The neural representation of facial-emotion categories reflects conceptual structure

Jeffrey A. Brooks[a,1], Junichi Chikazoe[b], Norihiro Sadato[b], and Jonathan B. Freeman[a,c,1]

[a]Department of Psychology, New York University, New York, NY 10003; [b]Department of System Neuroscience, National Institute for Physiological Sciences, Okazaki 444-8585, Japan; and [c]Center for Neural Science, New York University, New York, NY 10003

Edited by Lisa Feldman Barrett, Northeastern University, Boston, MA, and accepted by Editorial Board Member Renée Baillargeon June 18, 2019 (received for review September 21, 2018)

Humans reliably categorize configurations of facial actions into specific emotion categories, leading some to argue that this process is invariant between individuals and cultures. However, growing behavioral evidence suggests that factors such as emotion-concept knowledge may shape the way emotions are visually perceived, leading to variability—rather than universality—in facial-emotion perception. Understanding variability in emotion perception is only emerging, and the neural basis of any impact from the structure of emotion-concept knowledge remains unknown. In a neuroimaging study, we used a representational similarity analysis (RSA) approach to measure the correspondence between the conceptual, perceptual, and neural representational structures of the six emotion categories Anger, Disgust, Fear, Happiness, Sadness, and Surprise. We found that subjects exhibited individual differences in their conceptual structure of emotions, which predicted their own unique perceptual structure. When viewing faces, the representational structure of multivoxel patterns in the right fusiform gyrus was significantly predicted by a subject's unique conceptual structure, even when controlling for potential physical similarity in the faces themselves. Finally, cross-cultural differences in emotion perception were also observed, which could be explained by individual differences in conceptual structure. Our results suggest that the representational structure of emotion expressions in visual face-processing regions may be shaped by idiosyncratic conceptual understanding of emotion categories.

emotion perception | facial expressions | conceptual knowledge | functional magnetic resonance imaging | fusiform gyrus

A fundamental debate about human emotion concerns the nature of emotion-concept knowledge and the manner in which emotion concepts are involved in experiencing emotions and perceiving them in other people. Classic theories of emotion assume that emotion categories map onto biologically and psychologically distinct states with specific behavioral and expressive profiles (1, 2). Facial expressions in particular are typically assumed to inherently signal specific emotions, triggering relatively stable and accurate categorizations in human perceivers, including those from different cultures, due to their high motivational relevance. As such, this approach tends to minimize the possible influence of top-down factors such as conceptual knowledge on facial-emotion perception (3–5). A number of findings support this idea by showing that perceivers often rapidly and reliably classify configurations of facial actions into specific emotion categories (3, 5–10).

This research has gone a long way in establishing that individuals show high agreement in their categorizations of specific facial expressions, but the tasks used typically present stereotyped facial expressions to perceivers without consideration of the context. Contrasting work demonstrates that emotion perception is highly sensitive to visual and social contextual factors, such as body posture (11, 12) and visual scenes (13, 14), suggesting that facial-emotion perception cannot be fully understood from sensitivity to facial cues alone (15–17). Accumulating evidence suggests that an additional influence on emotion perception comes from the internal context of the perceiver—that an individual's own conceptual associations with specific emotion categories may be implicitly used to make sense of visual information conveyed by the face, influencing the perceptual process (18–22). Thus, recent theoretical accounts and computational models highlight the variety of top-down conceptual, contextual, and associative factors that may weigh in on visual processing before categorizations of facial emotion stabilize (23–26).

These contrasting approaches to understanding emotion perception touch on fundamental debates about the degree to which visual perception is able to be influenced by more "cognitive" resources, such as conceptual knowledge, in the first place (see ref. 27 for a recent discussion). Classic views of emotion assume that facial actions associated with particular emotions (e.g., a scowling face for Anger) trigger an invariant perceptual process that leads to recognition of Anger in an individual displaying these facial actions and subsequent activation of Anger-related conceptual knowledge (such as what actions that person is likely to take next; refs. 1 and 28). This approach treats emotion concepts as abstractions that come online at the end stage of a feed-forward perceptual process. Alternative approaches view them as more dynamic multimodal constructs with a functional role in resolving visual processing to rapidly predict and guide behavior (23, 29–31). These recent approaches would assume that early processing of facial cues associated with a given emotion would trigger an interactive perceptual process that utilizes contextual

Significance

Classic theories of emotion hold that emotion categories (e.g., Anger and Sadness) each have corresponding facial expressions that can be universally recognized. Alternative approaches emphasize that a perceiver's unique conceptual knowledge (e.g., memories, associations, and expectations) about emotions can substantially interact with processing of facial cues, leading to interindividual variability—rather than universality—in facial-emotion perception. We find that each individual's conceptual structure significantly predicts the brain's representational structure, over and above the influence of facial features. Conceptual structure also predicts multiple behavioral patterns of emotion perception, including cross-cultural differences in patterns of emotion categorizations. These findings suggest that emotion perception, and the brain's representations of face categories, can be flexibly influenced by conceptual understanding of emotions.

Author contributions: J.A.B., J.C., N.S., and J.B.F. designed research; J.A.B. and J.C. performed research; J.A.B. analyzed data; and J.A.B. and J.B.F. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. L.F.B. is a guest editor invited by the Editorial Board.

Data deposition: Data and code relevant to the results in this manuscript are publicly available and hosted by the Open Science Framework (https://osf.io/vurqd/).

[1]To whom correspondence may be addressed. Email: jab1148@nyu.edu or jon.freeman@nyu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1816408116/-/DCSupplemental.

Published online July 22, 2019.

cues, prior experiences, and conceptual knowledge to resolve visual input (23–26, 32).

Existing evidence for conceptual knowledge influencing facial-emotion perception comes from behavioral tasks that manipulate the presence or accessibility of emotion concepts, showing that reduced access impairs emotion perception, while increased access facilitates speed, accuracy, and memory in emotion-perception tasks (19, 20, 33, 34). Additional research shows that individuals with semantic dementia—who have dramatically reduced access to emotion-concept knowledge—have an impaired ability to perceive distinct emotion categories from facial expressions at all (21). These studies show that emotion-concept knowledge is flexible enough that even briefly manipulating emotion concepts can impact emotion perception in experimental tasks. One possible implication of this work is that emotion-concept knowledge is dynamic and flexible in general, such that idiosyncrasies in conceptual structure between individuals could differentially shape emotion perception, leading to variability between individuals in how emotions are perceived and categorized. Consistent with this idea, emotion perception exhibits well-documented variability between social groups (35, 36) and cultures (37–42). It is possible that such variability could be partly due to corresponding variability in the conceptual structure of emotion.

Despite considerable interest in the role of conceptual knowledge in facial-emotion perception, and its relationship with cultural and individual variability in perceptions, the neural basis of such a flexible influence remains unknown. In particular, while behavioral studies have been valuable, questions remain as to how deeply such conceptual impacts might manifest in perception. Specifying the level of processing at which perceptual representations reflect the influence of conceptual knowledge is a task particularly well-suited for neuroimaging to help address. Multivoxel pattern-analysis approaches to functional neuroimaging data, which measure distributed patterns of activity across voxels that can isolate different stimulus conditions, have shown that emotion categories can be decoded from responses to facial expressions in regions such as V1 (43), the posterior superior temporal sulcus (44), and the fusiform gyrus (FG; refs. 45 and 46). In a study comparing facial-emotion category classification performance in multiple brain regions, Wegrzyn et al. (46) found that neural patterns in the right FG (rFG) were best able to discriminate between emotion categories, emphasizing the important role of this region in facial-emotion perception.

While these studies show that brain regions involved in visual processing contain information about facial-emotion categories, it is unclear to what extent such representations may reflect the influence of conceptual knowledge. For example, a plausible alternative explanation of existing behavioral findings is that conceptual structure affects perceptual judgments of facial emotion due to response biases or other postperceptual processes, rather than affecting perception itself. Existing neuroimaging findings have been unable to directly address this concern, but recent uses of representational similarity analysis (RSA; ref. 47) have proven useful in disentangling various influences (e.g., visual vs. conceptual) on multivoxel representations in other domains (e.g., refs. 47–54). In a recent set of behavioral studies, we used the RSA technique to show that individual differences in the conceptual structure of emotion categories are related to the structure of how those categories are perceived using perceptual judgment data (55). However, these results cannot necessarily inform at what level of representation such conceptual impacts may manifest. Determining whether conceptual knowledge shows an influence on perceptual brain regions (rather than regions that would indicate primarily postperceptual processing) is a critical step in understanding the manner in which conceptual knowledge is involved in emotion perception.

The present study (n = 40) used functional neuroimaging to test whether the conceptual structure of emotion categories may shape the structure of those categories' multivoxel representations in regions involved in facial-emotion perception. We hypothesized that differences in the extent to which any given pair of emotions (e.g., Anger and Disgust) are deemed conceptually more similar would predict a corresponding similarity in multivoxel response patterns to faces displaying those emotions. Thus, for all pairwise combinations of the six emotion categories Anger, Disgust, Fear, Happiness, Sadness, and Surprise, we measured perceptual similarity, conceptual similarity, and neural-pattern similarity using independent tasks in an RSA approach. To demonstrate an effect of conceptual similarity above and beyond any potential physical resemblances in the emotion expressions themselves, we additionally derived measures of visual similarity between categories based on stimuli's low-level image properties as well as internal representations from a biologically based computational model of object recognition (*Materials and Methods*).

We were additionally interested in the role of culture, given previous research findings that East Asian perceivers showed differential perception of high-arousal negative facial expressions compared with Western perceivers (42, 56–60). Indeed, research has shown that these patterns of categorizations yield perceptual representations of emotion expressions that are less discrete (i.e., not clustered as neatly into specific categories, both within and between individuals) in East Asian relative to Western perceivers (59, 61, 62). Here, we hypothesized that Japanese subjects (n = 20) would also show less discrete perceptual structure of emotions than American subjects (n = 20), but that this difference could be explained by differences in conceptual structure. More generally, investigating culture within the RSA framework allowed us to disentangle multiple influences on cultural differences in perceptual judgments. Most importantly, regardless of any such cultural differences, we hypothesized that whatever unique conceptual structure of emotion a given subject has will predict their unique perceptual structure and neural pattern structure involved in representing emotion categories.

