nature human behaviour

LETTERS

https://doi.org/10.1038/s41562-018-0376-6

Conceptual knowledge predicts the representational structure of facial emotion perception

Jeffrey A. Brooks[1*] and Jonathan B. Freeman[1,2*]

Recent theoretical accounts argue that conceptual knowledge dynamically interacts with processing of facial cues, fundamentally influencing visual perception of social and emotion categories. Evidence is accumulating for the idea that a perceiver's conceptual knowledge about emotion is involved in emotion perception, even when stereotypic facial expressions are presented in isolation[1–4]. However, existing methods have not allowed a comprehensive assessment of the relationship between conceptual knowledge and emotion perception across individuals and emotion categories. Here we use a representational similarity analysis approach to show that conceptual knowledge predicts the representational structure of facial emotion perception. We conducted three studies using computer mouse-tracking[5] and reverse-correlation[6] paradigms. Overall, we found that when individuals believed two emotions to be conceptually more similar, faces from those categories were perceived with a corresponding similarity, even when controlling for any physical similarity in the stimuli themselves. When emotions were rated conceptually more similar, computer-mouse trajectories during emotion perception exhibited a greater simultaneous attraction to both category responses (despite only one emotion being depicted; studies 1 and 2), and reverse-correlated face prototypes exhibited a greater visual resemblance (study 3). Together, our findings suggest that differences in conceptual knowledge are reflected in the perceptual processing of facial emotion.

Without effort or a moment's deliberation, we perceive the emotions of others based on facial actions that are often subtle and fleeting. This everyday phenomenon is a feat of information processing and perceptual efficiency that presents a problem for researchers: how does the perceptual system reach these categorizations so quickly and confidently? A classic and influential solution to this problem is the 'basic emotion' approach, which treats emotional facial expressions as one example of unambiguous signals produced by the body during emotional experiences. These signals are typically assumed to cohere to specific emotional states, projecting these states to nearby conspecifics that are so evolutionarily attuned to these signals that they spontaneously read them off the face as categorical instances of discrete emotion[7,8]. Evidence for this approach comes from the remarkable speed and consistency that perceivers often show in categorizing facial emotion expressions[9–15].

Although such work traditionally focused on how specific and unambiguous facial actions drive particular emotion categorizations, factors such as context, prior experience and emotion concept knowledge have been increasingly acknowledged to play a role in emotion perception as well. Much research acknowledges the role of context and conceptual knowledge when facial expressions are ambiguous, incongruent with the surrounding visual context or as a regulatory influence (for reviews see refs [16,17]), but in the absence of such top-down biasing factors or extreme bottom-up ambiguity, a common view is that prototypical facial expressions of emotion (for example, a scowling face for Anger, a frowning face for Sadness) are directly tied to perceptions of discrete categories of emotion. However, the strong and rapid impact of context on perception of even prototypical expressions with very little ambiguity[18–22] suggests that facial emotion perception may be integrative, routinely utilizing available cognitive resources and associative cues such as context and conceptual knowledge to make meaning of facial expressions as instances of emotion.

Based on these empirical insights, approaches such as the theory of constructed emotion posit that emotion perception is an 'active inference' process in which early processing of facial cues tentatively linked to emotion categories through prior experiences (for example, a scowling face for Anger) implicitly activate related conceptual knowledge, which plays a crucial role in resolving perceptual input and integrating it with aspects of the situation and context for a categorization to solidify[23]. Similarly, computational models of social perception such as the dynamic interactive model[24] argue that a stable percept is the end result of an interplay between facial features, categories and associated conceptual knowledge, with recurrent feedback from higher-level conceptual knowledge influencing processing of facial cues before categorization is complete. Critically, these approaches assume that conceptual knowledge is integrated into the perceptual process well before a categorical percept (Angry, Disgusted) is reached, potentially allowing subtle differences in this conceptual structure to influence the way these categories are perceived from facial expressions. Thus, these approaches hypothesize that emotion concept knowledge not only comes into play to resolve ambiguity, or to communicate or label perceptions, but also provides associations that are able to influence the perceptual process itself[23–25]. An important implication of this perspective is that inter-individual variability in conceptual knowledge about emotion categories may shape how even prototypical facial expressions of emotion are perceived.

Despite considerable theoretical work positing a constitutive role for conceptual knowledge in facial emotion perception, relatively few studies have directly assessed the relationship between conceptual knowledge and emotion perception. The strongest evidence comes from behavioural studies showing that individuals who have reduced access to emotion concepts, either due to experimental manipulations[1,2] or neurological disease[3], show poor encoding of emotion percepts and diminished speed and accuracy in emotion perception tasks. A similar line of research shows that when access to emotion concepts is increased (usually through conceptual priming with emotion-related language), speed and accuracy of categorization increases[4,26], and memory for facial actions is biased

[1]Department of Psychology, New York University, New York, NY, USA. [2]Center for Neural Science, New York University, New York, NY, USA.
*e-mail: jab1148@nyu.edu; jon.freeman@nyu.edu

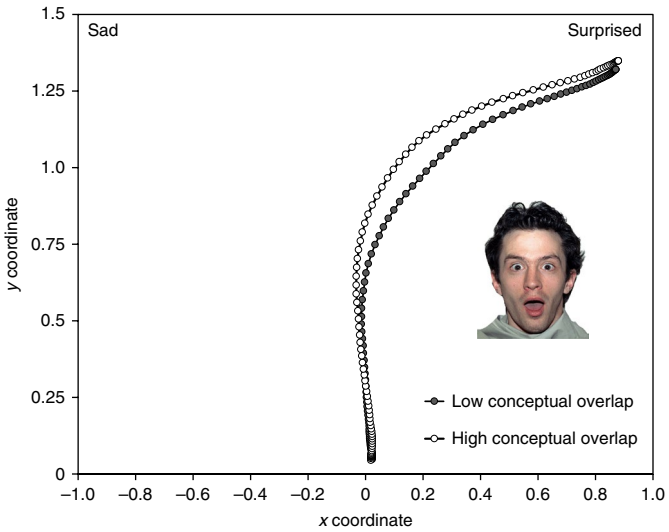

Fig. 1 | Mouse-tracking effects. On each trial of the mouse-tracking tasks, subjects were presented with a face stimulus and categorized it as one of two emotion categories, one of which corresponded to the posed emotion in the stimulus. MD of subjects' hand trajectories towards an unselected response is an index of how much that category was co-activated during perception. Complete mouse-tracking results for studies 1 and 2 are presented in Figs. 2 and 3 and Supplementary Fig. 3. To provide a visual example of the pattern of results for one category pair (Sad–Surprised), mean mouse trajectories during Sad versus Surprised trials in study 2 are depicted, separately for subjects with low and high conceptual similarity between Sad and Surprised categories (using median split). Trajectories are averaged across both Sad and Surprised responses and only arbitrarily depicted here as selecting the Surprised response. Subjects with high conceptual similarity exhibited a greater simultaneous attraction to select both emotion categories (even though only one was depicted on a given trial), manifesting in the mean trajectories and MD. Note that median split is only used here for visualization purposes; all analyses were conducted treating conceptual overlap as a continuous variable.

towards categorical facial expressions[27,28]. These findings suggest that conceptual knowledge weighs in on processing of facial cues during facial emotion perception. However, examining a patient population or experimentally manipulating perceivers to have more or less access to conceptual knowledge than they usually have access to in daily life does not address the fundamental role that conceptual knowledge may have in naturalistic emotion perception. Moreover, while such approaches have been valuable in manipulating conceptual knowledge of specific emotions in isolation, a more comprehensive assessment of how individual differences in conceptual knowledge relate to the representational space of facial emotion perception across emotion categories is still needed.

To permit this kind of comprehensive assessment, we adopted a representational similarity analysis (RSA[29]) approach. The RSA approach, originating from systems neuroscience, is to measure similarity (for example, correlation, distance) for all pairwise combinations of conditions based on one variable (for example, neural activity, reaction time, mouse trajectory) and correlate this pattern of similarity values with what one would expect from a specific model (for example, derived from another experimental modality, such as conceptual similarity ratings). This allows researchers to measure correspondence between different modalities or levels of analysis and adjudicate between competing models[29]. To study conceptual influences on emotion perception, RSA would permit an analysis of how the entire similarity structure or representational space of various emotion categories maps across conceptual, perceptual and visual levels. Here, we used RSA to predict the similarity in how any two emotion categories (for example, Anger and Disgust) are perceived from the similarity in how those categories are conceived conceptually, even when acknowledging the contribution of visual similarity between face stimuli belonging to those categories. Indeed, the RSA approach has already been a useful tool in other studies on emotion and person perception, allowing researchers to parse out the roles of visual features versus stereotypes in the neural representation of social categories[30] and test how well different models of affective experience predict neural patterns elicited by reading emotionally evocative scenarios[31].

Thus, in the present research, we aimed to provide a comprehensive test of the relationship between conceptual and perceptual similarity in facial emotion perception. In studies 1 and 2, we measured each subject's idiosyncratic conceptual similarity between each pairwise combination of the six 'basic' emotions commonly studied in the literature: Anger, Disgust, Fear, Happiness, Sadness and Surprise. Study 1 used a single rating of subjective similarity and study 2 used a more sophisticated method of assessing each category's conceptual contents. To measure perceptual similarity, we used computer mouse-tracking (Fig. 1), which is a well-validated measure of how multiple perceptual categories activate and resolve over hundreds of milliseconds, allowing a measure of the early processing of facial features before categorizations are complete. During two-choice categorization tasks (for example, Angry versus Disgusted), maximum deviation (MD) in a subject's hand trajectory towards an unselected category response provides an indirect measure of the degree to which that category was simultaneously co-activated during perception, despite not being explicitly selected. If conceptual knowledge about a given emotion category (for example, Anger) overlaps with conceptual knowledge about an alternative category (for example, Disgust), we hypothesize that subjects' perceptions will be biased towards that category and, consequently, their hand trajectories will deviate towards that category response in mouse-tracking tasks[32]. Indeed, such trajectory-deviation effects have been recently linked to neural markers of co-activated categories in brain regions involved in perceptual processing[33]. Thus, in studies 1 and 2, we hypothesized that conceptual similarity would significantly predict perceptual similarity, above and beyond any possible inherent physical similarity in the stimuli themselves. Finally, in study 3, we sought converging evidence using a reverse-correlation technique that can measure subjects' visual 'prototype' faces for each emotion category in a data-driven fashion. This allowed a test of how conceptual similarity predicts perceptual similarity with a more unconstrained approach and without making assumptions about stimuli or emotion-related features in advance (Fig. 2).

