in white font and centred on a black background.

Task instructions. Participants were asked to actively think about the concepts to which the stimulus words refer while they were displayed, thinking about their main properties (and filling in details that come to mind) and attempting consistency across presentations.

Image acquisition and preprocessing. The fMRI data were acquired on a Siemens Verio 3.0 Tesla scanner (20 slices, voxel size $3.125 \times 3.125 \times 5$ mm^3, repetition time 1 s). The data were preprocessed and converted to a standard Montreal Neurological Institute (MNI) space using SPM8 (Wellcome Department of Cognitive Neurology), and a single mean value was computed for each voxel and stimulus item (see Supplementary Information for details).

fMRI data analytic approach. Three analyses are described here: (1) selecting voxels with stable semantic tuning curves; (2) spatial clustering of the stable voxels at the group level to determine the brain locations that contain the neural representations of the concepts; and (3) developing a resulting machine-learning classification model from the reduced data and attempting to classify the group membership of participants using the model.

1. Selecting voxels with stable semantic tuning curves**.** These analyses focus on a subset of all the voxels (each ~50 mm^3) whose semantic tuning curve of activation over the set of stimulus items is stable across the multiple presentations of the set of items (see Supplementary Information for details).

2. Obtaining group-level clusters of stable voxels**.** A fixed number of the most stable voxels are selected in each participant (excluding bilateral occipital lobes), and a group hit map is computed and thresholded by the number of contributing participants and spatial proximity (see Supplementary Information for details). The clusters of stable voxels in the group hit maps indicate where the set of neural representations (including all of the concepts) are located for the two groups, as shown in Fig. 1. Preliminary testing identified which of the clusters best discriminated between the two groups (see Supplementary Information). The features of the classifer included voxels in clusters that are common between the two groups as well as voxels from unshared clusters.

3. Machine-learning methods**.** Machine learning entails training a classifier on a subset of the data and testing the classifier on an independent subset. The cross-validation procedure iterates through all possible partitionings (folds) of the data, always keeping the training and test sets separate from each other. The main machine learning here uses a GNB classifier (using pooled variance).
The main type of classifications performed in this study was a group membership classification that assigned each participant to one of the two groups: (1) the accuracy was the proportion of correctly classified participants, and significance levels were obtained using a binomial distribution, and (2) the identification of which of the 30 concepts a participant was thinking about; in this case, rank accuracy was computed (see Supplementary Information for details) and compared to a chance level of accuracy obtained by random permutation testing.

The main reason that classification was used rather than the GLM is that classification is multivariate, whereas GLM uses univariate analysis of fMRI data (assessing each voxel independently). The phenomena of interest here (and in many fMRI studies of cognition) are inherently multivariate, in the sense that such cognitively related phenomena typically occur in several different voxels or voxel clusters that do not need to be proximal to each other. In particular, the neural representations of individual concepts such as 'apple' or 'death' correspond to activation in a set of spatially distributed voxel clusters, and the groups here differ in the collective pattern of activation levels in these spatially distributed voxels. GLM, because of its univariate nature, fails to assess the collective pattern and the group differences in the collective pattern. By contrast, the features of the classifier are the set of activation levels of a set of spatially distributed voxels. Many other studies have shown greater sensitivity of classification over GLM where the phenomena of interest consist of a spatially distributed pattern of activation.

Group membership classification. Two types of group membership classification were performed: (1) suicidal ideator versus control group, consisting of 17 participants in each group, and (2) within the suicidal ideator group, attempters ($n=9$) versus non-attempters ($n=8$). Both types of classification were based on fMRI data in the sets of group-level stable clusters that were identified for both groups.

The features used by the classifier to characterize a participant consisted of a vector of activation levels for several (discriminating) concepts in a set of (discriminating) brain locations. To determine how many and which concepts were most discriminating between ideators and controls, a reiterative procedure analogous to stepwise regression was used, first finding the single most discriminating concept and then the second most discriminating concept, reiterating until the next step reduced the accuracy. A similar procedure was used to determine the most discriminating locations (clusters). The procedure is further described in Supplementary Information. The activation level in each brain location was computed as a mean activation of the five most stable voxels in that location. The classifier was trained on the data of all but one participant, and the group membership of the remaining participant was predicted.