Results

We took an RSA approach, measuring conceptual similarity, perceptual similarity, and neural-pattern similarity between each pairwise combination of the emotion categories Anger, Disgust, Fear, Happiness, Sadness, and Surprise (n = 40). RSA allows direct comparison of representational spaces from different modalities [e.g., behavioral measures, functional MRI (fMRI) activation patterns, and computational models] by mapping the correspondence between their similarity structures. Dissimilarity matrices (DMs) were used to model the similarity structure within each modality, and these DMs can be directly compared by using standard statistical techniques that measure the variance explained in one variable from another (e.g., correlation and regression). Given a DM derived from patterns of neural data, one can predict this neural DM from different candidate models, determining which model explains most of the variance in the neural DM. Thus, this approach allows researchers to adjudicate between competing explanations of the brain's representational structure. This technique has already proven informative in a variety of domains, including object recognition (51), social cognition (54), social categorization (53), affect (48), and emotional inference (52).

We measured neural pattern similarity through an fMRI task in which subjects passively viewed faces displaying expressions commonly associated with the six emotion categories Anger, Disgust, Fear, Happiness, Sadness, and Surprise—specifically, posed emotional expressions from the Japanese and Caucasian Facial Expressions of Emotion (JACFEE) database (63, 64). Japanese and American subjects were scanned in their respective countries, but with identical scanner protocols. Comparison of image-quality metrics derived from MRIQC (version 0.10.1) (65) revealed comparable signal and data quality between sites (*SI Appendix*, Table S1).

Following the scan, we conducted two emotion-categorization tasks that provided complementary measures of perceptual similarity. The first was an explicit ratings task that assessed any systematic "confusions" or discordance in the ways that facial expressions are categorized between individuals and cultures; in

the task, subjects made explicit categorizations of each face they previously saw in the scanner, with six choices corresponding to each emotion category. The second task was a mouse-tracking categorization task in which they made forced-choice speeded categorizations of each face, with two emotion categories as response options. In this task, subjects categorized using a computer-mouse click, and the trajectories of their response-directed hand movements were recorded on each trial. This permitted an assessment of the temporal dynamics leading up to explicit categorizations, which previous research has shown indexes multiple category coactivations during perception (66–68). (Throughout the paper, we refer to the measures derived from these tasks as reflecting "perceptual similarity," but, of course, the behavioral tasks cannot be construed as "purely" perceptual, in that they still inevitably rely on a perceptual judgment. The mouse-tracking measures how the perceptual process evolves over time before an explicit perceptual judgment, but, nevertheless, a judgment is still required. However, the fMRI task consisted of passive and unconstrained viewing of facial expressions, without any perceptual judgment or task demand.) Finally, subjects also completed a conceptual ratings task for each emotion category, used to measure conceptual similarity. Overall, we hypothesized that, for any given pair of emotion categories, conceptual similarity would predict perceptual similarity as well as neural-pattern similarity in regions involved in facial-emotion perception. To control for potential confounding influences from the visual stimuli themselves, we also computed measures of the inherent visual similarity of the face stimuli belonging to each category.

Behavioral Results.

Cultural differences in emotion perception. Our one cross-cultural hypothesis was to replicate in our paradigm previous studies finding differential perception of high-arousal negative facial expressions in Eastern vs. Western cultures (42, 56–60), which yields a less discrete perceptual structure in Japanese vs. American subjects (59, 61, 62). In the present work, when referring to categorization responses, we use the term "discordance" to refer to responses inconsistent with the intended emotion display, per the JACFEE stimulus database, so as not to imply that they are erroneous. We found significantly different rates of discordance in categorizations between cultures, with Japanese ($n = 20$) showing higher discordance ($M = 18.125\%$, $SD = 7.86$) than Americans [$n = 20$; $M = 6.875\%$, $SD = 7.56$), $t(38) = 4.614$, $P = 0.0000439$] (Fig. 1). In particular, Japanese subjects frequently categorized posed Angry facial expressions as Disgusted (and vice versa) and posed Fearful facial expressions as Surprised (and vice versa), relative to American subjects.

Although not of primary interest, since the stimulus set included both Caucasian and Japanese target faces, we also tested a possible interaction between subjects' culture and the culture of the target faces impacting rates of discordance. A 2 × 2 repeated-measures ANOVA on accuracy rates showed a significant interaction between a subject's culture and the target face's culture, $F(1,38) = 4.265$, $P = 0.046$. Follow-up paired-samples t tests demonstrated that Japanese subjects had higher accuracy for Caucasian faces ($M = 0.84$) compared with Japanese faces ($M = 0.79$), $t(19) = 2.6$, $P = 0.017$. American subjects' accuracy, however, did not differ between Japanese ($M = 0.93$) and Caucasian faces ($M = 0.93$), $t(19) = 0.27$, $P = 0.79$. It is worth noting that the outgroup advantage observed for Japanese perceivers in this analysis was consistent with recent metaanalytic work showing that emotion displays are more often recognized in targets from historically heterogeneous (vs. homogeneous) societies (69).

Consistent with previous work (59, 61, 62), these results suggest that Japanese subjects have a less discrete perceptual structure (as measured by patterns of discordant responses) of emotion categories. Our primary analyses focused on whether such differences in perceptual structure could be explained by individual variability in conceptual structure.

Behavioral RSA. Our primary behavioral analyses focused on assessing the relationship between conceptual and perceptual similarity for each individual subject and for each pairwise combination of the emotion categories Anger, Disgust, Fear,

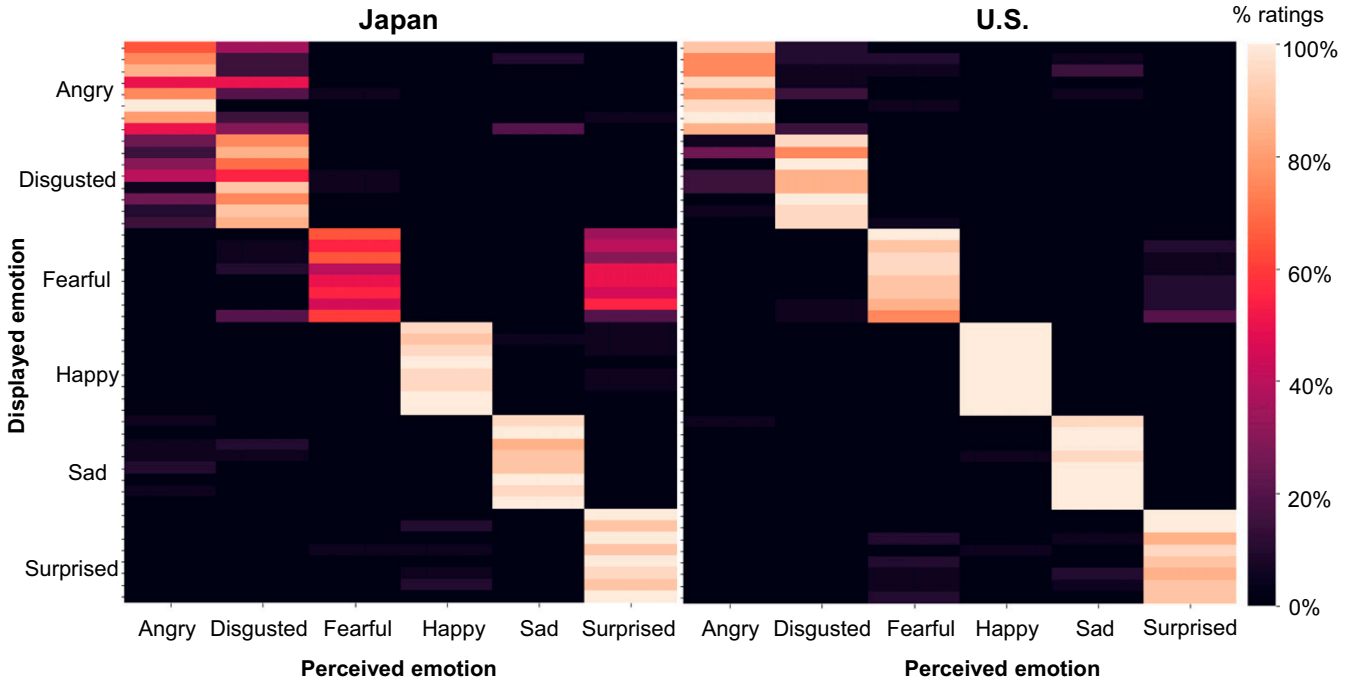

Fig. 1. Discordance between cultures in explicit categorizations of facial-emotion expressions. In an explicit categorization task, subjects categorized each face as one of the six emotion categories: Anger, Disgust, Fear, Happiness, Sadness, or Surprise. We found that the resulting perceptual structures showed systematically more "discordance," or categorizations not in accordance with the intended facial expression, in Japanese vs. American subjects. Rates of categorizations for each culture are shown for each image.