In study 1 ($N = 100$), we measured perceptual and conceptual similarity for each subject and for each pairwise combination of the emotion categories Anger, Disgust, Fear, Happiness, Sadness and Surprise (thus totalling 15 unique emotion category pairs under the diagonal of the 6×6 dissimilarity matrices (DMs); see Fig. 2). We also computed the inherent visual similarity of the emotion category pairs to statistically control for any physical resemblances among them (using Facial Action Coding System (FACS) measurement of the face stimuli used; see Methods). We hypothesized that, for each of the 15 emotion category pairs, conceptual similarity (conceptual DM) would predict perceptual similarity (perceptual DM), even when controlling for visual similarity. Conceptual similarity in study 1 was assessed with subjects providing a rating of the conceptual similarity of each emotion category pair (for example, Anger and Disgust) on a 10-point scale, comprising the conceptual DM (Fig. 2). Perceptual similarity of each emotion category pair was indexed by the trajectory-deviation effect (MD) on mouse-tracking trials where subjects made a categorization between the two categories (for example, Angry versus Disgusted; see Figs. 1 and 2), comprising the perceptual DM. Visual similarity of each emotion category

pair was indexed by the extent to which face stimuli belonging to the two categories (for example, Anger and Disgust) shared overlapping facial action units (AUs), as assessed by FACS coding (see Methods), comprising the visual DM (Supplementary Fig. 2a).

Multiple regression RSA was used to predict the perceptual DM from the conceptual DM, while controlling for potential visual similarity (visual DM), for each subject and for each of the 15 emotion category pairs. Conceptual, perceptual and visual similarities were calculated and recoded into comparable distance metrics, such that higher values indicate greater dissimilarity and lower values indicate greater similarity between any given emotion category pair (see Methods). Due to the multilevel nature of our design (15 similarity values for each variable nested within each subject), we conducted multilevel regression analyses using generalized estimating equations (GEEs), which can incorporate nested data while accounting for the intra-correlations in repeated-measures designs[34], consistent with previous RSA studies[30]. For all analyses, we report Wald Z tests with unstandardized regression coefficients (Bs). Note that these regression coefficients indicate the expected change in the outcome variable given a one-unit change in the predictor. As such, the B values reported provide a measure of effect size in the original units of the outcome variable.

Since our analyses aimed to predict the perceptual DM (15 unique emotion category pairs per subject) from the conceptual DM across subjects, noise (that is, inter-subject variability) in subjects' perceptual DMs can limit the ability of the perceptual DM to be explained by other models, such as the conceptual or visual DMs. Thus, we computed an estimate of the noise ceiling[35], which represents the range of variance that could be possibly explained if the unknown 'true' generative model was used as a predictor, given the noise in the perceptual DM. To estimate the lower bound of the noise ceiling (as in, for example, refs [31,36]), multilevel regressions analogous to our primary analyses were used to predict the group-average perceptual DM from each subject's perceptual DM. The effect size for this analysis (Wald Z) thereby served as the lower bound on the theoretical maximum effect size we could expect to observe given the noise in individual subjects' perceptual DMs. In all analyses, the effect size for each model (conceptual and visual DM) is provided as a proportion of this value, given as the percent of noise ceiling (%NC). The %NC thereby serves as an estimate of the relative importance of each model to variance in the outcome variable, constrained by any limitations inherent to measurement of this variable.

An inspection of the distribution of conceptual and perceptual similarity in emotion category pairs revealed substantial inter-subject variability in all studies (Supplementary Figs. 3 and 4). For study 1, we first regressed each subject's perceptual DM onto their own conceptual DM. The noise ceiling for this analysis was $Z = 5.54$, representing the lower bound on the theoretical maximum effect size we could expect to observe given the noise in individual subjects' perceptual DMs. Consistent with our predictions, conceptual similarity values predicted perceptual similarity values, $B = 8.5 \times 10^{-3}$, s.e. $= 1.1 \times 10^{-3}$, 95% confidence interval (CI) $[6.4 \times 10^{-3}, 1.1 \times 10^{-2}]$, $Z = 8.00$, %NC = 144.4%, $P < 0.0001$. Further, when including the visual DM (based on FACS AUs) as an additional predictor, the relationship between conceptual and perceptual similarity remained highly significant, $B = 7.9 \times 10^{-3}$, s.e. $= 1.1 \times 10^{-3}$, 95% CI $[5.7 \times 10^{-3}, 1.0 \times 10^{-2}]$, $Z = 7.29$, %NC = 131.6%, $P < 0.0001$ (Figs. 1–3). The model of visual similarity (FACS) did not significantly predict perceptual similarity, $B = 2.4 \times 10^{-3}$, s.e. $= 1.5 \times 10^{-3}$, 95% CI $[-5.0 \times 10^{-4}, 5.3 \times 10^{-3}]$, $Z = 1.64$, NC% = 29.6%, $P = 0.102$. Average conceptual, perceptual and visual DMs are found in Fig. 2 and Supplementary Fig. 2a. Note that individual subjects' conceptual and perceptual DMs were used in the multilevel regression analyses.

To provide more conservative estimates of visual similarity, we generated two additional visual DMs based on subsets of the FACS

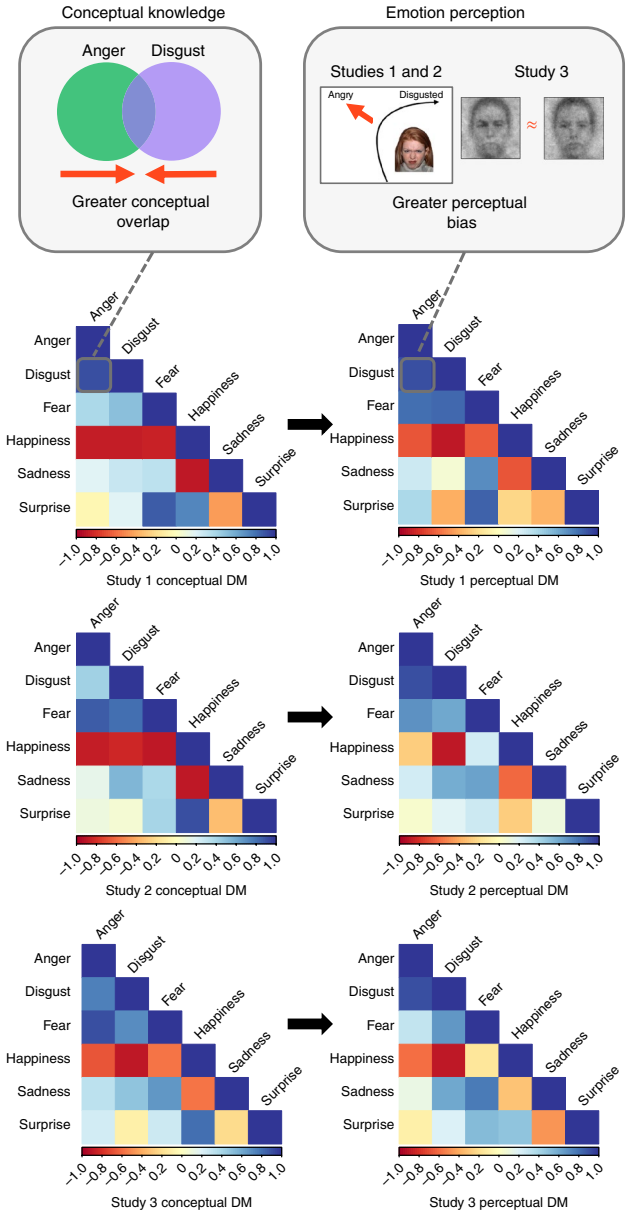

Fig. 2 | Average conceptual and perceptual DMs for studies 1–3 and schematic of the analytic approach. In all studies, conceptual similarity and perceptual similarity were assessed for all pairwise emotion combinations (for example, Anger–Disgust). Each subject's 15 unique emotion category pairs (unique values under the diagonal) for both conceptual structure and perceptual structure were vectorized and submitted to multilevel regression analyses, predicting perceptual similarity values from conceptual similarity values. Average conceptual and perceptual similarity structure (dissimilarity matrices (DMs)) are shown for all studies (study 1, $N = 100$; study 2, $N = 91$; study 3, $N = 368$). Average DMs for the two additional measures of perceptual similarity collected in study 3 are presented in Supplementary Fig. 1. Note that the average DMs for study 3 are presented for illustrative purposes only, as each cell under the diagonal included a different set of subjects. Due to the length of the task in study 3, each participant was randomly assigned to a different condition, where each condition was a given emotion category pair (for example, Anger–Disgust). The overall hypothesis was that a greater conceptual similarity between any two emotion categories (for example, Anger–Disgust) would correspond to a greater bias to perceive those emotions more similarly, measured by a simultaneous attraction to select both emotions during face perception with mouse tracking (studies 1 and 2) and a greater resemblance in estimated visual prototypes for the two emotions using reverse correlation (study 3).

