American Psychological Association

Emotion

1528-3542/20/$12.00

2020, Vol. 20, No. 7, 1165–1184
http://dx.doi.org/10.1037/emo0000623

Emotional Resemblance: Perception of Facial Emotion in Written English

Max Weisbuch, Jeremy R. Reynolds, Sarah Lamer, and Masako Kikuchi
University of Denver

Toko Kiyonari
Aoyama Gakuin University

Written language is comprised of simple line configurations (i.e., letters) that, in theory, elicit affect by virtue of the concepts they symbolize, rather than their physical features. However, we propose that the line configurations that comprise letters vary in their visual resemblance to canonical features of facial emotion and, through such *emotional resemblance,* influence affective responses to written language. We first describe our data-driven approach to indexing emotional resemblance in each letter according to its visual signature. This approach includes cross-cultural validation and neural-network modeling. Based on the resulting weights, we examine the extent to which emotional resemblance in Latin letters is incidentally processed in a flanker paradigm (Study 1), shapes unintentional affective responses to letters (Study 2), accounts for affective responses to orthographically controlled letter strings (Study 3), and shapes affective responses to real English words (Study 4). Results were supportive of hypotheses. We discuss mechanisms, limitations, and implications.

Keywords: social vision, facial expressions, emotion, visual word perception

Supplemental materials: http://dx.doi.org/10.1037/emo0000623.supp

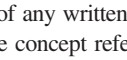
Written language can be emotionally powerful, and this emotional influence is typically attributed to the conceptual meaning of words, rather their visual appearance. For example, attitude researchers often assume that concepts like <LOVE> or <MUSLIM> drive affective responses to words such as "love" and "Muslim." Yet written words are visual stimuli and, like other such stimuli, might generate affect by virtue of their visual signatures, independent of their conceptual meanings. For example, geometric shapes and household products that visually resemble angry faces are evaluated negatively (Aggarwal & McGill, 2007; Ichikawa, Kanazawa, & Yamaguchi, 2011; Landwehr, McGill, & Herrmann, 2011; Watson, Blagrove, Evans, & Moore, 2012; Windhager et al., 2008) and curvy objects are evaluated positively (Bar & Neta, 2006). Such effects are rooted in the visual signature of these objects (e.g., ∇), rather than their conceptual identity (e.g., <TRIANGLE>; e.g., Watson et al., 2012). The visual signature of any written word may likewise generate affect independent of the concept referenced by that visual signature. We propose that visual signatures of Latin letters resemble affectively meaningful objects and through such *emotional resemblance*, quickly shape readers' affective responses to written words. More broadly, we hypothesize that affective processes are sensitive to the pictorial meaning of written language.

This article was published Online First July 1, 2019.

Max Weisbuch, Jeremy R. Reynolds, Sarah Lamer, Masako Kikuchi, Department of Psychology, University of Denver; Toko Kiyonari, School of Social Informatics, Aoyama Gakuin University.

Correspondence concerning this article should be addressed to Max Weisbuch, Department of Psychology, University of Denver, 2155 S Race Street, Denver, CO 80210. E-mail: max.weisbuch@du.edu

A Historical Perspective on Pictures and Text in Visual Communication

Ancient history and modern methods of communication both support the idea that people are sensitive to pictorial meaning in written language. People have always used pictorial imagery to communicate: The earliest records of pictorial imagery coincide with the earliest records of the human species. More recently, ancient Egyptians and Mesoamericans shared the intuition to use pictorial imagery to systematically communicate symbolic thought. In fact, modern alphabetic letters, including the Latin letters on this page, can be traced from (1) ancient hieroglyphics through (2a) heriatics and (2b) protosinaitic to (3) phoenician and finally, to (4) the letters you see in front of you (cf. Fischer, 2001; Gelb, 1963; Trigger, 1998). For example, it is thought that the Latin letter *m* can be traced to the hieroglyph  for "water" or "wave" (Pflughaupt, 2007; Sacks, 2003). Indeed, most writing systems across the world and over time appear to have been preceded by visual communication systems in which people write and read via pictures (Fischer, 2001, 2003).