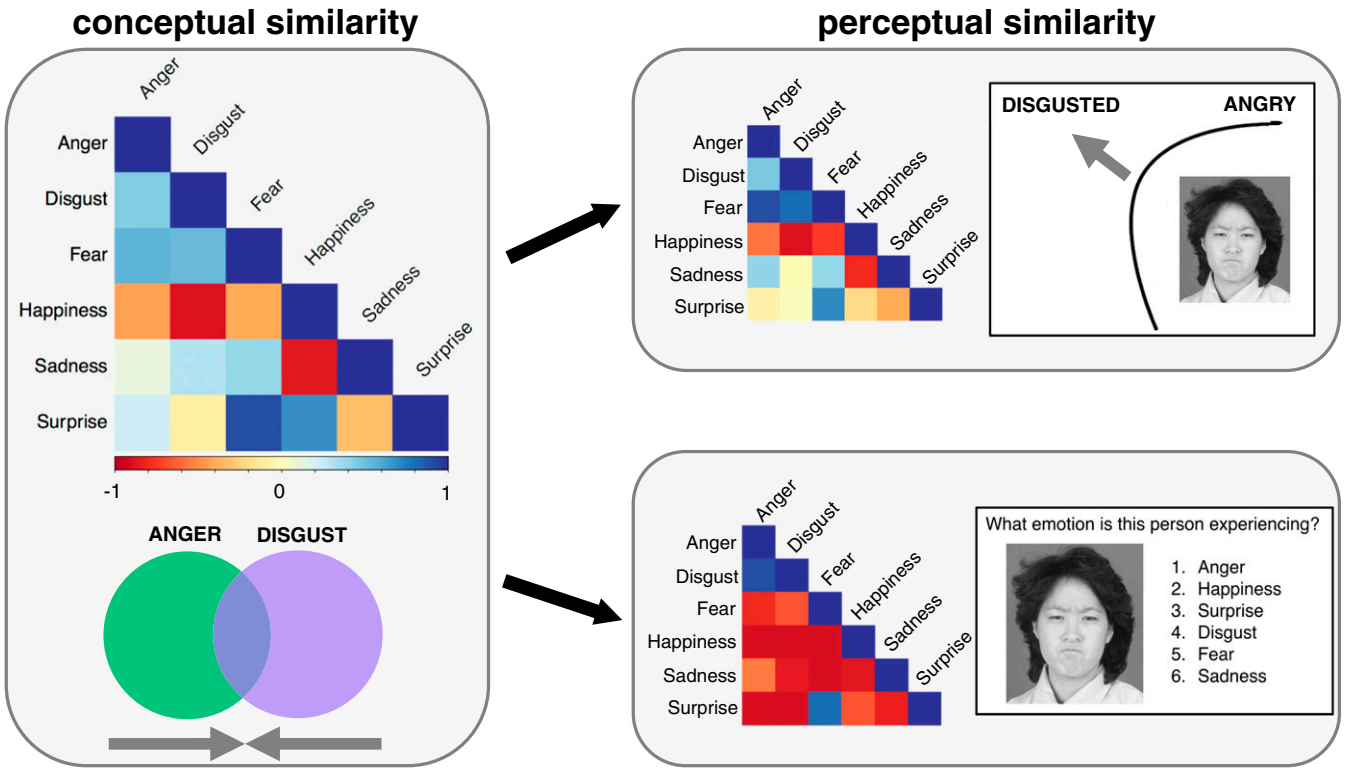

Fig. 2. Behavioral RSA results. We measured each subjects' conceptual similarity between each pairwise combination of the emotions Anger, Disgust, Fear, Happiness, Sadness, and Surprise through a conceptual ratings task. Perceptual similarity was measured two ways. One measure of perceptual similarity used computer mouse-tracking, which indexed participants' response-directed hand movements en route to an eventual category response in a facial-emotion perception task. Perceptual similarity was also assessed as participants' tendency to explicitly categorize faces in a "discordant" manner—that is, to select a certain category response (e.g., Disgust) instead of the intended category displayed by the image (e.g., Anger). Using each measure of perceptual similarity, we found that emotion categories that were conceptually more similar in the mind of a given subject were perceived with a corresponding similarity, controlling for the visual similarity of the stimuli in each category. For illustrative purposes, the whole-sample average conceptual DM, mouse-tracking perceptual DM, and explicit perceptual DMs are depicted.

Happiness, Sadness, and Surprise (yielding 15 unique pairs of emotion categories under the diagonal of the 6 × 6 DMs; Fig. 2).

We hypothesized that each individual's model of conceptual similarity would significantly predict their model of perceptual similarity, controlling for models of the visual similarity of the face stimuli used in each category. Conceptual DMs were created for each subject based on their data from a conceptual rating task in which they rated each emotion category on its conceptual relationship with a 40-item battery of emotion features used in previous research (55), including thoughts, bodily feelings, and action tendencies (items such as "crying," "heart racing," "clenching fists," etc.; *Materials and Methods* and *SI Appendix*, Table S2). We measured the overlap in patterns of responses for each pair of emotion categories, yielding a unique conceptual DM for each subject (Fig. 2 and *SI Appendix*, Figs. S1 and S4).

Separate perceptual DMs were created for each subject based on responses in the two perceptual tasks: the six-choice emotion categorization task (perceptual outcomes; Fig. 1) and the two-choice mouse-tracking task (perceptual process; Fig. 2 and *SI Appendix*, Figs. S2 and S5). Computer-mouse tracking is a well-validated measure of the temporal dynamics of perceptual categorization, providing an index of how multiple categories simultaneously coactivate and resolve during real-time perception (66–68). On each trial, subjects were presented with a face stimulus displaying an emotional facial expression and categorized it as one of two emotion categories (e.g., "Angry" vs. "Disgusted") by clicking on a response in either top corner of the screen. On every trial, one of the response options corresponded to the intended emotion display of the face stimulus. Maximum deviation (MD) of subjects' response trajectories toward the unselected category response provided an indirect measure of the degree to which the unselected category was simultaneously coactivated with the ultimately selected category during perception, despite only one facial emotion being depicted (*Materials and Methods*). Thus, each subject's average MD within each category pair (e.g., Anger–Disgust or Fear–Sadness) served as our measure of the perceptual similarity between those emotion categories (their degree of coactivation from the same facial stimuli; *SI Appendix*, Figs. S2 and S5).

As with the explicit categorization data, although not of primary interest, we submitted rates of accuracy as well as MD to a 2 × 2 repeated-measures ANOVA to assess a potential interaction between the subjects' culture and the culture of the face stimulus impacting these measures. No interaction was observed for either accuracy [$F(1,38) = 1.661$, $P = 0.205$] or MD [$F(1,38) = 0.006$, $P = 0.941$].

To control for the potential contribution of visual similarity between the stimuli in each pair of emotion categories, we computed two visual models of similarity in the low-level image properties of the stimuli in each category, as well as a third model derived from the similarity structure of the stimuli's internal representations from the HMAX computational model of visual object recognition (ref. 70; *Materials and Methods*). The resulting visual DMs comprehensively mapped the pairwise similarities between emotion categories in the visual features of their associated stimuli (*SI Appendix*, Fig. S3).

Conceptual, perceptual, and visual DMs were all recoded into comparable distance metrics, so that higher values in each model indicated greater dissimilarity and lower values indicated greater similarity between each emotion category pair (for ease of communication, we continue to use the term "similarity," but all

values analyzed in our models were in units of dissimilarity). To account for repeated measurements within subjects, we used a multilevel regression approach with generalized estimating equations (GEEs; ref. 71) to predict perceptual similarity (MD) from conceptual similarity, adjusting for the contribution of the three visual models. Unstandardized regression coefficients are reported.

From the explicit six-choice categorization task, rates of discordant categorizations were calculated for each subject and each emotion-category pair (e.g., for Anger–Disgust, how often a subject categorized a posed Angry facial expression as Disgusted or vice versa) and recoded into a distance metric to be used as a measure of perceptual similarity (*Materials and Methods*). Consistent with our predictions, we found that conceptual similarity significantly predicted perceptual similarity, controlling for the three different measures of visual similarity: $B = 0.1105$, $SE = 0.0235$, 95% CI [0.0644, 0.1567], $Z = 4.70$, $P < 0.0001$ (Fig. 2). From the two-choice mouse-tracking task, MD (i.e., the degree of category coactivation) for each subject and each category pair served as another measure of perceptual similarity. Again, conceptual similarity significantly predicted perceptual similarity, controlling for the three measures of visual similarity, $B = 0.0714$, $SE = 0.0096$, 95% CI [0.0525, 0.0903], $Z = 7.41$, $P < 0.0001$.

Conceptual DMs were constructed from subjects' ratings on a set of 40 features that included thoughts and bodily feelings (e.g., "nausea" or "heart racing") but also physical attributes of faces (e.g., "laughing" or "wide eyes"). To ensure that the association between conceptual and perceptual DMs was not spuriously produced by mere similarity in physical attributes of faces, we partitioned the 40 features into those which were face-related (12 features) and non-face-related (28 features; *SI Appendix*, Table S2) and computed two additional conceptual DMs from ratings on only face-related and non-face-related features. Separate models using each conceptual DM revealed that, expectedly, the face-related conceptual DM significantly predicted perceptual similarity (mouse tracking), controlling for the three different measures of visual similarity ($B = 0.0477$, $SE = 0.0068$, 95% CI [0.0344, 0.0610], $Z = 7.03$, $P < 0.0001$), as did the non-face-related conceptual DM ($B = 0.0669$, $SE = 0.0096$, 95% CI [0.0481, 0.0857], $Z = 6.98$, $P < 0.0001$). Critically, when including both DMs as predictors in the same model, the non-face-related conceptual DM remained a significant predictor ($B = 0.0401$, $SE = 0.0129$, 95% CI [0.0148, 0.0654], $Z = 3.11$, $P = 0.0019$), as well as the face-related conceptual DM ($B = 0.0276$, $SE = 0.0086$, 95% CI [0.0107, 0.0446], $Z = 3.19$, $P = 0.0014$). Separate models using each conceptual DM revealed a similar pattern of results for the perceptual DM derived from explicit categorization data. The face-related conceptual DM significantly predicted perceptual similarity (discordant responses), controlling for the three different measures of visual similarity ($B = 0.0379$, $SE = 0.0083$, 95% CI [0.0217, 0.0541], $Z = 4.58$, $P < 0.0001$), as did the non-face-related conceptual DM ($B = 0.0508$, $SE = 0.0113$, 95% CI [0.0286, 0.0729], $Z = 4.50$, $P < 0.0001$). As with the mouse-tracking perceptual DM, when including both DMs as predictors in the same model, the non-face-related conceptual DM remained a significant predictor ($B = 0.0274$, $SE = 0.0136$, 95% CI [0.0007, 0.0540], $Z = 2.01$, $P = 0.0440$), as well as the face-related conceptual DM ($B = 0.0245$, $SE = 0.0102$, 95% CI [0.0045, 0.0445], $Z = 2.40$, $P = 0.0162$). These analyses show that conceptual similarity unrelated to physical attributes of faces predicts perceptual similarity, above and beyond any effect of that related to physical attributes.