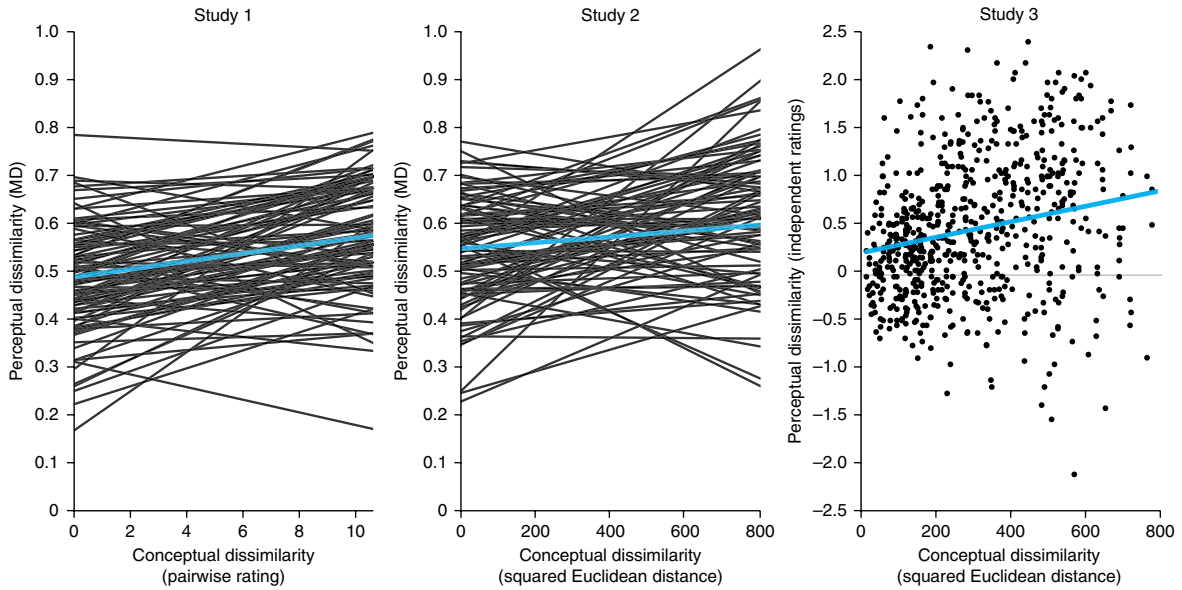

Fig. 3 | Multilevel regression results for studies 1–3. For illustrative purposes only, each subject's relationship between conceptual and perceptual similarity values are plotted as linear slopes (using ordinary least squares). Actual analyses were conducted using GEE multilevel regression. Individual subject slopes are not depicted for study 3 (instead, all data points are depicted) due to study 3's design in which no individual subject completed all experimental conditions. A positive relationship between conceptual and perceptual similarity was observed across studies. In studies 1 ($N = 100$) and 2 ($N = 91$), perceptual similarity was measured as average MD towards the unselected category response on mouse-tracking trials with the two categories in question as response options (for example, Angry–Disgusted). In study 3 ($N = 368$), perceptual similarity was measured through independent ratings of reverse-correlated prototype faces from each category. Mean slopes are shown in blue.

AUs that were specific to the emotion categories in question (see Methods; Supplementary Fig. 2b,c). Rerunning the regression models using these more conservative visual DMs did not change the significance of the conceptual DM (Supplementary Table 1). Note that the more conservative visual DMs did significantly predict the perceptual DM (Supplementary Table 1), which is sensible as physical resemblances in AUs among emotion categories should help drive perceptions. Thus, the more similar two emotion categories' stimuli were visually also predicted how similarly the categories were perceived. Critically, however, regardless of which variant of the visual DM was used, the effect of the conceptual DM remained highly significant and outperformed the visual DM in all cases. Thus, these results show that the extent to which any given pair of two emotions (for example, Anger and Fear) was deemed conceptually more similar predicted the extent to which those categories were perceived more similarly as reflected in a greater simultaneous attraction towards both categories during facial emotion perception (despite only one emotion being depicted; Figs. 1–3). Further, this effect of conceptual similarity between pairs of emotion categories holds above and beyond any effects of inherent visual similarity between the categories as assessed with three different visual models.

As the conceptual similarity ratings of study 1 were somewhat unconstrained, it is difficult to know exactly what drove the relationship between this conceptual similarity measure and the mouse-tracking data. One possibility is that subjects spontaneously chose to imagine how similar they find the stereotypical facial expressions associated with each category, and used this mental image as the main criterion for their judgements of similarity. While simulation of facial expressions may indeed be an important aspect of emotion concept knowledge (and conceptual knowledge more generally; see, for example, refs [37,38]), for our specific purposes we wanted to control as much as possible for the potential influence of physical similarity between categories. In study 2 ($N = 91$), we sought to replicate and extend the results of study 1 using a more fine-grained measure of conceptual similarity that minimizes the potential contribution of imagined physical overlap between categories. In particular, subjects rated each emotion category on its conceptual relationship with a large set of traits including thoughts, bodily feelings and associated actions (40 items). From the category-specific responses, we measured the conceptual similarity (that is, overlap) between all category pairs (see Methods). This measure therefore enabled us to capture a wide range of conceptual contents for each category and the overlap of those contents between categories (rather than similarity ratings, which may be more influenced by perceptual imagery effects).

As in study 1, we first regressed subjects' perceptual similarity values onto their conceptual similarity values. The lower bound of the estimated noise ceiling for this analysis was $Z = 5.56$. Replicating the previous results, the conceptual DM strongly predicted the perceptual DM, $B = 7.1 \times 10^{-5}$, s.e. $= 1.7 \times 10^{-5}$, 95% CI $[3.7 \times 10^{-5}, 1.0 \times 10^{-4}]$, $Z = 4.12$, %NC = 74.1%, $P < 0.0001$. Including visual similarity (FACS) as an additional predictor did not affect the relationship between conceptual and perceptual similarity, $B = 6.6 \times 10^{-5}$, s.e. $= 1.7 \times 10^{-5}$, 95% CI $[3.3 \times 10^{-5}, 9.9 \times 10^{-5}]$, $Z = 3.92$, %NC = 70.5%, $P < 0.0001$ (Figs. 1 and 3). The model of visual similarity predicted perceptual similarity at marginal significance, $B = 2.3 \times 10^{-3}$, s.e. $= 1.3 \times 10^{-3}$, 95% CI $[-2.0 \times 10^{-4}, 4.8 \times 10^{-3}]$, $Z = 1.83$, %NC = 32.9%, $P = 0.067$. As in study 1, we re-ran the model including more stringent visual DMs, which had a negligible impact on the effect of the conceptual DM (Supplementary Table 2). As in study 1, the two additional visual DMs were significant predictors of the perceptual DM (Supplementary Table 2), which is expected given that such visual resemblances in AUs should help drive perception. Most importantly, the conceptual DM remained a strong and significant predictor of the perceptual DM regardless of including any of the three visual DMs. These results extend the findings of study 1, showing they hold even when using a more comprehensive measure of conceptual similarity that minimizes possible imagined physical overlap. Thus, in studies 1 and 2, we have shown that when emotion

categories are more conceptually similar, they co-activate more strongly during perception of facial emotion as reflected in perceptual dynamics (Figs. 1–3). Further, this relationship between conceptual and perceptual structure held even when controlling for any bottom-up physical overlap that may be present between the categories in question.

Such findings demonstrate a fundamental correspondence between idiosyncratic conceptual and perceptual spaces of emotion perception, suggesting that early processing of facial cues is subject to input from conceptual knowledge about emotion categories. Further, idiosyncratic differences in emotion concept knowledge may partly shape how emotion is perceived from a face. However, during the mouse-tracking tasks, emotion category labels were present on the screen in every trial. A great deal of previous research has shown that the presence of emotion-related language increases the accessibility of emotion concept knowledge, impacting judgements of facial emotion[4], memory for facial emotion[27,28] and potentially decreasing affective reactivity to emotional stimuli in general[39]. As two emotion category labels were present on the screen in every trial, the correspondence between the conceptual similarity between any given pair of categories and their perception could, in theory, be explained by the pairs of linguistic labels priming related conceptual knowledge on each trial, causing a spurious influence on mouse trajectories that does not reflect an actual influence on perception. Study 3 aimed to address this issue. It also permitted a test of our predictions using a more data-driven approach that is not constrained to a particular face-stimulus set or normative assumptions about how the six emotion categories ought to appear on a face.

In study 3 ($N = 368$, see Methods for details on sample size), we used a reverse-correlation paradigm to assess correspondence between conceptual similarity and perception. On each trial of a reverse-correlation task, subjects must decide which of two faces are most likely to belong to a given category (for example, Anger, Disgust) even though the two faces are actually the identical neutral face, overlaid with different patterns of random noise. Averaging the faces chosen by a subject on each of many trials yields a 'classification image', representing a subject's visual prototype for that category (see Methods). Importantly, in this task, each category in question is attended to in isolation. Thus, if a subject with a higher level of conceptual overlap between Anger and Disgust yields an Anger prototype with a greater resemblance to their Disgust prototype (and vice versa), this would provide evidence that conceptual similarity impacts perceptual similarity in a case where no emotion-related stimuli or features were specified a priori and in a case where priming from response labels (for example, the presence of a 'Disgust' label priming judgements of Anger) is not possible.

Subjects were randomly assigned to one of the 15 possible emotion category pairs (for example, Anger and Sadness), and completed the reverse-correlation task for the two categories in separate randomized blocks, permitting us to generate reverse-correlated classification images for the two categories. We also assessed subjects' conceptual overlap between the two categories in question in a manner identical to study 2. Following completion of the study, we ran two independent rating tasks on the classification images from the reverse-correlation task. The first set of independent raters ($N = 95$) rated each classification image on emotion using a 7-point scale, where the categories on either end of the scale corresponded to the condition to which the initial subject was assigned. For example, if a rater was judging a classification image that was produced by a subject in the Anger–Disgust condition, they would rate the image on a scale from Angry to Disgusted. To provide converging evidence using an approach that did not draw raters' explicit attention to emotion categories, a second set of independent raters ($N = 92$) provided pairwise similarity ratings on pairs of classification images using a 7-point scale with 1 = not at all similar and 7 = extremely similar (see Methods). All pairs of classification images came from a single subject from the reverse-correlation task (that is, a rater might have to rate the Angry and Disgusted classification images from a subject in the Anger–Disgust condition). Note that this task did not feature any emotion category labels and only required raters to attend to the perceived similarity between any two images. Perceptual DMs from the two independent ratings tasks are presented in Fig. 2 and Supplementary Fig. 1a.