We propose that human sensitivity to pictorial meaning during reading was not lost with the ancient Egyptians but instead persists in the visual perception of modern writing. Specifically, we suspect that affective processes may be especially sensitive to pictorial meaning in words. Indeed, affective processes are exceptionally sensitive to visual input such that people have lightning-fast affective responses to pictorial stimuli (Calvo & Nummenmaa, 2007; Codispoti, Bradley, & Lang, 2001; Dimberg, Thunberg, & Elmehed, 2000; Murphy & Zajonc, 1993; Weisbuch & Ambady, 2008). These processes are effortless (Fenske & Eastwood, 2003; Horstmann, Borgstedt, & Heumann, 2006; McAndrew, 1986; Maratos, Mogg, & Bradley, 2008) and appear fast enough to occur ahead of the speedy process of word identification (Yap, Balota,

Sibley, & Ratcliff, 2012). Indeed, with an increased emphasis on speed in modern text-based communication (e.g., e-mails, text messaging), humans are increasingly conveying and "reading" emotion through pictorial strategies. Yet beyond the modern and circumscribed usage of emoticons, letters and words may more broadly impact readers' affect by virtue of their visual signatures. Specifically, affective processes more typically observed in the visual perception of pictures (e.g., faces, animals, scenes) may operate in text perception. Like ancient Egyptian scribes responding to picture-based text, we propose that modern humans derive affective meaning from the pictorial content of letters and (thus) words. In support of this hypothesis, we next provide a brief overview of relevant empirical research.

Affective Responses to Images and Words

Mere milliseconds of exposure to an image of an angry face, spider, or other negative stimulus is sufficient to cause perceivers negative affect (Duckworth, Bargh, Garcia, & Chaiken, 2002; Globisch, Hamm, Esteves, & Öhman, 1999; Murphy & Zajonc, 1993). Affective processes thus operate quickly on the complex visual content of objects. Similarly, the complex visual content of written words may quickly shape human affective responses to those words. With a few notable exceptions (see below), however, it is more typically assumed that positive or negative responses to written words instead reflect the semantic meaning of those words. For example, positive and negative responses to words are often used to index attitudes or implicit associations between concepts and affect (Bargh, Chaiken, Raymond, & Hymes, 1996; Fazio, Sanbonmatsu, Powell, & Kardes, 1986; Greenwald, McGhee, & Schwartz, 1998).

In this context, it is noteworthy that perceivers' affective responses to written words can occur before perceivers become aware of a word's semantic meaning (Klinger, Burton, & Pitts, 2000; see also, Kouider & Dehaene, 2007). There is some evidence that such effects owe not to semantic meaning, but rather affect associated with the visual appearance of letter strings. For example, participants in one study (Abrams & Greenwald, 2000) were first asked to classify the affective meaning of unambiguously positive and negative words during a training phase. This phase presumably strengthened an association between the individual letters (or letter strings) and the affect generated by the words in which they appeared, so strings such as "tu" (tulip) and "mor" (humor) became increasingly associated with positive affect. In a second phase, participants completed an affective priming task in which masked prime words such as "tumor" and "smile" were generated by combining letter strings from the training task (e.g., tulip + humor = tumor, smut + bile = smile). Participants had more positive responses to masked words such as "tumor" than words such as "smile," reversing the pattern expected by a semantic meaning approach (smiles seem more pleasant than tumors). In this study, meaningless word-parts (e.g., "tu") served as a visual stimulus for eliciting affect and such effects appeared to overshadow the semantic meaning of masked words. Less controversially, these and other findings (see Kouider & Dehane, 2007) raise the possibility that speedy affective responses to written words derive not *only* from semantic meaning but—at least in part—from the visual properties that uniquely characterize any written word.