Thus, the behavioral results show that the more a subject believed two emotion categories to be conceptually similar predicted a greater number of categorization discordances and greater coactivation of the two categories (simultaneous attraction in hand movement to both responses). These results replicate and extend prior work (55), demonstrating that when emotion categories are more conceptually similar in the mind of a perceiver, their facial expressions are perceived with a corresponding similarity, as reflected in both an explicit and more indirect measure of emotion categories' perceptual similarity. In both cases, conceptual similarity predicted perceptual similarity over and above the physical similarity in the facial expressions themselves, and these results could not be explained by conceptual associations about faces' physical attributes alone. Moreover, with respect to the cultural differences reported earlier, these results suggest that, while there are differences in explicit categorizations of facial-emotion categories between cultures overall, whatever unique conceptual structure a given subject has is reflected in their perceptual structure, regardless of what culture they are from.

fMRI Results.

Searchlight RSA. We computed whole-brain activation maps for each subject, comprising their average neural response patterns to Angry, Disgusted, Fearful, Happy, Sad, and Surprised facial expressions (*Materials and Methods*). First, we aimed to identify any regions in which multivoxel response patterns to faces showed a similarity structure corresponding to the conceptual similarity structure of the subject's culture. We conducted a whole-brain searchlight analysis with multiple-regression RSA, testing whether neural-pattern similarity of local response patterns was significantly predicted by conceptual similarity, controlling for the three models of visual similarity ($P < 0.05$, corrected; *Materials and Methods*). This analysis revealed a region of the rFG ($x = 38$, $y = -43.6$, $z = -25$; mean $t = 4.54$; 146 voxels) (Fig. 3). No other regions survived correction.

To determine whether this rFG region reflected individual variability in emotion-concept knowledge, a region of interest (ROI) analysis was used to test whether rFG neural-pattern similarity was significantly related to each subject's own idiosyncratic conceptual DM. To ensure independence between the data used to generate the ROI and the data tested within the ROI (72), we employed a leave-one-out procedure in which the corrected-group statistical map was recomputed $n = 40$ times, with one subject left out of the analysis on each iteration. Each analysis yielded a similar cluster in the rFG to the one revealed by the analysis on the full sample. For each subject-specific rFG ROI, we tested whether the left-out subject's neural DM within this ROI was predicted by their idiosyncratic conceptual DM, once again controlling for the three models of visual similarity. The beta values from these analyses were submitted to a one-sample t test. Indeed, neural-pattern structure in the independent rFG ROI was significantly related to each subject's idiosyncratic conceptual structure, one-sample $t(39) = 2.867$, $P = 0.007$, and this could not be explained by inherent visual structure as described by three visual models. These results suggest that a subject's unique conceptual knowledge about emotion is reflected in rFG representational structure when viewing faces displaying emotional facial expressions.

Discussion

We found that neural patterns in the rFG exhibit a similarity structure of emotion categories that conforms to culturally and individually held conceptual knowledge about emotion, even when controlling for the potential contribution of visual cues. Moreover, conceptual similarity predicted multiple behaviorally derived estimates of the representational structure of emotion perception. These findings support recent theoretical and computational models of emotion and social perception which posit a fundamental relationship between conceptual knowledge and emotion perception (23–26). The results also dovetail with the literature on object recognition and visual cognition more generally, which suggest that perceptual category representations in the ventral temporal cortex may be subject to the influence of predictions derived from prior knowledge, memory, and contextual associations (32, 73–79).

Recent behavioral evidence has begun to support the idea that conceptual knowledge about emotion scaffolds facial-emotion perception, in contrast to classic approaches which assume that conceptual knowledge is more or less separable from emotion-related events. However, behavioral studies alone have been unable to identify which levels of perceptual processing are impacted by conceptual knowledge, in theory running the risk of

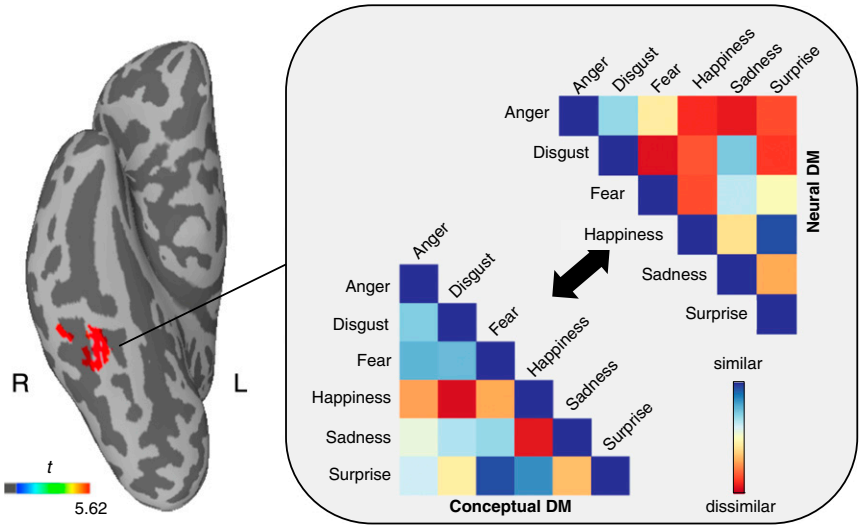

Fig. 3. Searchlight RSA results. At each searchlight sphere, the conceptual DM was used to predict the neural DM, controlling for three visual DMs. This analysis revealed a region of the rFG which showed a representational similarity structure that was predicted by the conceptual DMs, controlling for the visual DMs ($P < 0.05$, corrected).

capturing postperceptual outcomes. By integrating behavioral measures with multivariate fMRI, here we provided evidence suggesting that conceptual knowledge shapes representations of emotion categories in the rFG, a region critically involved in the perceptual processing of faces. These findings were afforded by an RSA approach that permitted a broad and comprehensive test of the correspondence between conceptual structure with perceptual and neural-pattern structure, but it was necessarily correlational. Thus, strong inferences about any causal relationship between conceptual knowledge and emotion perception would be unwarranted. Still, the possibility of top-down conceptual impact on perception via rFG representations is consistent with recent theory and evidence suggesting that low-level processing of visual input associated with a particular category may trigger relevant cognitive resources (i.e., conceptual knowledge shaped by past experience) that guide higher-level visual perception via regions, including the rFG (24, 25, 32, 73–79).

Indeed, recent work utilizing RSA finds that exclusively bottom-up, feature-based models of visual percepts cannot fully account for their representations in ventral temporal cortex (50, 80, 81) and that social expectations about face stimuli partly predict their representational structure in the rFG (53). Moreover, univariate responses in regions such as the rFG are sensitive to expectations about visual stimuli (32), including face stimuli in particular (82). Neural representations in this region can also be shaped by subjective categorizations made about face stimuli (83, 84). Such results suggest that ventral–temporal cortical representations do not purely reflect visual encoding of stimulus features, but comprise conceptually shaped or "visuosemantic" representations (50). Taken together with this previous work, the present findings suggest that, like other kinds of visual information, rFG representations of a given facial-emotion display may be tuned in part by conceptual understanding of that emotion.

While our results are consistent with recent developments in understanding the flexible nature of rFG representations, the stringency of our whole-brain searchlight analysis may have precluded an investigation into other stages of visual processing that could be more subtly influenced by conceptual knowledge. For example, research using eye-tracking shows that cultural differences in early visual processing of faces arise from spontaneous differences in the way that visual attention is deployed to extract information from faces (85). Moreover, while the passive viewing paradigm used in the fMRI task allowed for a more unconstrained perceptual task, this approach limits any inferences about whether different individuals were processing the images differently in real time. An important task for future work will be to determine whether conceptual knowledge has a broader impact on visual-processing regions, including earlier regions that feed output to the rFG. However, previous studies have found the rFG to be a primary site of top-down effects on various kinds of face processing without any broader cortical participation, and effective connectivity analyses suggest that the rFG may readily be modulated by higher-order regions with conceptual access during face processing (32, 82). [It is additionally possible that low statistical power or low signal-to-noise ratio may have resulted in only a single localized region of the rFG, such that improving power may have yielded additional regions. While we cannot exclude this possibility, quantified estimates of data quality revealed no issues at either scanning site (*SI Appendix*, Table S1).]

Cultural differences in emotion perception between Eastern and Western countries (and Japan vs. the United States in particular) have been widely studied, as they can often speak to fundamental debates regarding the universality of emotion. Our behavioral results replicate a well-established finding in this domain—differential perception of high-arousal negative facial expressions. Due to the widely held assumption that emotions are universally recognized from facial features alone, these findings are often taken to simply reflect different cultural norms. In particular, a prominent idea is that Japanese individuals have different "rules" for displaying and interpreting emotion, such that they regulate their level of expressiveness and their willingness to explicitly label others with certain emotion categories (37, 40, 41, 86). However, in the present study, an additional predictor of an individual's rate of discordance in categorization was not only their culture, but their internal conceptual model of emotions. While we expect that cultural norms still play a role, our findings suggest that "incorrect" responses in emotion-perception tasks may reflect genuine discordance in perception rather than the outcome of a regulatory strategy.

Indeed, rather than relying on broad distinctions between cultures, our analyses make use of individual subjects' conceptual structure, demonstrating substantial within-cultural variability between individuals that is reflected in emotion perception. We hope that our approach may inspire future research to explore variability within cultures, as well as between them, to understand variation and flexibility in a variety of psychological domains. One task for future research is to further understand the nature and origins of subtle interindividual variability in emotion-concept knowledge. Recent developmental work has focused on understanding how children acquire and differentiate emotion concepts (e.g., ref. 87). To determine how conceptual structure can differ subtly between individuals, continued efforts in this domain may benefit from more attention to the differences in verbal and nonverbal displays of emotion that occur during sensitive periods of development. More generally, interindividual variability in

emotion has previously been studied with constructs such as emotion "differentiation," "granularity," and "complexity," which capture different aspects of variability in individuals' use of emotion categories to describe their affective experiences (88–90). An important task for future research will be to determine whether similar latent constructs underlie variability in emotion perception, or if differences in conceptual structure may globally shape an individual's relationship to emotion experience and perception. To detect these differences (which may manifest in instability of the underlying dimensions of conceptual structure), future studies could also benefit from studying more emotion categories beyond the six emotion categories tested in the present study. Indeed, while the fMRI task did not impose any constraints on subjects while they viewed faces, both behavioral tasks had a fixed set of category response options on each trial. It is possible that different patterns of cultural and individual differences may emerge when subjects are given a wider variety of emotion category labels to choose from, or when experimenter constraints on category options are removed entirely.