As in studies 1 and 2, we hypothesized that conceptual similarity between emotion categories would predict a greater similarity in how those categories are perceived. Perceptual similarity here was indexed by the bias in how the reverse-correlated classification images were perceived by independent raters. For example, we predicted that a subject with greater conceptual overlap between Anger and Sadness would yield an Anger classification image appearing more Sad, and a Sadness classification image appearing more Angry (as assessed by independent raters). Further, we predicted that pairs of classification images would appear more similar to independent raters when they were produced by a subject with greater conceptual overlap between those categories. To provide a comparable distance metric with the conceptual similarity measure for RSA, ratings of classification images from the independent emotion rating task were averaged across raters and recoded such that more negative values indicate similarity (greater resemblance between the two categories) and more positive values indicate dissimilarity (less resemblance between the two categories) (see Methods). Using ratings from the emotion rating task, we regressed perceptual similarity in subjects' classification images (that is, biased resemblance between the two classification images) onto subjects' conceptual similarity value (that is, conceptual overlap between the two categories), which revealed a highly significant effect, $B = 8.2 \times 10^{-4}$, s.e. $= 1.4 \times 10^{-4}$, 95% CI [5.5×10^{-4}, 1.1×10^{-3}], $Z = 5.85$, $P < 0.0001$ (Figs. 2–4). Using ratings from the similarity rating task, we regressed average perceptual similarity ratings onto subjects' conceptual similarity value, which similarly revealed a highly significant effect, $B = 6.6 \times 10^{-4}$, s.e. $= 1.5 \times 10^{-4}$, 95% CI [3.6×10^{-4}, 9.6×10^{-4}], $Z = 4.34$, $P < 0.0001$. Unlike studies 1 and 2, noise ceilings were not computed for these analyses as each subject completed different conditions and there was no estimate of the 'true effect' (that is, group-average DM) that we could use to compute an estimate of the noise ceiling.

Finally, to assess perceptual similarity in a manner that did not rely on subjective judgements from human perceivers, we analysed the physical structure of the classification images themselves by computing the Pearson correlation distance between the flattened pixel maps for each pair of images from each subject in the reverse-correlation task (see Methods). The resulting perceptual DM is presented in Supplementary Fig. 1b. These values were regressed onto each subject's conceptual similarity values, revealing a highly significant effect, $B = 3.4 \times 10^{-5}$, s.e. $= 5.6 \times 10^{-6}$, 95% CI [2.3×10^{-5}, 4.5×10^{-5}], $Z = 6.30$, $P < 0.0001$. Thus, when subjects held a greater conceptual overlap between two emotion categories, they yielded visual prototypes for those categories that exhibited a biased resemblance towards either category, as measured through emotion labelling judgements, pairwise similarity judgements and an objective measure of the physical similarity of the classification images (Fig. 4). These results converge with the results of studies 1 and 2, showing a fundamental correspondence between how emotions are conceptualized and perceived, but here using a more unconstrained task that makes no a priori assumptions about particular stimuli or emotion-related features.

To better assess the consistency and size of the effect of conceptual structure on perceptual structure obtained across studies 1–3, we conducted a local meta-analysis using all data across the three studies relevant for the hypothesis that conceptual knowledge predicts facial emotion perception (see ref. [40]). Specifically, we aggregated data across the three studies in a GEE multilevel model and

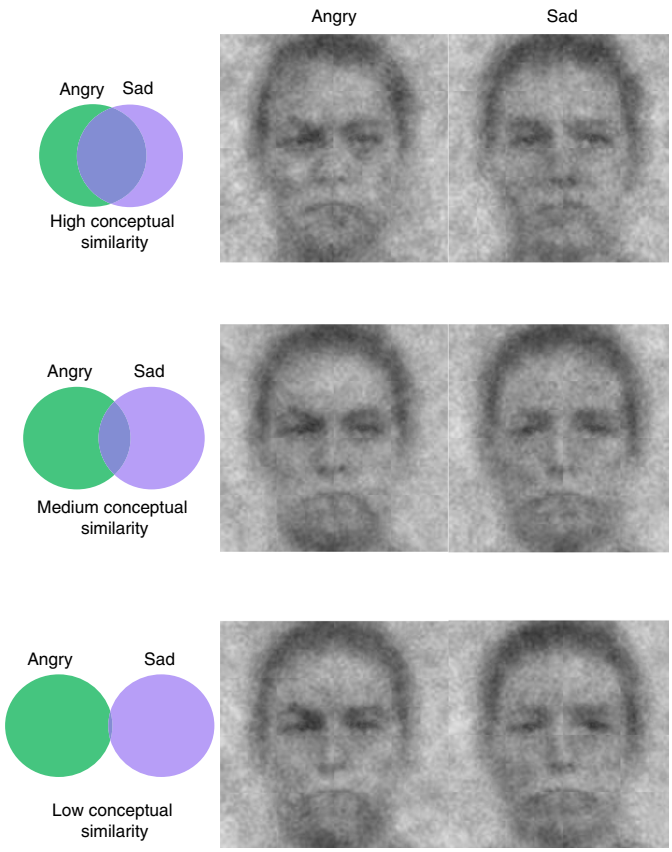

Fig. 4 | Reverse-correlation effects. Reverse-correlation allows an estimation of each subject's visual prototype for a given emotion category. Complete reverse-correlation results for study 3 are presented in Fig. 3. To provide a visual example of the pattern of results for one category pair (Angry–Sad), reverse-correlated classification images are depicted, separately for subjects in tertiles of high, average and low conceptual similarity between the Angry and Sad categories. Subjects with higher conceptual similarity between two categories exhibited a greater resemblance in the appearance of their classification images for those two categories, as assessed by independent ratings of emotion category, independent ratings of subjective perceptual similarity and an objective measure of pixel-based similarity of the classification images themselves.

added an additional nesting factor of study (subjects nested within study), and regressed subjects' perceptual similarity values onto their conceptual similarity values. This indicated a strong and consistent effect across the three studies, $B = 1.8 \times 10^{-2}$, s.e. $= 2.3 \times 10^{-3}$, 95% CI $[1.4 \times 10^{-2}, 2.3 \times 10^{-2}]$, $Z = 7.81$, $P < .0001$. Note that, as the visual similarity controls were only relevant for studies 1 and 2, they were not included in this meta-analysis. Further details on the correspondence across studies may be found in Supplementary Table 3.

Supplementary Figs. 3 and 4 indicate considerable inter-subject variability in perceptual and conceptual similarity for the various category pairs. The multilevel regression analyses presented thus far have all factored by subject, thereby showing that within a given subject, category pairs with higher conceptual similarity values tend to also have higher perceptual similarity values. However, we additionally wanted to more directly assess the role that idiosyncratic differences across subjects have in the relationship between conceptual and perceptual structure. To do so, we conducted an additional set of analyses that instead factor by category pair, thereby aiming to show that, within a given category pair (for example, Anger–Disgust), subjects with higher conceptual similarity values tend to also have higher perceptual similarity values for that category pair. Although such analyses may have limited power due to $N = 15$ (total number of category pairs), they provide a complementary means to assess the overall effects while better characterizing the role of individual differences in similarity values across subjects.

Rerunning the analyses of studies 1–3 factoring by category pair rather than subject indicated that conceptual similarity significantly predicted perceptual similarity in study 1 ($B = 4.2 \times 10^{-3}$, s.e. $= 1.5 \times 10^{-3}$, 95% CI $[1.2 \times 10^{-3}, 7.1 \times 10^{-3}]$, $Z = 2.78$, %NC $= 49.91\%$, $P = 0.0055$), study 2 (at marginal significance: $B = 4.3 \times 10^{-5}$, s.e. $= 2.3 \times 10^{-5}$, 95% CI $[-2.0 \times 10^{-7}, 8.7 \times 10^{-5}]$, $Z = 1.89$, %NC $= 62.79\%$, $P = 0.0591$) and study 3 ($B = 2.9 \times 10^{-4}$, s.e. $= 1.1 \times 10^{-4}$, 95% CI $[7.3 \times 10^{-5}, 5.0 \times 10^{-4}]$, $Z = 2.63$, $P = 0.0085$) (Figs. 2 and 3). Estimates of the noise ceiling for models factored by category pair showed a lower bound on the theoretical maximum possible effect size Z of 5.57 in study 1 and 3.01 in study 2. Note that because the analyses are factored at the level of the stimulus condition (category pair), the visual similarity control of studies 1 and 2 was unnecessary. Conducting an analogous local meta-analysis by aggregating data across the three studies and nesting by an additional factor of study (category pairs nested within study) revealed a strong and consistent effect of conceptual similarity on perceptual similarity across studies, $B = 9.3 \times 10^{-3}$, s.e. $= 3.0 \times 10^{-3}$, 95% CI $[3.5 \times 10^{-3}, 1.5 \times 10^{-2}]$, $Z = 3.16$, $P = 0.0016$. These results complement the primary analyses and suggest that idiosyncratic differences in conceptual knowledge across subjects manifest in facial emotion perception.

To provide an even more conservative test that individual differences in conceptual similarity predict perceptual similarity, we conducted an additional set of analyses nested within category pair. We wished to control for the possibility that individual differences in conceptual and perceptual similarity reflected differences in magnitude or scale in these variables (that is, some subjects have higher or lower similarity values overall, or less or more dispersion, across all 15 category pairs), rather than relative differences in the ordering of which emotion category pairs are more or less similar across subjects. To control for this possibility, we rank ordered each subject's conceptual similarity and perceptual similarity variables in studies 1 and 2. (This approach was not undertaken for study 3 as only one conceptual similarity variable was collected per subject, making rank ordering impossible.) Thus, these analyses isolate inter-subject variability in the form of relative ordering of which emotion category pairs are more or less similar and remove inter-subject variability in the form of any absolute differences in similarity values' magnitude and scale.