In addition to group membership classification based on the neural representations of the stimulus concepts themselves, another classification was based on the emotional content of the neural representations of the discriminating concepts. The activation of the discriminating concepts was represented as a weighted sum of activation vectors that characterized the involvement of four emotions: 'sadness', 'shame', 'anger' and 'pride'. Each participant was characterized by a vector consisting of the weights associated with these emotions for each discriminating concept, and the group membership of participants was classified using a machine-learning procedure similar to the one described above (see Supplementary Information for details).

Classification of 30 concepts. This procedure attempted to identify which of the 30 concepts a participant was thinking about, given an independent sample of its neural signature. This measure provided an index of the inconsistency or noise level in the neural signature data from a participant.

Life Sciences Reporting Summary. Further information on experimental design and reagents is available in the Life Sciences Reporting Summary.

Code availability**.** The custom computer code that was used in the main analysis of this study is available from the corresponding author upon reasonable request.

Data availability. The de-identified data that support the main findings of this study are available from the corresponding author upon reasonable request.

Received: 6 February 2017; Accepted: 4 October 2017;
Published online: 30 October 2017

References

1. *WISQARS Data* (CDC, 2016); http://webappa.cdc.gov/sasweb/ncipc/leadcaus10_us.html.
2. Glenn, C. R. & Nock, M. K. Improving the short-term prediction of suicidal behavior. *Am. J. Prev. Med.* **47**, S176–S180 (2014).
3. Nock, M. K. et al. Measuring the suicidal mind: implicit cognition predicts suicidal behavior. *Psychol. Sci.* **21**, 511–517 (2010).
4. Busch, K. A., Fawcett, J. & Jacobs, D. G. Clinical correlates of inpatient suicide. *J. Clin. Psychiatry* **64**, 14–19 (2003).
5. Mann, J. J. et al. Candidate endophenotypes for genetic studies of suicidal behavior. *Biol. Psychiatry* **65**, 556–563 (2009).
6. Ribeiro, J. D. et al. Self-injurious thoughts and behaviors as risk factors for future suicide ideation, attempts, and death: a meta-analysis of longitudinal studies. *Psychol. Med.* **46**, 225–236 (2016).
7. Just, M. A., Cherkassky, V. L., Aryal, S. & Mitchell, T. M. A neurosemantic theory of concrete noun representation based on the underlying brain codes. *PLoS ONE* **5**, e8622 (2010).
8. Kassam, K. S., Markey, A. R., Cherkassky, V. L., Loewenstein, G. & Just, M. A. Identifying emotions on the basis of neural activation. *PLoS ONE* **8**, e66032 (2013).
9. Mitchell, T. M. et al. Predicting human brain activity associated with the meanings of nouns. *Science* **320**, 1191–1195 (2008).
10. Just, M. A., Cherkassky, V. L., Buchweitz, A., Keller, T. A. & Mitchell, T. M. Identifying autism from neural representations of social interactions: neurocognitive markers of autism. *PLoS ONE* **9**, e113879 (2014).
11. Cha, C. B., Najmi, S., Park, J. M., Finn, C. T. & Nock, M. K. Attentional bias toward suicide-related stimuli predicts suicidal behavior. *J. Abnorm. Psychol.* **119**,616–622 (2010).
12. Armey, M. F., Crowther, J. H. & Miller, I. W. Changes in ecological momentary assessment reported affect associated with episodes of nonsuicidal self-injury. *Behav. Ther.* **42**, 579–588 (2011).
13. Bresin, K., Carter, D. L. & Gordon, K. H. The relationship between trait impulsivity, negative affective states, and urge for nonsuicidal self-injury: a daily diary study. *Psychiatry Res.* **205**, 227–231 (2013).