A related limitation of this research concerns our materials, some of which were produced in a Western context. For example, the emotion features used to calculate conceptual similarity came from American participants in an earlier study (55). To match the demographic composition of our sample, we used images from the only facial-expression database to our knowledge including validated and controlled emotion displays from both Japanese and Caucasian individuals. As in other databases of emotional facial expressions, the facial displays in these images are posed to minimize ambiguity and maximize categorization performance in human subjects. However, the validity of the assumption that these facial expressions are "recognized" universally has been weakened by work showing that facial expressions themselves—the configurations of facial actions associated with each specific category—may differ substantially between Eastern and Western cultures (59, 61, 62). Therefore, the images may primarily conform to Western norms and expectations about facial expressions. Indeed, some recent work shows that, when asked to pose emotion expressions, Japanese perceivers do not typically generate the stereotypical expressions associated with Western norms (91). Cross-cultural studies on emotion perception using more data-driven techniques to produce face stimuli have been able to estimate more nuanced patterns of consistency and discordance in facial-emotion displays between cultures (e.g., ref. 92). However, this limitation does not only impact cross-cultural investigations. Even within the same culture, few studies find that individuals spontaneously produce these stereotyped facial poses during real-world instances of emotion (93). Future work should take care to introduce variability into the facial actions displayed in the stimulus set to increase ecological validity when measuring variability in emotion perception between individuals, cultures, and contexts.

Finally, our findings have relevance for artificial intelligence and computer vision—fields that are rapidly converging with psychology and neuroscience. Enormous effort has focused on developing computational models that perform as well as humans in classifying the emotions of others (94, 95). However, these approaches have largely been inspired by classic theories of emotion perception that emphasize the role of sensitivity to particular combinations of facial cues. As a result, existing computational models primarily focus on extraction of facial features (95–98). Our findings suggest that, if computational models wish to capture the full range of human performance in emotion-perception tasks, they may need to incorporate computational implementations of emotion-concept knowledge, including contextual situational associations and related appraisals. Developing such computational accounts of emotion concepts promises to benefit multiple fields (see refs. 52, 99, and 100 for early efforts in this area), but finding the exact specifications of these models will be a demanding task for future research.

Materials and Methods

Subjects. A total of 40 subjects participated in the study, all of whom were right-handed, with normal or corrected-to-normal vision and no history of neurological or psychiatric disease. All subjects were financially compensated and provided informed consent in a manner approved by the New York University Institutional Review Board and/or the Ethics Committee at the National Institute for Physiological Sciences, both of which approved the experimental procedures described here. Japanese subjects (n = 20) were recruited from the surrounding community of the National Institute for Physiological Sciences in Okazaki, Japan (10 female, M_{age} = 21.75, SD_{age} = 1.65). To minimize any potential influence of cultural and linguistic experiences on task performance, Japanese subjects were only recruited if they reported little to no English language ability and no time spent abroad where they would have had firsthand experience of the English language or Western culture. American subjects (n = 20) were recruited from the surrounding community of New York University (12 female, M_{age} = 24.5, SD_{age} = 6.16). American subjects had no Japanese language ability and no time spent abroad where they would have had firsthand experience of Japanese culture.

Stimuli. Face stimuli were 48 photographs from the JACFEE (64) stimulus set. For each of the six facial-emotion categories Anger, Disgust, Fear, Happiness, Sadness, and Surprise, the stimuli comprised eight images, including four Caucasian (two female) and four Japanese individuals (two female) portraying stereotyped emotional facial expressions (e.g., scowls for Anger and smiles for Happiness). This particular stimulus set was chosen to match the demographic composition of the sample (half Caucasian and half Japanese). No single identity was depicted in more than one image. Stimuli were converted to grayscale and matched on luminance and contrast by using the SHINE toolbox to control for low-level image properties (101).

Word and phrase stimuli used to measure conceptual similarity were taken from a previous study (55). Subjects in this study were instructed to "list the top 5 bodily feelings, thoughts, or actions" they personally associated with the six emotion categories under study. We took the 40 words and phrases that were reported most frequently across all emotions and subjects and used those as stimuli in the conceptual rating task. For the Japanese sample, all text-based materials were translated into Japanese by J.C., who speaks both Japanese and English and has spent substantial time living in both Japan and the United States. The translated tasks were pretested on several bilingual (Japanese/English) researchers to ensure translational equivalence.

MRI Acquisition. Japanese and American subjects were scanned by using identical acquisition protocols. Japanese subjects were scanned by using a Siemens 3T Magnetom Verio with a 32-channel head coil at the National Institute for Physiological Sciences. American subjects were scanned on a Siemens 3T Magnetom Prisma with a 32-channel head coil at the New York University Center for Brain Imaging. Structural images were acquired by using a 3D MPRAGE T1-weighted sequence with the following parameters: 1,800-ms repetition time (TR); 1.97-ms echo time (TE); 1.0-mm^3 voxel size; 256-mm field of view (FOV); 176 slices with no gap; anterior–posterior phase encoding direction. Functional images were acquired by using a multiband echo-planar imaging sequence with the following parameters: 1,000 ms TR, 35 ms TE, 2.0 mm^3 voxel size; 192 mm FOV; 60 slices with no gap; anterior–posterior phase encoding direction; multiband acceleration factor of 6. Gradient spin-echo field maps were also acquired in both the anterior–posterior and posterior–anterior phase encoding directions for use in correcting for potential susceptibility artifacts. Diffusion-weighted images were also collected, but those data are not presently reported.

fMRI Task. The fMRI task used an event-related design that largely followed the procedures used in refs. 49 and 53. Across 10 functional runs lasting 5 min and 24 s each (thus totaling 54 min of scanning time), subjects passively viewed faces displaying posed facial expressions commonly associated with the six emotion categories under study. Each functional run included six trials, each of which consisted of six encoding events, one null event (fixation), and one probe event, in which subjects were instructed to make a "yes" or "no" recognition judgment about whether the probe face appeared in the same trial. Probe events were included to ensure subjects' attention to the face stimuli. Encoding events were presented in a pseudorandomized order to ensure a similar stimulus order and distribution of probe and intertrial intervals (ITIs) between subjects. Trials were separated by ITIs ranging from 2,000 to 6,000 ms. Each encoding event presented one face stimulus for 1,500 ms, followed by a 4,500-ms fixation cross. Probe events also followed

the same structure, presenting the face probe for 1,500 ms followed by a 4,500-ms fixation cross.

Behavioral Tasks.

Explicit categorization task. Subjects completed a task in which they provided explicit emotion categorizations for each of the 48 stimuli from the fMRI task. Subjects were instructed to choose which specific emotion they thought the person in each photograph was experiencing, given the choice of all six emotion categories (Anger, Disgust, Fear, Happiness, Sadness, and Surprise).

Mouse-tracking categorization task. Mouse-tracking data were collected with a standard two-choice categorization paradigm implemented in MouseTracker software (102). Stimuli in the mouse-tracking task were the same faces from the fMRI and explicit categorization tasks. On each of 240 trials, subjects clicked a start button at the bottom center of the screen, which revealed a face stimulus. Each stimulus stayed on the screen until subjects chose one of two response options located in either top corner of the screen. On each trial, the response options were two emotion categories (e.g., "Angry" or "Disgusted"), one of which always corresponded to the posed expression displayed in the face stimulus. Response options changed on every trial and always included the ostensibly correct response. Trials were randomized, and the position of response options (left/right) was counterbalanced across trials. The specific number of trials was chosen to ensure an equal number of stimulus repetitions and trials per category pair, resulting in each stimulus being presented five times throughout the task and 16 trials per emotion category pair condition (e.g., Anger–Disgust).

Conceptual ratings task. Subjects completed a task in which they rated each emotion category (Anger, Disgust, Fear, Happiness, Sadness, and Surprise) on its conceptual relationship with a set of 40 traits including thoughts, bodily feelings, and associated actions, as used in prior work (ref. 55; *SI Appendix*, Table S2). Subjects attended to one emotion category at a time, for a total of six blocks presented in a randomized order. In each block, subjects rated each of the 40 word/phrase stimuli for how related it was to the emotion category in question (e.g., "On a scale from 1 = not at all to 7 = extremely, how related is 'tension' to the emotion Sadness?"), for a total of 240 trials. No faces were involved in this task.

Perceptual DMs. We used data from the explicit categorization task to produce a perceptual DM for each subject. In particular, for each subject, we measured pairwise discordance between emotion categories—the tendency to categorize a face as a different emotion category than the one it is posed to display. For example, the Anger–Disgust cell in this perceptual DM would capture the percentage of times that a subject categorized an Angry face as Disgusted or a Disgusted face as Angry. To use these values as a dissimilarity metric, we subtracted them from 1, such that a value of 0 would mean that all Angry faces were rated as Disgusted and all Disgusted faces were rated as Angry (high perceptual similarity), and a value of 1 would mean that all categorizations were in accordance with the posed facial expression (high perceptual dissimilarity).

Data from the mouse-tracking task were also used to estimate the similarity between emotion categories in how they are perceived. Any mouse trajectories with response times >3 SDs above the mean for a given subject were excluded. This procedure resulted in at most 3.33% of trials being excluded for a given subject. Since we were interested in mouse-trajectory deviation toward unselected responses regardless of the eventual response (which reflects greater similarity in how a face was perceived between the two emotion-category options), "incorrect" responses (i.e., final responses not in concordance with the posed facial expression) were included in the dataset.