We regressed rank-ordered perceptual similarity onto rank-ordered conceptual similarity using a GEE multinomial logistic regression, nesting by emotion category pair. These analyses indicated that, within a given emotion category pair, subjects whose conceptual structure placed that emotion category pair at a higher rank (that is, more conceptual similarity), were significantly more likely to have that emotion category pair more highly ranked in their perceptual structure in study 1 ($B = 8.9 \times 10^{-2}$, s.e. $= 1.9 \times 10^{-2}$, 95% CI $[5.1 \times 10^{-2}, 0.127]$, $Z = 4.57$, %NC $= 42.43\%$, $P < 0.0001$) and in study 2 ($B = 5.1 \times 10^{-2}$, s.e. $= 1.8 \times 10^{-2}$, 95% CI $[1.6 \times 10^{-2}, 8.6 \times 10^{-2}]$, $Z = 2.86$, %NC $= 44.06\%$, $P = 0.0042$). Estimates of the noise ceiling for these analyses showed a theoretical maximum possible effect size Z of 10.77 in study 1 and 6.49 in study 2. These analyses show that variance in perceptual similarity values was largely due to inter-subject differences—specifically where each category pair fit into an individual's idiosyncratic conceptual structure. Thus, within each of the 15 emotion category pairs (for example, Anger–Disgust), these results show that the relative ranking of that category pair's conceptual similarity for a given subject predicts the relative ranking of that pair's perceptual similarity for a given subject. If such relative ordering were stable across subjects (no inter-subject variability), these analyses nesting by category pair would not yield a significant relationship.

In sum, across three studies, we observed that the structure of emotion conceptual knowledge manifests in emotion perception. In studies 1 and 2, we showed that idiosyncratic differences in emotion concept knowledge can predict subtle differences in how those emotions are perceived. Specifically, we found that the extent to which any two emotion categories are conceptually more similar in the mind of a subject predicts a stronger co-activation of the two categories and a conceptual-biasing effect during emotion perception. Our results suggest that, even within the same culture, people may perceive identical facial expressions differently based on their individual differences in conceptual knowledge. In study 3, we obtained converging evidence that a greater conceptual similarity between emotion categories predicts a greater resemblance in the facial prototypes of those categories and how they may be visually represented. This was accomplished using a data-driven task that did not rely on emotion linguistic labels, potential artefacts of a given stimulus set or on normative assumptions of how different facial emotion expressions should appear. Together, converging evidence across both the dynamic process of perceiving faces (mouse tracking) and in estimating facial prototypes (reverse correlation) suggests that emotion concept knowledge is a cognitive resource that comes into play during emotion perception, providing a structure for how facial expressions are visually perceived.

Such findings bolster recent models of emotion and person perception[23–25], which suggest that emotion perception is so rapid and flexible because early processing of facial cues activate related top-down expectations that effectively take on some of the visual processing load. This top-down guiding of emotion perception supplements the visual input, in some cases biasing it to be in line with prior expectations. Importantly, we view these results not necessarily as revealing a special role for conceptual knowledge in emotion perception, but rather showing that emotion perception is subject to domain-general characteristics of a dynamic, predictive and interactive perceptual system[41]. These findings dovetail with research on visual perception more generally[42,43], suggesting that emotion perception relies on the same kinds of top-down predictions as those observed in perceiving objects, words, non-emotional social categories and other non-social stimuli alike.

Few studies to date have used RSA to study facial emotion perception specifically, but those that do have also suggested that more abstract conceptual information may affect the perceptual representation of emotion. Neuroimaging work using classification approaches similar to RSA shows a correspondence between the neural representations of valence information from perceived human facial expressions and inferences from situations[44]. More closely related to the current approach, one study measured the representational similarity between emotion categories in their perception from faces and voices[45], showing high correspondence between modalities even when controlling for low-level stimulus features, suggesting that the representational structure of emotion perception may be shaped by more abstract conceptual features. The present work was able to build on these previous approaches, providing an estimate of how well an explicit conceptual model fits the representational space of facial emotion perception, as well as an estimate of how much this conceptual structure may vary between individuals.

One question of interest concerns the origins of individual differences in emotion concept knowledge. In the cross-cultural domain, researchers have interpreted differences in performance on emotion perception tasks as reflecting culture-specific knowledge about the situational and normative factors that constrain facial expressiveness, which is instantiated in subtle 'dialects' in how facial actions are interpreted as instances of emotion[46,47]. Indeed, recent research shows that regardless of explicit customs, normative facial expressions of emotion may simply consist of different configurations of facial actions between cultures[48,49]. However, the present work suggests that there may be more subtle and idiosyncratic within-cultural variability as well, and future work is necessary to directly investigate where these individual conceptual differences come from in the first place. Given the recent increase in interest into the role of conceptual knowledge in emotion perception, ongoing research is working to describe the process by which developing minds acquire emotion concepts. Recent work suggests that the strongest predictor of emotion concept acquisition is a child's verbal development more generally[50]. While speculative, it is possible that subtle differences in conceptual knowledge between individuals could emerge from the nonverbal and verbal information about emotion that is present in the social environment during this sensitive period in conceptual development. Future studies could investigate this possibility directly. Beyond the developmental origins of individual differences in conceptual knowledge, an additional task for future research is to further understand the nature of these differences. One possibility is that individual differences in conceptual structure are due to individual differences in emotion differentiation, emotion granularity or emotional complexity, variables that have primarily been used in affective science to study individual differences in the structure of emotion experience[51–53]. An important question for future research is whether similar constructs shape the structure of emotion perception as well.

The studies are not without their limitations. As the studies are correlational, it is not clear whether conceptual knowledge has a causal relationship with perceptual similarity. For example, while the mouse-tracking technique allows an index of the temporal dynamics of emotion perception, it is difficult to know precisely at what level of representation the mouse-trajectory-deviation effects reflect. An alternative explanation of the correspondence between conceptual knowledge and perceptual similarity (as measured with mouse tracking in studies 1 and 2) could be that the unchosen category response acted as a distractor when it was held as conceptually more similar to the chosen response. While study 3 was designed to be resistant to this particular concern, future work using mouse tracking to estimate models of perceptual similarity could manipulate conceptual knowledge for particular emotions to permit stronger causal claims (for example, combining the paradigms we used with previous 'semantic satiation' procedures) or correlate with additional methodologies (for example, neuroimaging) to better understand at what level of representation such impacts manifest.

Our use of the six 'basic' emotions is also worth discussing. These emotion categories have been the most traditionally studied and their corresponding facial expressions depicted in stimulus sets are unambiguous and extreme. This can differ from naturalistic emotion perception, where facial actions are often more subtle, ambiguous and fleeting[54] (for a discussion, see ref. [55]). In our view, the fact that we are able to demonstrate such top-down effects of conceptual knowledge on clear-cut and canonical facial expressions of the six basic emotions only speaks to the strength of these effects. While such stimuli are perhaps less ecologically valid, our findings speak to the routine integration of conceptual knowledge in emotion perception, and with more naturalistic and ambiguous stimuli we would expect such top-down effects only to be magnified. But one goal of the present work was to show that these idiosyncratic conceptual differences can manifest even in the most unambiguous and traditional categories of emotion.

These findings also likely have implications for social interaction. In each of our studies, a perceiver never saw a face again after placing it into an emotion category, but this of course rarely happens in daily life. Future work could explore whether the structure of conceptual knowledge influences downstream consequences of emotion perception, including affective and behavioural reactions to conceptually shaped emotional expressions. If so, this approach could have both theoretical and practical implications for understanding a variety of emotion-related responses to other people.

Methods

Subjects in all studies were financially compensated and provided informed consent in a manner approved by the New York University Institutional Review Board.

Study 1. *Participants.* Given no strong precedent for calculating sample size, we aimed to collect roughly 100 subjects. One hundred and ten individuals participated in exchange for monetary compensation on Amazon's Mechanical Turk. Ten subjects were excluded for not following instructions correctly on the mouse-tracking task, resulting in a final sample of 100 (62 female; mean age (M_{age}) = 37.25; s.d.$_{age}$ = 10.86; all White, 2 Hispanic).

Stimuli. Stimuli in the mouse-tracking task were 150 colour photographs of posed human emotional expressions from the NimStim stimulus set[56]. To minimize complexity due to the perceptual interdependence between race and emotion[57], we only included images depicting Caucasian individuals. The resulting stimulus set included 25 images for each of the 6 emotion categories (that is, depicting posed emotional expressions corresponding to normatively Angry, Disgusted, Fearful, Happy, Sad and Surprised facial expressions). No single identity was shown more than once per emotion condition.

Procedure. Mouse-tracking data were collected using MouseTracker software[5], implementing a standard two-choice design. On each of 150 trials, subjects clicked a 'Start' button at the bottom centre of the screen to reveal a face stimulus, which stayed on the screen until they chose one of two response options located in either top corner. On each trial, the response options were two emotion categories (for example, Angry, Fearful), one of which corresponded to the posed emotional expression of the face stimulus. Each stimulus was seen only once, and the response options changed on every trial, resulting in a total of 10 trials for each of the 15 possible pairwise combinations of the 6 emotion categories included (for example, Anger–Fear, Surprise–Happiness). Trials were randomized and the position of response options (left/right) was counterbalanced across subjects. After completing the mouse-tracking task, subjects completed a conceptual similarity task in which they made 15 similarity judgements corresponding to each pairwise combination of emotions, on a 10-point scale (for example, "From 1 = not at all to 10 = extremely, how similar do you find the emotions Anger and Fear?").

After completing these tasks, subjects also completed three survey measures for use as potential covariates: the 20-item Toronto Alexithymia Scale (TAS-20[58]), the Positive and Negative Affective Schedule (PANAS[59]), and a single question asking for self-reported political ideology on a scale from 1 = extremely liberal to 7 = extremely conservative. The inclusion of these variables did not change the significance of the reported results, so they are not discussed further.

Data preprocessing. Any trials with response times exceeding 2,000 ms were excluded from analysis. As we were interested in mouse-trajectory deviation towards the unselected category regardless of eventual response (reflecting greater similarity in how a face was perceived between the two emotion categories), exclusions were not made based on eventual response. This also obviates any need to define trials as 'correct' or 'incorrect' simply because they do not conform to an a priori normative category label. For instance, even if the ostensibly correct and normative answer is Angry on a trial where Disgusted was ultimately selected, deviation towards the Angry response still reflects a greater Anger–Disgust co-activation during perception that we aim to measure. Per standard procedures[5], trajectories were normalized into 100 time bins using linear interpolation and rescaled into a coordinate space with [−1.0,1.5] at the top left and [1,0] at the bottom right, leaving [0,0] at the start location. MD of each mouse trajectory towards the unselected response option on the opposite side of the screen was calculated as the maximum perpendicular deviation from an idealized straight line between its start and end point. MD has been well-validated by previous behavioural and neuroimaging research as an index of how much the unselected category response was co-activated during perception[33,60].