Wholly independent of the meaning of words in which they appear, letters may elicit affective responses because they resemble affectively meaningful objects. Specifically, each Latin letter carries with it a standard visual appearance that allows it to be uniquely identified across textual manipulations, such as manipulations due to *italics,* bold, font type, and visual contrast (e.g., *b,* **b,** b, b, and b), and this appearance may therefore be considered a letter's visual signature. The shape and features of any given letter may be present in other frequently observed positive or negative objects, such that the visual signature of any letter can be characterized for its resemblance to other affective stimuli (e.g., angry faces, spiders). Subsequently, each word in a dictionary could be characterized by the degree to which its letters resemble affective stimuli. Any written word may thus elicit affect from readers because its letters resemble positive or negative objects (*emotional resemblance*), because its semantic meaning is positive or negative, or because of both effects. Put differently, the affective power of written language may owe in part to the visual signature of specific letters and words.

The Current Research: Emotional Resemblance in Latin Letters

In the current research, we examined whether affective responses to written words can reflect the extent to which component letters resemble affective stimuli. As the first investigation of emotional resemblance in written language, we focused on resemblance to an especially powerful and well-learned affective stimulus: facial emotion. There were several reasons for this choice. First, prior research suggests that adults' visual sensitivity to facial emotion is exceptional. For example, when presented so quickly as to be impossible to subjectively identify (e.g., 1/300th of a second), images of facial joy and facial anger evoke positive and negative affective responses, respectively (Dimberg et al., 2000; Murphy & Zajonc, 1993). Even as compared with the attention-grabbing power of neutral faces, there is evidence that emotional faces are prioritized in visual perception (Adams, Gray, Garner, & Graf, 2010; Alpers & Gerdes, 2007; Fox, Russo, & Dutton, 2002; Yoon, Hong, Joormann, & Kang, 2009). The speedy and effortless perception of facial emotions suggests that they may be quickly perceived in other objects. Indeed, simple geometric shapes and household objects that include features resembling happy or angry faces generate the same speedy attentional and affective responses as the faces themselves (Aggarwal & McGill, 2007; Ichikawa et al., 2011; Landwehr et al., 2011; Watson et al., 2012; Windhager et al., 2008). Hence, existing evidence is consistent with the view that affective processes are sensitive to emotional resemblance in nonword objects. We thus expected affective/evaluative responses to written words to be influenced by the resemblance between Latin letters and facial emotions.

We took a data-driven approach to identifying emotional resemblance for each lowercase and UPPERCASE Latin letter. As detailed in the next section, we drew emotional resemblance estimates from (a) the judgments of naïve study participants and (b) weights assigned by a neural network previously trained only to identify facial emotion. These two estimates thus do not reflect our ideas—as researchers—of which letters "should" look like facial joy or facial anger, but rather which letters *did* look like facial joy or facial anger to laypersons or a neural network. Both types of

estimates were used to index emotional resemblance in the studies reported here. In Study 1, we used a flanker task (Eriksen & Eriksen, 1974) to examine the extent to which people spontaneously process emotional resemblance in Latin letters. In Study 2, we used the affect misattribution paradigm (AMP; Payne, Cheng, Govorun, & Stewart, 2005) to examine whether emotional resemblance influenced participants unintentional affective responses to letters. In Study 3, we examined if emotional resemblance in letters influenced evaluations of letter strings—an important question given that letters are processed differently when presented alone than when presented in a string (cf. Grainger & Holcomb, 2010). In Study 4, we examined the extent to which emotional resemblance in letters influenced (a) evaluations of real words and (b) response-times to make those evaluations.

Generation of Emotional Resemblance Scores

Prior to conducting Studies 1–4, we sought to establish convergent validity in emotional resemblance estimates for each of 52 Latin letters (26 letters, both cases). Accordingly, we identified two different means of indexing emotional resemblance in letters, one derived from human judgments and one derived from a neural network trained only to identify facial emotion. We used these estimates in Studies 1–4 to predict affective responses to letters, letter strings, and written words. We here provide a summary of our procedures for generating emotional resemblance weights, and a more detailed description appears in the online supplementary materials.