Per standard preprocessing procedures for mouse-tracking data (102), trajectories were normalized into 100 time bins by using linear interpolation and rescaled into an x,y coordinate space with [0,0] as the start location. MD was calculated for each mouse trajectory as the trajectory's maximum perpendicular deviation toward the unselected response option on the opposite side of the screen from the ultimately chosen response. To construct perceptual DMs from the mouse-tracking data, MD was rescaled for each subject within a range of [0,1] such that 0 corresponded to a subject's largest MD (reflecting high category coactivation and thus perceptual similarity) and 1 corresponded to their smallest (reflecting perceptual dissimilarity). MD was then averaged within category pair (e.g., for the Anger–Disgust cell, average rescaled MD on all trials with "Angry" and "Disgusted" as response options).

Conceptual DM. To obtain a conceptual DM for each subject and for each of the six emotion categories Anger, Disgust, Fear, Happiness, Sadness, and Surprise, we calculated the Pearson correlation distance between vectors of responses to the word and phrase stimuli for each emotion category in the conceptual ratings task. For example, to measure each subject's conceptual similarity between Anger and Disgust, we calculated the Pearson correlation distance between their Anger vector of 40 ratings and Fear vector of 40 ratings.

Visual DMs. To adjust for the possible contribution of bottom-up overlap in the physical features between images in two categories (e.g., physical resemblance between face stimuli in the Anger and Disgust categories), we included three visual controls in our model: silhouette, pixel-intensity map, and HMAX. To model low-level visual features in each image, we used custom MATLAB scripts to compute a silhouette model and pixel-intensity map for each image. For each stimulus, the silhouette model transformed the image into a silhouette (i.e., a matrix of 0 and 1 s with 0 corresponding to background pixels and 1 corresponding to face pixels) and then produced a flattened pixel-intensity map for the silhouette image (i.e., single vector of 1 and 0 s per image). Silhouette models typically perform well in modeling representations in the early visual cortex (47) and also capture retinotopic outlining of visual stimuli, accounting for any difference in facial shapes that may contribute to categorization responses (103). The additional pixel-intensity map was a model of general low-level image similarities computed on a pixel-by-pixel basis on the original (nonbinarized) stimuli.

To model higher-level visual features in each image, we submitted each image as input to the HMAX feed-forward computational model of object recognition (70) and extracted high-dimensional internal representations for each image from the C2 layer. We used a publicly available instantiation of the 2007 version of the model (104) implemented in MATLAB (https://maxlab.neuro.georgetown.edu/hmax.html). The HMAX model was designed to model the first 150 ms of visual processing in primate cortex, with the C2 layer accounting for position- and orientation-invariant visual object representations in posterior inferotemporal regions (70, 104). While newer convolutional neural networks can achieve higher-classification performance (105, 106), HMAX has the benefit of interpretability since it is designed with maximum fidelity to the known anatomical and physiological properties of neural computation. As such, HMAX representations provide a conservative estimate of how much representational content can be attributed to hierarchical feed-forward processing of features alone.

These three visual models were computed for each of the 48 stimuli used in the fMRI and mouse-tracking tasks. To compute dissimilarity values between emotion categories, we averaged representations from the visual models within-category (e.g., for each of the three models, the average representations for each of the images corresponding to Anger) and computed the Pearson correlation distance between the average values for each of the 15 pairs of emotion categories, resulting in 6×6 visual DMs for each measure of visual similarity.

fMRI Preprocessing. Functional data were first corrected for susceptibility artifacts by using TOPUP in FSL (107, 108). Subsequent preprocessing steps were performed by using FMRIPREP (Version 1.0.0) (109), a Nipype (110)-based tool. Each T1-weighted (T1w) volume was corrected for intensity nonuniformity by using N4BiasFieldCorrection (Version 2.1.0) (111). Skull-stripping and nonlinear spatial normalization was performed by using the antsBrainExtraction and antsRegistration tools in ANTs (Version 2.1.0) (112), using brain-extracted versions of both the T1w volume and template. Brain-tissue segmentation of cerebrospinal fluid (CSF), white matter (WM), and gray matter was performed on the brain-extracted T1w by using fast in FSL (Version 5.0.9) (113). By using the resulting tissue masks, physiological noise regressors were extracted for CSF, WM, and global signal at each functional volume by using CompCor (114).

Functional data were motion-corrected by using mcflirt in FSL (Version 5.0.9) (115). This was followed by coregistration to the corresponding T1w by using boundary-based registration (116) with 9° of freedom, using flirt in FSL. Motion-correcting transformations, BOLD-to-T1w transformation, and T1w-to-template (MNI) warps were concatenated and applied in a single step by using antsApplyTransforms in ANTs (Version 2.1.0) using Lanczos interpolation. Functional images were, finally, smoothed by using FSLSTATS with a 4-mm full-width at half-maximum Gaussian smoothing kernel.

RSA. In all behavioral and fMRI RSA analyses, the 15 unique dissimilarity values under the diagonal of the 6×6 conceptual, perceptual, visual, and neural DMs were vectorized and submitted to multiple regression. Behavioral RSAs were conducted by using GEE multilevel regression to account for intracorrelations due to repeated measurements (71). fMRI searchlight RSA was conducted by using custom Python code (see below). Vectors were z-normalized to isolate the relative pattern of each condition (removing absolute differences in vector magnitude and scale; refs. 50 and 117). Because multiple-regression RSA assumes a linear combination of multiple predictor DMs, these analyses require a dissimilarity measure that sums linearly; thus,

squared Euclidean distance is an appropriate measure (118). However, because squared Euclidean distances of normalized pattern vectors are equivalent (i.e., linearly proportional) to Pearson correlation distances (119), we report results in Pearson correlation distance for ease of understanding and for greater intuitiveness.

fMRI Data Analysis. To generate a single whole-brain pattern of activation per emotion category, the average hemodynamic response for each condition was estimated for every voxel in the brain and for every run by using 3dDeconvolve in AFNI. BOLD responses were modeled by using a general linear model with a design matrix that included a total of 15 predictors: six predictors for each stimulus condition (Anger, Disgust, Fear, Happiness, Sadness, and Surprise); one predictor to model probe (recognition) events; and eight predictors to model effects of no interest (average signal at each time point attributable to WM, global signal, linear motion in three directions, and angular motion in three directions). The seven event-related predictors (six emotions + probe events) were modeled as boxcar functions across the first 2,000 ms of each event, during which the face stimuli were presented. These regressors were convolved with a gamma variate function (GAM in AFNI) to model the hemodynamic response. Brain responses associated with probe events were not included in any subsequent analyses. The voxelwise *t* statistics comparing each of the remaining six stimulus conditions of interest to baseline in each run were averaged across runs, and the resulting statistical maps comprised whole-brain patterns of activation for each emotion category for use in searchlight RSA.

For each subject, multiple-regression searchlight RSA was performed by using PyMVPA (120). A searchlight sphere was centered on every voxel in the brain, and the similarity structure of the multivoxel patterns within the sphere (neural DM) was tested against the similarity structure of predictor models (conceptual and visual DMs). Specifically, at each 5-voxel (10-mm) radius searchlight sphere, the multivoxel response pattern for each of the six emotion categories was vectorized, and Pearson correlation distance was used to compute the neural dissimilarity of each category pair, yielding a 6×6 neural DM. The neural, conceptual, and visual DMs were rank-ordered (which is preferable when comparing DMs derived from different measures; ref. 47), and ordinary least-squares regression was used to predict the neural DM from the conceptual DM within each searchlight sphere, with the three visual DMs included as covariates. The resulting regression coefficient for the conceptual DM was then mapped back to the center voxel of the searchlight sphere. The resulting subject-level maps depict the whole-brain correspondence between the conceptual DM and neural DM, controlling for three models of visual feature-based similarity.

These subject-level maps were then tested at the group level by using a one-sample *t* test in conjunction with maximum statistic permutation testing using Randomize in FSL (121), which tested significance of the raw *t* statistic with 5,000 Monte Carlo simulations. The resulting group-level statistical maps are significant at the $P < 0.05$ level, corrected for multiple comparisons using threshold-free cluster enhancement (122), which controls the family-wise error rate without setting arbitrary cluster-forming thresholds.

Data and Code Availability. Data and code relevant to the results in this manuscript are publicly available and hosted by the Open Science Framework (123).

ACKNOWLEDGMENTS. We thank Ryan Tracy and Yoshimoto Takaaki for assistance with data collection and Ryan Stolier for assistance with data analysis. This work was supported in part by National Science Foundation East Asia and Pacific Summer Institutes Grant NSF-OISE-1713960 (to J.A.B.) and National Institutes of Health Research Grant NIH-R01-MH112640 (to J.B.F.).