Visual controls. To account for the potential contribution of bottom-up overlap in physical features between images in two categories (for example, physical overlap between faces belonging to the Fear and Surprise categories), we included a visual control in our model. To model the similarity between emotion-related facial features, we used FACS[61]. FACS is a widely studied anatomically based technique for quantifying the precise activity of facial musculature (delineated into independent facial AUs, each of which represents the action of one or more specific facial muscles, for example, 'nose wrinkler', 'lip tightener') in facial expressions and dynamic facial movements. An independent certified expert FACS coder, who was blind to our hypotheses, coded each of the 150 stimuli on the presence or absence of 30 AUs. Using this coding, we also assessed how well our stimuli agreed with EMFACS, a subset of FACS meant to isolate only emotionally meaningful facial actions, including codes which the creators of FACS tentatively deemed 'critical' for perception of each basic emotion[62,63]. For each EMFACS code, the same independent certified FACS coder coded each image on which AUs it displayed that were critical for perception of its stimulus category, and whether it displayed any AUs critical for perception of other categories. Across all of the stimuli used in studies 1 and 2, this approach showed an 85.85% agreement between the AUs present in the stimuli and the 'critical' AUs traditionally described by the authors of FACS. Full results for each stimulus category are presented in Supplementary Table 4.

Analytic approach. Our analytic approach was to use multiple regression RSA, predicting perceptual similarity (MD) from conceptual similarity (conceptual ratings), while controlling for potential visual similarity in the stimuli themselves (overlap in facial AUs). Perceptual, conceptual and visual similarity was calculated for each of the pairwise combinations of the six emotion categories (a total of 15 pairwise combinations). Calculating similarity in the present study was intuitive as the primary independent variable (conceptual similarity rating) and dependent variable (MD) were already direct measures of similarity (this differs in study 2). Thus, subjects' rating of how similar two categories are conceptually (for example, Anger and Disgust) served as conceptual similarity, and subjects' trajectory-deviation effect when making a categorization between those two categories served as perceptual similarity. When similarity must be calculated based on patterns of a variable (rather than a single value), various similarity metrics may be used with RSA (for example, Pearson correlation distance, Euclidean distance). However, because multiple regression RSA in particular assumes a linear combination of multiple predictors, squared Euclidean distance (that is, sums of squared differences) is used to meet the linearity assumption[30,64]. Thus, to compute visual similarity, for each of the 15 pairwise combinations, we calculated the squared Euclidean distance between the two categories' 30 FACS AUs, thereby reflecting how dissimilar or similar any pair of two categories' face stimuli are in terms of configurations of all FACS AUs. We additionally computed two more stringent versions of visual similarity: one that captured the pairwise similarity of the stimuli in each condition across all EMFACS (that is, similarity between the conditions in all AUs traditionally defined as critical for Angry, Disgusted, Fearful, Happy, Sad, and Surprised expressions) and one that strictly captured the similarity of each pair of emotion categories on category-relevant EMFACS (for example, for Anger–Disgust, the similarity between the stimuli in these conditions on the specific AUs traditionally defined by EMFACS as critical for Anger and Disgust). All visual control DMs are depicted in Supplementary Fig. 2.

As squared Euclidean distance is a measure of dissimilarity rather than similarity (higher values = dissimilar, lower values = similar), for simplicity and ease of interpretation, we recoded conceptual and perceptual similarity measures to also reflect dissimilarity. In particular, conceptual similarity ratings were recoded such that a rating of 1 denoted maximum similarity and a rating of 10 denoted maximum dissimilarity. MD for all trials within a subject were rescaled to vary between [0,1], such that 0 corresponded to a subject's largest MD (reflecting perceptual similarity) and 1 corresponded to their smallest (reflecting perceptual dissimilarity). MD values were then averaged for each pairwise combination of the categories, resulting in 15 average MD values per subject.

Study 2. *Participants.* One hundred individuals completed the task in exchange for monetary compensation on Amazon's Mechanical Turk (continuing the targeted sample size from study 1). Nine subjects were excluded from analysis due to not following instructions on the mouse-tracking task, resulting in a final sample of 91 (49 female; M_{age} = 40.451; s.d.$_{age}$ = 13.03; all White, 5 Hispanic).

Stimuli. The mouse-tracking task used the same 150 images (25 per emotion category) as those used in Study 1. Word stimuli were generated for the conceptual similarity rating task through a pre-test administered on Mechanical Turk (N = 47). Using a 'feature listing' approach[65], in which the conceptual contents of a category is elucidated by having subjects generate and list features of category exemplars, we asked subjects to "list the top 5 bodily feelings, thoughts or actions" they associate with each of the six emotion categories used in study 1. We then took the top 40 words and phrases that occurred most frequently across all emotions and all subjects and used these as stimuli in the conceptual rating portion of the main experiment (Supplementary Table 5).

Procedure. The mouse-tracking task followed an identical procedure to that in Study 1. The conceptual similarity task followed the mouse-tracking task and required subjects to attend to each emotion category in isolation, for a total of six blocks presented in a randomized order. In each block, subjects rated each of the 40 word and phrase stimuli in a randomized order on how related they were to the category in question, on a 7-point scale (for example, "On a scale from 1 = not to all to 7 = extremely, how related is 'crying' to the emotion Fear?"), for a total of 240 trials. Finally, subjects completed the three survey tasks for use as potential covariates, as in study 1: TAS-20, PANAS and political ideology. As in study 1, including these variables as covariates did not change the pattern of results, so are not discussed further.

Mouse-trajectory preprocessing was conducted in a manner identical to that in study 1, as was the rescaling of MD. As the stimuli used in the mouse-tracking task were the same as those used in study 1, the same FACS-based visual models were also included in the present analysis. Unlike study 1 where we had a direct single-value measure of conceptual similarity, here conceptual similarity was calculated as the squared Euclidean distance between vectors of

responses to the word and phrase stimuli for each emotion category. For example, to measure each subject's conceptual similarity between Anger and Fear, we calculated the squared Euclidean distance between their Anger vector of 40 ratings and Fear vector of 40 ratings (items such as 'crying', 'clenching fists' and so on; see Supplementary Table 5).

Study 3. *Participants.* As reverse-correlation tasks require a large number of trials per condition, it was unfeasible to have a single subject complete the task for all 6 emotion categories (which would total 1,200 trials). Instead, we randomly assigned subjects to 1 of 15 conditions corresponding to each pairwise combination of emotions (for example Anger–Fear, Disgust–Happiness). Analyses would be aggregated across these conditions, but for adequate representativeness of all pairwise combinations we recruited 25 subjects per condition, leading to a total sample size of 375. The 375 subjects completed the initial reverse-correlation task in exchange for monetary compensation on Mechanical Turk. Seven subjects were excluded for not following instructions, resulting in a final sample of 368 (234 female; M_{age} = 37.35; s.d.$_{age}$ = 12.37; 74.7% White, 8.97% Black, 7.34% Asian, 8.97% other). Once reverse-correlated images were generated for the subjects, a separate group of independent raters (N = 95) were recruited to rate their images on emotion category (48 female; M_{age} = 36.38; s.d.$_{age}$ = 11.19; 72.63% White, 11.58% Black, 8.42% Asian, 7.37% other). Due to a technical error, classification images from 15 subjects (that is, 30 classification images) were not properly presented to these independent raters and thus were excluded from the final analysis. A second separate group of independent raters (N = 92) were recruited to provide pairwise similarity ratings (40 female; M_{age} = 38.58; s.d.$_{age}$ = 13.52; 69.56% White, 15.22% Black, 9.78% Asian, 5.44% other).

Stimuli. The base face used in the reverse-correlation task was the same image used in the first published study to use the reverse-correlation technique[6], which was also used in our recent work taking a similar approach[66]. The image is the average neutral male face from the Karolinska Directed Emotional Faces database[67].

Procedure. The initial reverse-correlation task followed the standard procedure implemented in other studies[6,66,68]. Each subject in the task was randomly assigned to 1 of 15 conditions corresponding to one of the possible pairwise combinations of the 6 emotion categories (for example, Anger and Disgust). Participants in each condition classified each of the two emotion categories in the reverse-correlation task, and also provided conceptual similarity data for these two categories. On each trial in the reverse-correlation task, subjects were presented with two side-by-side faces obscured with different patterns of randomly generated visual noise. The task was split into 200 trials per emotion category. For a subject in the Anger–Disgust condition, this would require them to complete 200 trials where they were instructed to "Choose the face that looks more Angry" and 200 trials where they were instructed to "Choose the face that looks more Disgusted" (order was counterbalanced). Participants also completed the conceptual similarity rating task (40 word and phrase stimuli) used in study 2. However, instead of completing ratings for all six emotion categories, they only completed ratings for the two emotion categories in their randomly assigned condition. Due to the existing length of the task and given that they did not influence the results of studies 1 and 2, we did not include the previous survey measures in this study.

Following the standard data preprocessing approach for reverse-correlation research[68], we averaged the face selected on each trial for each subject, resulting in classification images for each category that each subject classified (that is, for each subject in the Anger–Disgust condition, we would have subject-specific classification images produced for the Anger and Disgust categories). Across all subjects, this resulted in a total of 736 classification images.

In the first independent rating task, independent groups of subjects rated the generated classification images on emotion category labels. Due to the large number of classification images, we split the classification images into three groups corresponding to a randomly chosen subset of five of the fifteen conditions from the reverse-correlation task. Three separate groups of raters rated the images in a series of five randomized blocks. For example, one rater might have to rate classification images from the Anger–Disgust, Fear–Sadness, Happiness–Surprise, Disgust–Fear and Surprise–Fear conditions. Raters were instructed to rate the images on a 7-point scale that ranged from one emotion category to another, where the categories on either end of the scale corresponded to the condition to which the initial subject was assigned. For example, if a rater was judging an image that was produced by a subject in the Anger–Disgust condition, they would rate the image on a scale from 1 = Angry to 7 = Disgusted. We chose to use bipolar scales to measure independent ratings as we were interested in the relative effects between the two emotion categories depending on conceptual overlap. Since individual classification images can be relatively noisy and ambiguous, we were concerned about the possibility of floor effects if we had given raters separate unipolar scales with a 'neutral' option.