14. Bryan, C. J., Morrow, C. E., Etienne, N. & Ray-Sannerud, B. Guilt, shame, and suicidal ideation in a military outpatient clinical sample. *Depress. Anxiety* **30**,55–60 (2013).
15. Bryan, C. J., Ray-Sannerud, B., Morrow, C. E. & Etienne, N. Shame, pride, and suicidal ideation in a military clinical sample. *J. Affect. Disord.* **147**, 212–216 (2013).
16. Humber, N., Emsley, R., Pratt, D. & Tarrier, N. Anger as a predictor of psychological distress and self-harm ideation in inmates: a structured self-assessment diary study. *Psychiatry Res.* **210**, 166–173 (2013).
17. Nock, M. K., Prinstein, M. J. & Sterba, S. K. Revealing the form and function of self-injurious thoughts and behaviors: a real-time ecological assessment study among adolescents and young adults. *J. Abnorm. Psychol.* **118**, 816–827 (2009).
18. Olié, E. et al. Processing of decision-making and social threat in patients with history of suicidal attempt: a neuroimaging replication study. *Psychiatry Res.* **234**,369–377 (2015).
19. Dukart, J., Schroeter, M. L. & Mueller, K. Age correction in dementia — matching to a healthy brain. *PLoS ONE* **6**, e22193 (2011).
20. Koikkalainen, J. et al. Improved classification of Alzheimer's disease data via removal of nuisance variability. *PLoS ONE* **7**, e31112 (2012).
21. Jacobson, C., Batejan, K., Kleinman, M. & Gould, M. Reasons for attempting suicide among a community sample of adolescents. *Suicide Life Threat. Behav.* **43**,646–662 (2013).
22. Rogers, M. L., Kelliher-Rabon, J., Hagan, C. R., Hirsch, J. K. & Joiner, T. E. Negative emotions in veterans relate to suicide risk through feelings of perceived burdensomeness and thwarted belongingness. *J. Affect. Disord.* **208**, 15–21 (2017).
23. Pestian, J., Matykiewicz, P. & Linn-Gust, M. What's in a note: construction of a suicide note corpus. *Biomed. Inform. Insights* **5**, 1–6 (2012).
24. Pan, L. A. et al. Differential patterns of activity and functional connectivity in emotion processing neural circuitry to angry and happy faces in adolescents with and without suicide attempt. *Psychol. Med.* **43**, 2129–2142 (2013).
25. Beck, A. T. & Haigh, E. A. P. Advances in cognitive theory and therapy: the generic cognitive model. *Annu. Rev. Clin. Psychol.* **10**, 1–24 (2014).
26. Adler, A. et al. A mixed methods approach to identify cognitive warning signs for suicide attempts. *Arch. Suicide Res.* **20**, 528–538 (2016).
27. Jager-Hyman, S. et al. Cognitive distortions and suicide attempts. *Cognit. Ther. Res.* **38**, 369–374 (2014).
28. Brown, G. K., Steer, R. A., Henriques, G. R. & Beck, A. T. The internal struggle between the wish to die and the wish to live: a risk factor for suicide. *Am. J. Psychiatry* **162**, 1977–1979 (2005).
29. Bakhiyi, C. L., Calati, R., Guillaume, S. & Courtet, P. Do reasons for living protect against suicidal thoughts and behaviors? A systematic review of the literature. *J. Psychiatr. Res.* **77**, 92–108 (2016).
30. Jollant, F. et al. Orbitofrontal cortex response to angry faces in men with histories of suicide attempts. *Am. J. Psychiatry* **165**, 740–748 (2008).
31. Mann, J. J., Waternaux, C., Haas, G. L. & Malone, K. M. Toward a clinical model of suicidal behavior in psychiatric patients. *Am. J. Psychiatry* **156**, 181–189 (1999).
32. Brent, D. A. et al. Familial pathways to early-onset suicide attempt: a 5.6 year prospective study. *JAMA Psychiatry* **72**, 160–168 (2015).
33. Minzenberg, M. J. & Carter, C. S. Developing treatments for impaired cognition in schizophrenia. *Trends Cogn. Sci.* **16**, 35–42 (2012).
34. Suppes, P., Han, B., Epelboim, J. & Lu, Z.-L. Invariance of brain-wave representations of simple visual images and their names. *Proc. Natl Acad. Sci. USA* **96**, 14658–14663 (1999).
35. Kessler, R. C. et al. Predicting suicides after psychiatric hospitalization in US Army soldiers. *JAMA Psychiatry* **72**, 49–57 (2015).
36. Wechsler, D. *Wechsler Abbreviated Scale of Intelligence* (Harcourt Assessment, 1999).
37. Beck, A. T., Schuyler, D. & Herman, I. in *The Prediction of Suicide* (eds Beck, A. T. et al.) 45–56 (Charles Press, Bowie, MD, 1974).
38. Oquendo, M. A., Halberstam, B. & Mann, J. J. in *Standardized Evaluation in Clinical Practice* (ed. First, M. B.) 103–129 (American Psychiatric Press, Washington DC, 2003).
39. Posner, K. et al. *Columbia-Suicide Severity Rating Scale (C-SSRS)* (Columbia Univ. Medical Center, New York, NY, 2008).
40. Reynolds, W. M. *Professional Manual for the Suicidal Ideation Questionnaire* (Psychological Assessment Resources, Odessa, FL, 1987).
41. Achenbach, T. M., Howell, C. T., McConaughy, S. H. & Stanger, C. Six-year predictors of problems in a national sample: IV. Young adult signs of disturbance. *J. Am. Acad. Child Adolesc. Psychiatry* **37**, 718–727 (1998).
42. Achenbach, T. *Adult Self Report Measure for Ages 18–59* (Univ. Vermont, Burlington, VT, 2003).
43. Kroenke, K., Spitzer, R. L. & Williams, J. B. W. The PHQ-9: validity of a brief depression severity measure. *J. Gen. Intern. Med.* **16**, 606–613 (2001).
44. Spielberger, C. D. *State Trait Anxiety Inventory for Adults: Sampler Set: Manual, Test, Scoring Key;[form Y]* (Mind Garden, Menlo Park, CA, 1983).
45. Bernstein, D. P. et al. Initial reliability and validity of a new retrospective measure of child abuse and neglect (CTQ). *Am. J. Psychiatry* **151**, 1132–1136 (1994).