1. P. Ekman, D. Cordaro, What is meant by calling emotions basic. *Emot. Rev.* **3**, 364–370 (2011).
2. J. L. Tracy, D. Randles, Four models of basic emotions: A review of Ekman and Cordaro, Izard, Levenson, and Panskepp and Watt. *Emot. Rev.* **3**, 397–405 (2011).
3. P. Ekman, Facial expression and emotion. *Am. Psychol.* **48**, 384–392 (1993).
4. C. Montag, J. Panksepp, Primal emotional-affective expressive foundations of human facial expression. *Motiv. Emot.* **40**, 760–766 (2016).
5. J. L. Tracy, R. W. Robins, The automaticity of emotion recognition. *Emotion* **8**, 81–95 (2008).
6. R. Adolphs, Neural systems for recognizing emotion. *Curr. Opin. Neurobiol.* **12**, 169–177 (2002).
7. R. Adolphs, Cognitive neuroscience of human social behaviour. *Nat. Rev. Neurosci.* **4**, 165–178 (2003).
8. C. Darwin, *The Expression of the Emotions in Man and Animals* (Oxford University Press, New York, NY, 1872).
9. P. Ekman, Facial expressions of emotion: New findings, new questions. *Psychol. Sci.* **3**, 34–38 (1992).
10. M. L. Smith, G. W. Cottrell, F. Gosselin, P. G. Schyns, Transmitting and decoding facial expressions. *Psychol. Sci.* **16**, 184–189 (2005).
11. H. K. M. Meeren, C. C. R. J. van Heijnsbergen, B. de Gelder, Rapid perceptual integration of facial expression and emotional body language. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 16518–16523 (2005).
12. J. Van den Stock, R. Righart, B. de Gelder, Body expressions influence recognition of emotions in the face and voice. *Emotion* **7**, 487–494 (2007).
13. R. Righart, B. de Gelder, Context influences early perceptual analysis of faces—An electrophysiological study. *Cereb. Cortex* **16**, 1249–1257 (2006).
14. R. Righart, B. de Gelder, Rapid influence of emotional scenes on encoding of facial expressions: An ERP study. *Soc. Cogn. Affect. Neurosci.* **3**, 270–278 (2008).
15. H. Aviezer, S. Bentin, V. Dudarev, R. R. Hassin, The automaticity of emotional face-context integration. *Emotion* **11**, 1406–1414 (2011).
16. H. Aviezer, R. Hassin, S. Bentin, Y. Trope, "Putting facial expressions into context" in *First Impressions*, N. Ambady, J. Skowronski, Eds. (Guilford Press, New York, NY, 2008).
17. R. R. Hassin, H. Aviezer, S. Bentin, Inherently ambiguous: Facial expressions of emotions, in context. *Emot. Rev.* **5**, 60–65 (2013).
18. L. F. Barrett, B. Mesquita, M. Gendron, Context in emotion perception. *Curr. Dir. Psychol. Sci.* **20**, 286–290 (2011).
19. M. Gendron, K. A. Lindquist, L. Barsalou, L. F. Barrett, Emotion words shape emotion percepts. *Emotion* **12**, 314–325 (2012).
20. K. A. Lindquist, L. F. Barrett, E. Bliss-Moreau, J. A. Russell, Language and the perception of emotion. *Emotion* **6**, 125–138 (2006).
21. K. A. Lindquist, M. Gendron, L. F. Barrett, B. C. Dickerson, Emotion perception, but not affect perception, is impaired with semantic memory loss. *Emotion* **14**, 375–387 (2014).
22. A. B. Satpute *et al.*, Emotions in "black and white" or shades of gray? How we think about emotion shapes our perception and neural representation of emotion. *Psychol. Sci.* **27**, 1428–1442 (2016).
23. L. F. Barrett, The theory of constructed emotion: An active inference account of interoception and categorization. *Soc. Cogn. Affect. Neurosci.* **12**, 1833 (2017).
24. J. B. Freeman, N. Ambady, A dynamic interactive theory of person construal. *Psychol. Rev.* **118**, 247–279 (2011).
25. J. B. Freeman, K. L. Johnson, More than meets the eye: Split-second social perception. *Trends Cogn. Sci.* **20**, 362–374 (2016).
26. K. A. Lindquist, Emotions emerge from more basic psychological ingredients: A modern psychological constructionist model. *Emot. Rev.* **5**, 356–368 (2013).
27. C. Firestone, B. J. Scholl, Cognition does not affect perception: Evaluating the evidence for 'top-down' effects. *Behav. Brain Sci.* **229**, 1–77 (2016).
28. A. F. Shariff, J. L. Tracy, What are emotion expressions for? *Curr. Dir. Psychol. Sci.* **20**, 395–399 (2011).
29. L. W. Barsalou, Situated simulation in the human conceptual system. *Lang. Cogn. Process.* **18**, 513–562 (2003).
30. L. W. Barsalou, "Abstraction as dynamic interpretation in perceptual symbol systems", L. Gershkoff-Stowe, D. Rakison, Eds. *Building Object Categories in Developmental Time* (Carnegie Mellon Symposia on Cognition, Erlbaum, Mahwah, NJ, 2005), pp. 389–431.
31. C. D. Wilson-Mendenhall, L. F. Barrett, W. K. Simmons, L. W. Barsalou, Grounding emotion in situated conceptualization. *Neuropsychologia* **49**, 1105–1127 (2011).
32. C. Summerfield, T. Egner, Expectation (and attention) in visual cognition. *Trends Cogn. Sci.* **13**, 403–409 (2009).
33. N. C. Carroll, A. W. Young, Priming of emotion recognition. *Q. J. Exp. Psychol. A* **58**, 1173–1197 (2005).
34. E. C. Nook, K. A. Lindquist, J. Zaki, A new look at emotion perception: Concepts speed and shape facial emotion recognition. *Emotion* **15**, 569–578 (2015).
35. P. Thibault, P. Bourgeois, U. Hess, The effect of group identification on emotion recognition: The case of cats and basketball players. *J. Exp. Soc. Psychol.* **42**, 676–683 (2006).
36. S. G. Young, K. Hugenberg, Mere social categorization modulates identification of facial expressions of emotion. *J. Pers. Soc. Psychol.* **99**, 964–977 (2010).
37. H. A. Elfenbein, Nonverbal dialects and accents in facial expressions of emotion. *Emot. Rev.* **5**, 90–96 (2013).
38. H. A. Elfenbein, N. Ambady, On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychol. Bull.* **128**, 203–235 (2002).
39. H. A. Elfenbein, N. Ambady, When familiarity breeds accuracy: Cultural exposure and facial emotion recognition. *J. Pers. Soc. Psychol.* **85**, 276–290 (2003).
40. H. A. Elfenbein, M. Beaupré, M. Lévesque, U. Hess, Toward a dialect theory: Cultural differences in the expression and recognition of posed facial expressions. *Emotion* **7**, 131–146 (2007).
41. D. Matsumoto, Cultural influences on the perception of emotion. *J. Cross Cult. Psychol.* **20**, 92–105 (1989).
42. D. Matsumoto, P. Ekman, American–Japanese culture differences in intensity ratings of facial expressions of emotion. *Motiv. Emot.* **13**, 143–157 (1989).
43. L. S. Petro, F. W. Smith, P. G. Schyns, L. Muckli, Decoding face categories in diagnostic subregions of primary visual cortex. *Eur. J. Neurosci.* **37**, 1130–1139 (2013).
44. C. P. Said, C. D. Moore, A. D. Engell, A. Todorov, J. V. Haxby, Distributed representations of dynamic facial expressions in the superior temporal sulcus. *J. Vis.* **10**, 11 (2010).
45. B. Harry, M. A. Williams, C. Davis, J. Kim, Emotional expressions evoke a differential response in the fusiform face area. *Front. Hum. Neurosci.* **7**, 692 (2013).
46. M. Wegrzyn *et al.*, Investigating the brain basis of facial expression perception using multi-voxel pattern analysis. *Cortex* **69**, 131–140 (2015).
47. N. Kriegeskorte, M. Mur, P. Bandettini, Representational similarity analysis—connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* **2**, 4 (2008).
48. J. Chikazoe, D. H. Lee, N. Kriegeskorte, A. K. Anderson, Population coding of affect across stimuli, modalities and individuals. *Nat. Neurosci.* **17**, 1114–1122 (2014).
49. A. C. Connolly *et al.*, The representation of biological classes in the human brain. *J. Neurosci.* **32**, 2608–2618 (2012).