In the second independent rating task, independent groups of subjects rated pairs of the generated classification images on perceptual similarity. For consistency with the last independent rating task, we once again split the classification images into three groups of randomly chosen conditions from the reverse-correlation task. In this case, raters completed all ratings within one block, as response options did not change throughout the task. Raters saw pairs of classification images (where each pair corresponded to the two classification images produced by one subject in the reverse-correlation task) and were instructed to "Rate how similar the two images appear" on a 7-point scale that ranged from "Not at all similar" to "Extremely similar".

Analytic approach. We aimed to assess the relationship between each individual's conceptual associations between emotion categories (conceptual similarity) and how their classification images appeared (perceptual similarity). To do so, we conducted a regression analysis testing whether the degree of overlap in conceptual knowledge between a given pair of emotion categories (for example, conceptual similarity between Anger and Disgust) reliably predicted the extent to which their classification images were biased in physical appearance (for example, for an Anger–Disgust subject, how Disgusted their Angry classification image appeared and how Angry their Disgusted classification image appeared). Consistent with study 2, conceptual similarity was measured as the squared Euclidean distance between the 40 ratings for the two emotion categories in question. Independent ratings for each subject's two classification images served as the dependent measure of perceptual similarity. Emotion ratings for a given classification image were averaged across raters and recoded such that −3 indicated a maximally biased appearance (for example, an Angry classification image rated as Disgusted) and 3 indicated a maximally unbiased response (for example, an Angry classification image rated as Angry). Because higher values in squared Euclidean distance reflect dissimilarity and lower values reflect similarity, as in studies 1 and 2 this recoding ensured our measure of perceptual similarity was consistent (higher values = dissimilarity). With this approach, we expected a positive relationship between conceptual similarity and perceptual similarity, such that subjects with low conceptual similarity between emotion categories would produce classification images without biased appearance, and subjects with high conceptual overlap would produce classification images with biased appearance.

To analyse the relationship between conceptual similarity and ratings of subjective similarity, ratings of similarity for a given pair of classification images were averaged across raters. To increase interpretability of results, these ratings were also recoded such that −3 indicated maximum similarity (that is, pairs of classification images rated as 'Extremely similar') and 3 indicated maximum dissimilarity (that is, pairs of classification images rated as 'Not at all similar'). To analyse the relationship between conceptual similarity and objective physical similarity between classification images produced by each subject, classification images were read into MATLAB using the imread function and the resulting pixel intensity maps were flattened. The Pearson correlation distance ($1 - r$) was computed for each pair of images from subjects in the reverse-correlation task. As with independent ratings of subjective similarity, the resulting values were regressed onto subjects' conceptual similarity values (squared Euclidean distance).

Reporting Summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Code availability. Custom code used to produce analyses in this manuscript is available and hosted by the Open Science Framework (http://osf.io/q8yhs).

Data availability. Data and materials for all studies are available and hosted by the Open Science Framework (http://osf.io/q8yhs).

Received: 5 February 2018; Accepted: 14 June 2018;
Published online: 23 July 2018

References

1. Gendron, M., Lindquist, K. A., Barsalou, L. & Barrett, L. F. Emotion words shape emotion percepts. *Emotion* **12**, 314–325 (2012).
2. Lindquist, K. A., Barrett, L. F., Bliss-Moreau, E. & Russell, J. A. Language and the perception of emotion. *Emotion* **6**, 125–138 (2006).
3. Lindquist, K. A., Gendron, M., Barrett, L. F. & Dickerson, B. C. Emotion perception, but not affect perception, is impaired with semantic memory loss. *Emotion* **14**, 375–387 (2014).
4. Nook, E. C., Lindquist, K. A. & Zaki, J. A new look at emotion perception: concepts speed and shape facial emotion recognition. *Emotion* **15**, 569–578 (2015).
5. Freeman, J. B. & Ambady, N. MouseTracker: software for studying real-time mental processing using a computer mouse-tracking method. *Behav. Res. Methods* **42**, 226–241 (2010).
6. Dotsch, R., Wigboldus, D. H. J., Langner, O. & van Knippenberg, A. Ethnic out-group faces are biased in the prejudiced mind. *Psychol. Sci.* **19**, 978–980 (2008).

7. Ekman, P. Universals and cultural differences in facial expressions of emotion. In *Nebraska Symposium on Emotion and Motivation, 1971* (ed. Cole, J.) 207–283 (Univ. Nebraska Press, Lincoln, NE, 1972).
8. Ekman, P. & Cordaro, D. What is meant by calling emotions basic. *Emot. Rev.* **3**, 364–370 (2011).
9. Adolphs, R. Neural systems for recognizing emotion. *Curr. Opin. Neurobiol.* **12**, 169–177 (2002).
10. Adolphs, R. Cognitive neuroscience of human social behaviour. *Nat. Rev. Neurosci.* **4**, 165–178 (2003).
11. Darwin, C. *The Expression of the Emotions in Man and Animals* (Oxford Univ. Press, New York, NY, 1872).
12. Ekman, P. Facial expressions of emotion: new findings, new questions. *Psychol. Sci.* **3**, 34–38 (1992).
13. Ekman, P. Facial expression and emotion. *Am. Psychol.* **48**, 384–392 (1993).
14. Smith, M. L., Cottrell, G. W., Gosselin, F. & Schyns, P. G. Transmitting and decoding facial expressions. *Psychol. Sci.* **16**, 184–189 (2005).
15. Tracy, J. L. & Robins, R. W. The automaticity of emotion recognition. *Emotion* **8**, 81–95 (2008).
16. Ambady, N. & Weisbuch, M. in *Oxford Handbook of Face Perception* (eds Calder, A. J., Rhodes, G., Haxby, J. V. & Johnson, M. H.) 479–488 (Oxford Univ. Press, Oxford, 2011).
17. Hassin, R. R., Aviezer, H. & Bentin, S. Inherently ambiguous: facial expressions of emotions, in context. *Emot. Rev.* **5**, 60–65 (2013).
18. Aviezer, H., Hassin, R., Bentin, S. & Trope, Y. in *First Impressions* (eds Ambady, N. & Skowronski, J.) 255–286 (Guilford Press, New York, NY, 2008).
19. Aviezer, H. et al. Not on the face alone: perception of contextualized face expressions in Huntington's disease. *Brain* **132**, 1633–1644 (2009).
20. Aviezer, H., Dudarev, V., Bentin, S. & Hassin, R. R. The automaticity of emotional face-context integration. *Emotion* **11**, 1406–1414 (2011).
21. Meeren, H. K. M., van Heijnsbergen, C. C. R. J. & de Gelder, B. Rapid perceptual integration of facial expression and emotional body language. *Proc. Natl Acad. Sci. USA* **102**, 16518–16523 (2005).
22. Van den Stock, J., Righart, R. & de Gelder, B. Body expressions influence recognition of emotions in the face and voice. *Emotion* **7**, 487–494 (2007).
23. Barrett, L. F. The theory of constructed emotion: an active inference account of interoception and categorization. *Soc. Cogn. Affect. Neurosci* **12**, 1–23 (2017).
24. Freeman, J. B. & Ambady, N. A dynamic interactive theory of person construal. *Psychol. Rev.* **118**, 247–279 (2011).
25. Lindquist, K. A. Emotions emerge from more basic psychological ingredients: a modern psychological constructionist approach. *Emot. Rev.* **5**, 356–368 (2013).
26. Carroll, N. C. & Young, A. W. Priming of emotion recognition. *Q. J. Exp. Psychol. A* **58**, 1173–1197 (2005).
27. Doyle, C. M. & Lindquist, K. A. When a word is worth a thousand pictures: language shapes perceptual memory for emotion. *J. Exp. Psychol. Gen* **147**, 62–73 (2018).
28. Fugate, J. M. B., Gendron, M., Nakashima, S. & Barrett, L.F. Emotion words: adding face value. *Emotion* (in the press).
29. Kriegeskorte, N., Mur, M. & Bandettini, P. Representational similarity analysis—connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* **2**, 4 (2008).
30. Stolier, R. M. & Freeman, J. B. Neural pattern similarity reveals the inherent intersection of social categories. *Nat. Neurosci.* **19**, 795–797 (2016).
31. Skerry, A. E. & Saxe, R. Neural representations of emotion are organized around abstract event features. *Curr. Biol.* **25**, 1945–1954 (2015).
32. Freeman, J. B. Doing psychological science by hand. *Curr. Dir. Psychol. Sci.* (in the press).
33. Stolier, R. M. & Freeman, J. B. A neural mechanism of social categorization. *J. Neurosci.* **37**, 5711–5721 (2017).
34. Liang, K.-Y. & Zeger, S. L. Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22 (1986).
35. Nili, H. et al. A toolbox for representational similarity analysis. *PLoS Comput. Biol.* **10**, e1003553 (2014).
36. Khaligh-Razavi, S.-M., Henriksson, L., Kay, K. & Kriegeskorte, N. Fixed versus mixed RSA: explaining visual representations by fixed and mixed feature sets from shallow and deep computational models. *J. Math. Psychol.* **76B**, 184–197 (2017).
37. Barsalou, L. W. Grounded cognition. *Annu. Rev. Psychol.* **59**, 617–645 (2008).
38. Chanes, L., Wormwood, J. B., Betz, N. & Barrett, L. F. Facial expression predictions as drivers of social perception. *J. Pers. Soc. Psychol.* **114**, 380–396 (2018).
39. Lieberman, M. D. in *Social Neuroscience: Toward Understanding the Underpinnings of the Social Mind* (eds Todorov, A., Fiske, S. T. & Prentice, D.) 188–209 (Oxford Univ. Press, Oxford, 2011).
40. Riley, R. D., Lambert, P. C. & Abo-Zaid, G. Meta-analysis of individual subject data: rationale, conduct, and reporting. *Br. Med. J.* **340**, c221 (2010).
41. Freeman, J. B. & Johnson, K. L. More than meets the eye: split-second social perception. *Trends Cogn. Sci.* **20**, 362–374 (2016).
42. O'Callaghan, C., Kveraga, K., Shine, J. M., Adams, R. B. & Bar, M. Predictions penetrate perception: converging insights from brain, behavior and disorder. *Conscious. Cogn.* **47**, 63–74 (2017).
43. Summerfield, C. & Egner, T. Expectation (and attention) in visual cognition. *Trends Cogn. Sci.* **13**, 403–409 (2009).
44. Skerry, A. E. & Saxe, R. A common neural code for perceived and inferred emotion. *J. Neurosci.* **34**, 15997–16008 (2014).
45. Kuhn, L. K., Wydell, T., Lavan, N., McGettigan, C. & Garrido, L. Similar representations of emotions across faces and voices. *Emotion* **17**, 912–937 (2017).
46. Elfenbein, H. A. Nonverbal dialects and accents in facial expressions of emotion. *Emot. Rev.* **5**, 90–96 (2013).
47. Elfenbein, H. A., Beaupré, M., Lévesque, M. & Hess, U. Toward a dialect theory: cultural differences in the expression and recognition of posed facial expressions. *Emotion* **7**, 131–146 (2007).
48. Jack, R. E., Caldara, R. & Schyns, P. G. Internal representations reveal cultural diversity in expectations of facial expressions of emotion. *J. Exp. Psychol. Gen.* **141**, 19–25 (2012).
49. Jack, R. E., Garrod, O. G. B., Yu, H., Caldara, R. & Schyns, P. G. Facial expressions of emotion are not culturally universal. *Proc. Natl Acad. Sci. USA* **109**, 7241–7244 (2012).
50. Nook, E. C., Sasse, S. F., Lambert, H. K., McLaughlin, K. M. & Somerville, L. H. Increasing verbal knowledge mediates the development of multidimensional emotion representation. *Nat. Hum. Behav.* **1**, 881–889 (2017).
51. Barrett, L. F., Gross, J., Christensen, T. C. & Benvenuto, M. Knowing what you're feeling and knowing what to do about it: mapping the relation between emotion differentiation and emotion regulation. *Cogn. Emot.* **15**, 713–724 (2001).
52. Kang, S. M. & Shaver, P. R. Individual differences in emotional complexity: their psychological implications. *J. Pers.* **72**, 687–726 (2004).
53. Lindquist, K. A. & Barrett, L. F. in *Handbook of Emotions* 3rd edn (eds Lewis, M., Haviland-Jones, J. M. & Barrett, L. F.) 513–530 (Guilford, New York, NY, 2008).
54. Russell, J. A., Bachorowski, J. A. & Fernandez-Dols, J. M. Facial and vocal expressions of emotion. *Annu. Rev. Psychol.* **54**, 329–349 (2003).
55. Barrett, L. F., Mesquita, B. & Gendron, M. Context in emotion perception. *Curr. Dir. Psychol. Sci.* **20**, 286–290 (2011).
56. Tottenham, N. et al. The NimStim set of facial expressions: judgments from untrained research subjects. *Psychiatry Res.* **168**, 242–249 (2009).
57. Hugenberg, K. & Bodenhausen, G. V. Ambiguity in social categorization: the role of prejudice and facial affect in race categorization. *Psychol. Sci.* **15**, 342–345 (2004).
58. Bagby, R. M., Parker, J. D. A. & Taylor, G. J. The twenty-item Toronto Alexithymia scale—I. Item selection and cross-validation of the factor structure. *J. Psychosom. Res.* **38**, 23–32 (1994).
59. Watson, D., Clark, L. A. & Tellegen, A. Development and validation of brief measures of positive and negative affect: the PANAS scales. *J. Pers. Soc. Psychol.* **54**, 1063–70 (1988).
60. Freeman, J. B., Dale, R. & Farmer, T. A. Hand in motion reveals mind in motion. *Front. Psychol.* **2**, 59 (2011).
61. Ekman, P. & Friesen, W. V. *Facial Action Coding System: A Technique For The Measurement Of Facial Movement* (Consulting Psychologists Press, Palo Alto, CA, 1978).
62. Ekman, P., Irwin, W. & Rosenberg, E. *EMFACS: Coders Instructions (EMFACS-8)* (Univ. California San Francisco Press, San Francisco, CA, 1994).
63. Ekman, P., Friesen, W. & Hager, J. *Facial Action Coding System: Investigator's Guide* 2nd edn (Research Nexus eBook, Salt Lake City, UT, 2002).
64. Carlin, J. D. & Kriegeskorte, N. Adjudicating between face-coding models with individual-face fMRI responses. *PLoS Comput. Biol.* **13**, e1005604 (2017).
65. Barsalou, L.W. & Hale, C.R. in *Categories and Concepts: Theoretical Views and Inductive Data analysis* (eds Van Mechelen, I., Hampton, J., Michalski, R. & Theuns, P.) 97–144 (Academic Press, San Diego, CA, 1993).
66. Brooks, J. A., Stolier, R. M. & Freeman, J. B. Stereotypes bias visual prototypes for sex and emotion categories. *Soc. Cogn.* (in the press).
67. Lundqvist, D. & Litton, J. E. *The Averaged Karolinska Directed Emotional Faces—AKDEF (CD ROM)* (Karolinska Institute, Stockholm, 1998).
68. Dotsch, R. & Todorov, A. Reverse correlating social face perception. *Soc. Psychol. Personal. Sci.* **3**, 562–571 (2012).