Emotional Resemblance Scores: Human Perceptual Judgments

In pilot studies in North America and Japan, we first sought to establish that (a) humans agreed in their judgments of emotional resemblance in Latin letters, that (b) such agreement was not based on letter serial position or frequency, and that (c) these judgments were based on resemblance to facial *emotions* rather than resemblance to human faces in general. Participants were recruited from both North America and Japan to test the extent to which cross-cultural agreement could be reached among people who share neither a spoken language nor an alphabet.

Mechanical Turk (MT) was used to recruit 193 North American fluent English readers (seven additional participants did not complete this study), each of whom was randomly assigned to rate the extent to which each of 52 Latin letters visually resembled either (a) happy-looking faces or (b) angry-looking faces. These letters were embedded among 104 other "arbitrary" symbols (e.g., MS Word "Wingdings") so as to encourage participants to focus on visual details (they were explicitly asked to focus on visual details and not any other type of meaning they associated with those symbols). In the happy-rating condition, participants were asked to evaluate the extent to which each symbol looked like a happy face (1 = *not at all* to 10 = *extremely*). The angry-condition had identical instructions and rating scales but with respect to how much symbols looked like angry faces.

For each participant, ratings of Latin letters were standardized to reduce elevation in intraclass correlation (ICC) analyses (Shrout & Fleiss, 1979). We here describe ICC average-rater coefficients, corresponding to a conventional Cronbach's alpha (see online supplemental materials for single-rater coefficients and rationale). Reliability for raw ratings was high for both anger (ICC = .97) and joy (ICC = .97), remained high after controlling for letter frequency in English texts and serial position in the Latin alphabet (ICC's $\geq$ .94), and remained high in all of these analyses after excluding letters that, on average, were outliers (typically "x" and "z"; ICC's $\geq$ .92). These coefficients far exceed levels of reliability (e.g., $\alpha \geq .80$) appropriate for generating composite scores (for each letter, an average score across raters), which we do for Studies 1–4.

One standard of comparison for interrater consistency in emotion ratings of Latin letters are studies that examine emotion ratings of human faces. However, few emotion recognition studies both utilize Likert scales of facial affect and report interrater reliability. We were able to locate one study that examined 146 participants' Likert-ratings of faces on several emotions (Matsumoto et al., 2000). Matsumoto et al. (2000) observed αs of .77 and .79 for anger and happiness, respectfully, comparable with the interrater consistency observed in the current study (although Matsumoto et al., 2000 included more judges per rating). Hence, consensus in ratings of facial emotion is similar for Latin letters (in the current study) and for real faces (in Matsumoto et al., 2000). Finally, if participants were simply rating letters' resemblance to the structure of a typical (neutral) human face, rather than letters' resemblance to facial *emotions*, a positive correlation between facial joy and facial anger ratings should be observed. Conversely, and as with ratings of joy and anger in real faces, we expected and observed a negative correlation between ratings of facial joy and facial anger in Latin letters, $r(50) = -.66$, $p < .001$.

These findings were specific to a North American sample, so to what degree do people (more generally) agree in their judgments of emotional resemblance in Latin letters? Perhaps consensus in these judgments is sample-specific, culture-specific, or specific to people who frequently use of the Latin alphabet. We thus examined cross-cultural and cross-linguistic consensus in emotional resemblance judgments by comparing the judgments from the North American sample (see above) to a Japanese sample. Eighty-three undergraduates aged 20–23 were recruited at Aoyama Gakuin University near Tokyo, Japan, and participated in exchange for extra credit. These individuals reported a range of English-reading experience (see below), from no experience to frequent experience, and were randomly assigned to rate each letter's resemblance to facial anger or to facial joy. Levels of consensus (ICC's $\geq$ .87) within this Japanese sample were comparable to those observed in the North American sample and to those observed in Matsumoto et al. (2000) study of facial emotion ratings (see online supplemental materials).