50. S.-M. Khaligh-Razavi, N. Kriegeskorte, Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Comput. Biol.* **10**, e1003915 (2014).
51. N. Kriegeskorte *et al*., Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* **60**, 1126–1141 (2008).
52. A. E. Skerry, R. Saxe, Neural representations of emotion are organized around abstract event features. *Curr. Biol.* **25**, 1945–1954 (2015).
53. R. M. Stolier, J. B. Freeman, Neural pattern similarity reveals the inherent intersection of social categories. *Nat. Neurosci.* **19**, 795–797 (2016).
54. M. A. Thornton, J. P. Mitchell, Theories of person perception predict patterns of neural activity during mentalizing. *Cereb. Cortex* **28**, 3505–3520 (2018).
55. J. A. Brooks, J. B. Freeman, Conceptual knowledge predicts the representational structure of facial emotion perception. *Nat. Hum. Behav.* **2**, 581–591 (2018).
56. J. Y. Chiao *et al*., Cultural specificity in amygdala response to fear faces. *J. Cogn. Neurosci.* **20**, 2167–2174 (2008).
57. D. Matsumoto, F. Kasri, K. Kooken, American–Japanese cultural differences in judgments of expression intensity and subjective experience. *Cogn. Emotion* **13**, 201–218 (1999).
58. Y. Moriguchi *et al*., Specific brain activation in Japanese and Caucasian people to fearful faces. *Neuroreport* **16**, 133–136 (2005).
59. R. E. Jack, C. Blais, C. Scheepers, P. G. Schyns, R. Caldara, Cultural confusions show that facial expressions are not universal. *Curr. Biol.* **19**, 1543–1548 (2009).
60. N. Yrizarry, D. Matsumoto, C. Wilson-Cohn, American-Japanese differences in multiscalar intensity ratings of universal facial expressions of emotion. *Motiv. Emot.* **22**, 315–327 (1998).
61. R. E. Jack, R. Caldara, P. G. Schyns, Internal representations reveal cultural diversity in expectations of facial expressions of emotion. *J. Exp. Psychol. Gen.* **141**, 19–25 (2012).
62. R. E. Jack, O. G. B. Garrod, H. Yu, R. Caldara, P. G. Schyns, Facial expressions of emotion are not culturally universal. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 7241–7244 (2012).
63. M. Biehl *et al*., Matsumoto and Ekman's Japanese and Caucasian facial expressions of emotion (JACFEE): Reliability data and cross-national differences. *J. Nonverbal Behav.* **21**, 3–21 (1997).
64. D. Matsumoto, P. Ekman, *Japanese and Caucasian Facial Expressions of Emotion (JACFEE) and Neutral Faces (JACNeuF)* (San Francisco State University, San Francisco, CA, 1988).
65. O. Esteban *et al*., MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. *PLoS One* **12**, e0184661 (2017).
66. J. B. Freeman, Doing psychological science by hand. *Curr. Dir. Psychol. Sci.* **27**, 315–323 (2018).
67. R. M. Stolier, J. B. Freeman, A neural mechanism of social categorization. *J. Neurosci.* **37**, 5711–5721 (2017).
68. J. B. Freeman, R. Dale, T. A. Farmer, Hand in motion reveals mind in motion. *Front. Psychol.* **2**, 59 (2011).
69. A. Wood, M. Rychlowska, P. M. Niedenthal, Heterogeneity of long-history migration predicts emotion recognition accuracy. *Emotion* **16**, 413–420 (2016).
70. M. Riesenhuber, T. Poggio, Hierarchical models of object recognition in cortex. *Nat. Neurosci.* **2**, 1019–1025 (1999).
71. K.-Y. Liang, S. L. Zeger, Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22 (1986).
72. N. Kriegeskorte, W. K. Simmons, P. S. F. Bellgowan, C. I. Baker, Circular analysis in systems neuroscience: The dangers of double dipping. *Nat. Neurosci.* **12**, 535–540 (2009).
73. M. Bar, A cortical mechanism for triggering top-down facilitation in visual object recognition. *J. Cogn. Neurosci.* **15**, 600–609 (2003).
74. M. Bar *et al*., Top-down facilitation of visual recognition. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 449–454 (2006).
75. L. F. Barrett, M. Bar, See it with feeling: Affective predictions during object perception.*Philos. Trans. R Soc. Lond B Biol. Sci.* **364**, 1325–1334 (2009).
76. M. Chaumon, K. Kveraga, L. F. Barrett, M. Bar, Visual predictions in the orbitofrontal cortex rely on associative content. *Cereb. Cortex* **24**, 2899–2907 (2014).
77. K. Kveraga, J. Boshyan, M. Bar, Magnocellular projections as the trigger of top-down facilitation in recognition. *J. Neurosci.* **27**, 13232–13240 (2007).
78. K. Kveraga, A. S. Ghuman, M. Bar, Top-down predictions in the cognitive brain. *Brain Cogn.* **65**, 145–168 (2007).
79. C. O'Callaghan, K. Kveraga, J. M. Shine, R. B. Adams, , Jr, M. Bar, Predictions penetrate perception: Converging insights from brain, behaviour and disorder. *Conscious. Cogn.* **47**, 63–74 (2017).
80. K. M. Jóźwik, N. Kriegeskorte, K. R. Storrs, M. Mur, Deep convolutional neural networks outperform feature-based but not categorical models in explaining object similarity judgements. *Front. Psychol.* **8**, 1726 (2017).
81. K. Storrs, J. Mehrer, A. Walter, N. Kriegeskorte, Category-specialised neural networks best explain representations in category-selective visual areas. *Perception* **46**, 1217–1218 (2017).
82. C. Summerfield, T. Egner, J. Mangels, J. Hirsch, Mistaking a house for a face: Neural correlates of misperception in healthy humans. *Cereb. Cortex* **16**, 500–508 (2006).
83. C. J. Fox, S. Y. Moon, G. Iaria, J. J. Barton, The correlates of subjective perception of identity and expression in the face network: An fMRI adaptation study. *Neuroimage* **44**, 569–580 (2009).
84. A. Thielscher, L. Pessoa, Neural correlates of perceptual choice and decision making during fear-disgust discrimination. *J. Neurosci.* **27**, 2908–2917 (2007).
85. C. Blais, R. E. Jack, C. Scheepers, D. Fiset, R. Caldara, Culture shapes how we look at faces. *PLoS One* **3**, e3022 (2008).
86. D. Matsumoto, Cultural similarities and differences in display rules. *Motiv. Emot.* **14**, 195–214 (1990).
87. E. C. Nook, S. F. Sasse, H. K. Lambert, K. A. McLaughlin, L. H. Somerville, Increasing verbal knowledge mediates development of multidimensional emotion representations. *Nat. Hum. Behav.* **1**, 881–889 (2017).
88. L. F. Barrett, J. Gross, T. C. Christensen, M. Benvenuto, Knowing what you're feeling and knowing what to do about it: Mapping the relation between emotion differentiation and emotion regulation. *Cogn. Emotion* **15**, 713–724 (2001).
89. S. M. Kang, P. R. Shaver, Individual differences in emotional complexity: Their psychological implications. *J. Pers.* **72**, 687–726 (2004).
90. K. A. Lindquist, L. F. Barrett, "Emotional complexity" in *Handbook of Emotions*, M. Lewis, J. M. Haviland-Jones, L. F. Barrett, Eds. (Guilford, New York, NY, 2008).
91. W. Sato, S. Hyniewska, K. Minemoto, S. Yoshikawa, Facial expressions of basic emotions in Japanese laypeople. *Front. Psychol.* **10**, 259 (2019).
92. R. E. Jack, W. Sun, I. Delis, O. G. B. Garrod, P. G. Schyns, Four not six: Revealing culturally common facial expressions of emotion. *J. Exp. Psychol. Gen.* **145**, 708–730 (2016).
93. J. I. Durán, R. Reisenzein, J. Fernández-Dols, "Coherence between emotions and facial expressions: A research synthesis" in *The Science of Facial Expression*, J. M. Fernandez-Dols, J. A. Russell, Eds. (Oxford University Press, New York, NY, 2017).
94. B. C. Ko, A brief review of facial emotion recognition based on visual information. *Sensors (Basel)* **18**, E401 (2018).
95. A. M. Martinez, Computational models of face perception. *Curr. Dir. Psychol. Sci.* **26**, 263–269 (2017).
96. C. F. Benitez-Quiroz, R. Srinivasan, A. M. Martinez, EmotioNet: An accurate, realtime algorithm for the automatic annotation of a million facial expressions in the wild. *IEEE Conference on Computer Vision and Pattern Recognition* 16:5562–5570 (2016).
97. M. N. Dailey, G. W. Cottrell, C. Padgett, R. Adolphs, EMPATH: A neural network that categorizes facial expressions. *J. Cogn. Neurosci.* **14**, 1158–1173 (2002).
98. K. Zhao, W. S. Chu, F. De la Torre, J. F. Cohn, H. Zhang, Joint patch and multi-label learning for facial action unit detection. *IEEE Conference on Computer Vision and Pattern Recognition* 15:2207–2216 (2015).
99. B. Felbo, A. Mislove, A. Sogaard, I. Rahwan, S. Lehmann, "Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm" in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, M. Palmer, R. Hwa, S. Riedel, Eds. (Association for Computational Linguistics, Stroudsburg, PA, 2017).
100. D. C. Ong, J. Zaki, N. D. Goodman, Affective cognition: Exploring lay theories of emotion. *Cognition* **143**, 141–162 (2015).
101. V. Willenbockel *et al*., Controlling low-level image properties: The SHINE toolbox. *Behav. Res. Methods* **42**, 671–684 (2010).
102. J. B. Freeman, N. Ambady, MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behav. Res. Methods* **42**, 226–241 (2010).
103. J. B. Freeman, N. Ambady, Hand movements reveal the time-course of shape and pigmentation processing in face categorization. *Psychon. Bull. Rev.* **18**, 705–712 (2011).
104. T. Serre, A. Oliva, T. Poggio, A feedforward architecture accounts for rapid categorization. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 6424–6429 (2007).
105. S. R. Kheradpisheh, M. Ghodrati, M. Ganjtabesh, T. Masquelier, Deep networks can resemble human feed-forward vision in invariant object recognition. *Sci. Rep.* **6**, 32672 (2016).
106. Y. Li, W. Wu, B. Zhang, F. Li, Enhanced HMAX model with feedforward feature learning for multiclass categorization. *Front. Comput. Neurosci.* **9**, 123 (2015).
107. J. L. R. Andersson, S. Skare, J. Ashburner, How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor imaging. *Neuroimage* **20**, 870–888 (2003).
108. S. M. Smith *et al*., Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* **23** (suppl. 1), S208–S219 (2004).
109. O. Esteban *et al*., poldracklab/fmriprep, Version 1.0.0. Zenodo. https://zenodo.org/record/1095198#.XSOfd-tKhhE. Accessed 9 July 2019.
110. K. Gorgolewski *et al*., Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in python. *Front. Neuroinform.* **5**, 13 (2011).
111. N. J. Tustison *et al*., N4ITK: Improved N3 bias correction. *IEEE Trans. Med. Imaging* **29**, 1310–1320 (2010).
112. B. B. Avants, C. L. Epstein, M. Grossman, J. C. Gee, Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* **12**, 26–41 (2008).
113. Y. Zhang, M. Brady, S. Smith, Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imaging* **20**, 45–57 (2001).
114. Y. Behzadi, K. Restom, J. Liau, T. T. Liu, A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage* **37**, 90–101 (2007).
115. M. Jenkinson, P. Bannister, M. Brady, S. Smith, Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* **17**, 825–841 (2002).
116. D. N. Greve, B. Fischl, Accurate and robust brain image alignment using boundary-based registration. *Neuroimage* **48**, 63–72 (2009).
117. A. Alink, A. Walther, A. Krugliak, J. J. F. van den Bosch, N. Kriegeskorte, Mind the drift—Improving sensitivity to fMRI pattern information by accounting for temporal pattern drift. bioRxiv:10.1101/032391 (4 December 2015).
118. J. D. Carlin, N. Kriegeskorte, Adjudicating between face-coding models with individual-face fMRI responses. *PLoS Comput. Biol.* **13**, e1005604 (2017).
119. H. Nili *et al*., A toolbox for representational similarity analysis. *PLoS Comput. Biol.* **10**, e1003553 (2014).
120. M. Hanke *et al*., PyMVPA: A python toolbox for multivariate pattern analysis of fMRI data. *Neuroinformatics* **7**, 37–53 (2009).
121. A. M. Winkler, G. R. Ridgway, M. A. Webster, S. M. Smith, T. E. Nichols, Permutation inference for the general linear model. *Neuroimage* **92**, 381–397 (2014).
122. S. M. Smith, T. E. Nichols, Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage* **44**, 83–98 (2009).
123. J. A. Brooks, The neural representation of facial emotion categories reflects conceptual structure. Open Science Framework. https://osf.io/vurqd/. Deposited 9 July 2019.