Acknowledgements

This research was partially supported by the National Institute of Mental Health Grant MH029617 and an Endowed Chair in Suicide Studies at the University of Pittsburgh School of Medicine. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

The experiments were conceived and designed by M.A.J., L.P., V.L.C., D.L.M. and D.B. The experiments were performed by M.A.J., L.P. and D.B. The materials and analysis tools were contributed by M.A.J., V.L.C., C.C. and M.K.N. The data were analysed by V.L.C. The paper was written by M.A.J., V.L.C. and D.B.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41562-017-0234-y.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to M.J.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature research

Corresponding author(s): Marcel Adam Just

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

   Describe how sample size was determined.

   The sample size for the main analysis (17 suicidal ideators and 17 controls) matches the sizes used in our previous study (Just et.al, 2014). The high accuracy of the group membership classification (91%) indicates that the sample size was sufficient.

2. Data exclusions

   Describe any data exclusions.

   The data from 21 suicidal ideators and 24 controls were excluded from the main analysis on the basis of the pre-established minimal accuracy (0.6) of identifying which of the 30 words the participant was thinking about. The data for excluded suicidal ideators was included in the further analyses.

3. Replication

   Describe whether the experimental findings were reliably reproduced.

   The results of the main analysis were replicated for the initially excluded suicidal ideators.

4. Randomization

   Describe how samples/organisms/participants were allocated into experimental groups.

   The suicidal ideator participants were recruited from a clinically verified population with current suicidal ideation. The control group was recruited from the community with the stipulation that there was no personal or family history of psychiatric disorder or suicide attempt.