Acknowledgements

We thank L.I. Reed for assistance with FACS coding. This work was supported in part by research grant NIH-R01-MH112640 (J.B.F.). The funders of this research had no role in the conceptualization, design, data collection, analysis, decision to publish or preparation of the manuscript.

Author contributions

Both authors collaborated on the study concept, design, and interpretation of the data. J.A.B. collected and analysed the data. J.A.B. drafted the manuscript and J.B.F. provided critical revisions. Both authors approved the final version of the manuscript for submission.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41562-018-0376-6.

Correspondence and requests for materials should be addressed to J.A.B. or J.B.F.

nature research

Corresponding author(s): Jeffrey A. Brooks; Jonathan B. Freeman

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	Data were collected using MouseTracker software (Freeman & Ambady, 2010; Behavior Research Methods) and web-based surveys implemented on Mechanical Turk.
Data analysis	In Studies 1 and 2, mouse-tracking relevant variables (i.e. maximum deviation) were extracted from the data using the Analyzer feature in MouseTracker software, version 2.84 . Stimuli for the reverse-correlation task (Study 3), and classification images computed for this task, were created using the RCICR package in R, version 1.3.4 (http://www.rondotsch.nl/rcicr/). Pixel intensity maps for the classification images were computed using the imread function in MATLAB version 2017b. All regression analyses were run using the GENMOD procedure in SAS version 9.4.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All materials and data analyzed for this study are available through an Open Science Framework repository, https://osf.io/q8yhs/.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	The three studies reported in this manuscript utilized quantitative methods.
Research sample	The research sample consisted of workers from Amazon's Mechanical Turk. While not a fully representative sample, the participants are more diverse in age, race, and socioeconomic status than typical undergraduate research samples (Buhrmester, Kwang, & Gosling, 2011; Perspectives on Psychological Science).
Sampling strategy	All studies used random sampling from Mechanical Turk users. Given no strong precedent for sample size in study 1, we aimed to collect roughly 100 participants. We collected 110 participants and 10 participants were excluded for not following instructions, ending in a final sample of 100. In study 2, we replicated the effects of study 1 and continued the target sample size of 100. Nine participants were excluded for not following instructions, resulting in a final sample of 91. The task in sample 3 would have been prohibitively long (potentially impacting data quality) if all participants completed all conditions, so each subject was randomly assigned to one of 15 conditions. In order to have an adequate number of subjects per condition, we followed the precedent set by previous reverse correlation work and aimed to collect data from 25 individuals per condition, leading to a sample of 375. Seven participants were excluded for not following instructions, resulting in a final sample of 368.
Data collection	All data were collected in computer-based experiments, which measured participants responses, response times, and (in Studies 1 and 2), mouse trajectories.
Timing	Study 1 data was collected from June 27 - July 1, 2016. Study 2 data was collected on January 11, 2017. Study 3 data was collected from November 27, 2017 - March 24, 2018.
Data exclusions	In all 3 studies, participants were excluded for patterns of performance that suggested they were not paying attention. On the mouse-tracking tasks (Studies 1 and 2) this was quantified by a participant's failure on an attention check built into the task. On tasks using Likert scales (Studies 1, 2, and 3) participants were deemed to not be following instructions if they used only one response option on the scale for most of the duration of the task. As a result, 10 participants were excluded from Study 1; 9 participants were excluded from Study 2; and 7 participants were excluded from Study 3.
Non-participation	No participants dropped out or declined participation.
Randomization	In Studies 1 and 2, no experimental groups were created. In Study 3, participants were randomly assigned to one of fifteen conditions through a randomization procedure built into the experimental protocol.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	Study 1: n = 100 100 (62 female; Mage = 37.25; SDage = 10.86; all White, 2 Hispanic); Study 2: n = 91 (49 female; Mage = 40.451; SDage = 13.03; all White, 5 Hispanic); Study 3: reverse correlation task - n = 368 (234 female; Mage = 37.35; SDage = 12.37; 74.7% White, 8.97% Black, 7.34% Asian, 8.97% Other); emotion rating task - n = 95 (48 female; Mage = 36.38; SDage = 11.19; 72.63% White, 11.58% Black, 8.42% Asian, 7.37% Other); similarity rating task - n = 92 (40 female; Mage = 38.58; SDage = 13.52; 69.56% White, 15.22% Black, 9.78% Asian, 5.44% Other).
Recruitment	Participants were recruited from Amazon's Mechanical Turk user base. The only restrictions placed on the sample were age (above 18), nationality (born and raised in the United States), and an "approval rate" (indicating that the participant pays attention and follows instructions correctly in tasks) of over 90%. While all participants are self-selected due to interest and motivation to participate in research studies, this is unlikely to introduce bias into the sample since the participants are more diverse in age, race, and socioeconomic status than typical undergraduate research samples (Buhrmester, Kwang, & Gosling, 2011; Perspectives on Psychological Science), and studies have shown that MTurk workers provide high-quality data that replicates many classic findings in experimental psychology (Piolacci & Chandler, 2014; Current Directions in Psychological Science).