5. Blinding

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   The investigators were not blinded to the group membership of participants; however, the machine learning classifier was not given the group membership information for the test participant.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

Matlab version R2014a
SPM8
Custom Matlab code (most of this code was used in the previously published studies)

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

No unique materials were used.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

No antibodies were used.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

No eucaryotic cell lines were used.

b. Describe the method of cell line authentication used.

No eucaryotic cell lines were used.

c. Report whether the cell lines were tested for mycoplasma contamination.

No eucaryotic cell lines were used.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

No commonly misidentified cell lines were used.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	No animals were used.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	The covariates in the two groups are reported in detail and treated statistically in the manuscript.

nature research

Corresponding author(s): Marcel Adam Just

☐ Initial submission ☐ Revised version ☒ Final submission

MRI Studies Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

Experimental design

1. Describe the experimental design.	event-related design
2. Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.	30 concepts were presented in 6 blocks, each block in a different random order. Each concept was displayed for 3 sec followed by a 4 sec blank interval. 17 sec long fixation intervals were included periodically to provide an activation baseline
3. Describe how behavioral performance was measured.	The behavioral performance was not measured during the experiment; the performance was estimated by the sufficiently high accuracy of the machine learning classifier that identified which of the 30 concepts the participant was thinking about.

Acquisition

4. Imaging	
a. Specify the type(s) of imaging.	Functional MRI
b. Specify the field strength (in Tesla).	3T
c. Provide the essential sequence imaging parameters.	Echo-planar pulse sequence, TR (repetition time) = 1000 ms, TE (echo time) = 30 ms, flip angle = 60 degrees, FOV (field of view) = 20 cm, matrix size = 64 x 64, voxel size of 3.125 x 3.125 x 5 mm thick (skipping 1mm between slices), 20 AC-PC aligned brain slices
d. For diffusion MRI, provide full details of imaging parameters.	No diffusion MRI was acquired.
5. State area of acquisition.	The acquisition matrix covered all of the cerebrum.

Preprocessing

6. Describe the software used for preprocessing.	fMRI images were preprocessed with SPM8 (Wellcome Dept. of Cog. Neurology). The images were slice-timing- and motion-corrected, and spatially normalized to the MNI template without changing voxel size (3.125 x 3.125 x 6 mm)
7. Normalization	
a. If data were normalized/standardized, describe the approach(es).	Non-linear normalization to the functional MNI template
b. Describe the template used for normalization/ transformation.	MNI305
8. Describe your procedure for artifact and structured noise removal.	Motion parameters were estimated and corrected in original subject's space
9. Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.	The main analysis was performed on activation in the clusters of stable voxels. These clusters were identified algorithmically based exclusively on the training data inside the cross-validation folds.

Statistical modeling & inference

10. Define your model type and settings.	Multivariate analysis using machine learning classifier (Gaussian Naïve Bayes)
11. Specify the precise effect tested.	The main effect tested was the ability to identify group membership of a participant (suicidal ideator or control), based on the fMRI activation during processing of discriminative concepts.
12. Analysis	
a. Specify whether analysis is whole brain or ROI-based.	Whole-brain analysis restricted to the clusters of stable voxels that were identified algorithmically using the training data.
b. If ROI-based, describe how anatomical locations were determined.	No anatomical ROIs were used.
13. State the statistic type for inference. (See Eklund et al. 2016.)	The classifier performance was compared to the accuracy expected by chance (obtained using binomial distribution or random classifier).
14. Describe the type of correction and how it is obtained for multiple comparisons.	The main analysis uses cross-validated machine learning classification that does not require statistical correction.
15. Connectivity	
a. For functional and/or effective connectivity, report the measures of dependence used and the model details.	No connectivity analyses were performed.
b. For graph analysis, report the dependent variable and functional connectivity measure.	No graph analyses were performed.
16. For multivariate modeling and predictive analysis, specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.	In the main analysis, the independent variables (features) were the activation levels for a number of (discriminating) concepts in a set of (discriminating) brain locations. The classifier was trained on the data of all but one participant, and the group membership of the left out participant was predicted. The main evaluation metric was the accuracy of the classification.