We next turned our attention to cross-cultural consensus: Do the average-ratings of North Americans agree with the average-ratings of Japanese individuals? Correlations between composite American ratings and composite Japanese ratings were uniformly positive, large, and significant. Japanese and Americans exhibited consistency in their ratings of Latin letters' resemblance to facial anger, $r(51) = .75$, $p < .001$, and facial happiness, $r(51) = .53$, $p < .001$, even after controlling for serial position and frequency of occurrence in English texts ($pr(48) \geq .52$). North American judgments were even consistent with the judgments of Japanese participants who indicated that they never or rarely read English ($N = 18$), with cross-cultural correlations ranging from $r(51) =$

.48 to .61. Please see scatterplots in the online supplementary materials for each letter's emotional resemblance rating generated by American and Japanese judges.

Ultimately, emotional resemblance scores from the North American sample were highly correlated with those from the Japanese sample—even among Japanese who never or rarely read English. Such agreement held after controlling for nonvisual characteristics that could reasonably explain it (serial position, letter frequency). North Americans exhibited consensus in their judgments of the extent to which each individual letter resembles facial joy versus facial anger. Such consensus was comparable to that observed for ratings of emotion in faces and remained after controlling for letter serial position and frequency. Our primary index of emotional resemblance in Studies 1–4 is thus the average perceptual judgments of North Americans, accounting for letter serial position and frequency. Please see online supplementary materials for description of ICC single-rater analyses, details of how we controlled for various factors, several statistical treatments of extreme ratings, and detailed rationale for analyses.

Emotional Resemblance Scores: Neural Network Estimates

Many factors contribute to perceptual judgments, including low-level visual features but also including "top-down" influences from expectations, emotion, and other sensory modalities. Human judgments are excellent estimates of what people subjectively see and are therefore the primary predictor used in Studies 1–4. However, these subjective percepts may be informed by conceptual knowledge as well as low-level visual characteristics. For example, subjective percepts of emotional resemblance may be influenced by the pleasant (or unpleasant) acoustic properties of letters ("bee" for the letter b). Or a given letter may occur quite frequently in happy words (cf. Abrams & Greenwald, 2000) and on the basis of such repeated exposure, be rated as visually resembling a happy face.

Learning configurations versus curviness. Facial emotions are perceived by human observers but can also be defined by low-level configural patterns, and we expected perceptions of Latin letters to reflect these configural patterns. We tested this assumption through the use of a neural network trained only to identify facial emotions from low-level configurations. We expected neural network ratings to be associated with perceptual judgments of human raters.

Finally, facial emotions are complex visual stimuli that are not reducible to a single feature. Yet a neural network could conceivably use a single feature to distinguish positive from negative facial emotions and we sought to address this possibility in our model. Specifically, a network could primarily use curviness to evaluate facial emotion, as curviness is one feature known to be more prominent in facial joy than facial anger (Aronoff, Woike, & Human, 1992; Bar & Neta, 2006; Lundqvist, Esteves, & Ohman, 2004). The weights that a neural network assigns to letters would then index a single feature—the curviness of each letter—rather than the resemblance of each letter to a complex facial emotion per se. Instead, we intended for the neural network to identify letters' resemblance to facial emotion on the basis of its overall configuration. Thus, to examine the strict resemblance between the visual configuration of a letter with the visual configuration of facial joy (or facial anger), we trained a neural network to identify facial joy and facial anger but *equated the degree of curviness in facial joy versus facial anger*. Hence, the network did not learn to identify facial emotion on the basis of curviness. This network then assigned joy and anger ratings to each Latin letter, and we examined whether these curviness-controlled ratings correlated with perceptual judgments of emotional resemblance and whether these ratings predicted affective responses to words (Study 4).

Summary of network model. Please see the online supplementary materials for details on the neural network model and rationale for its input, hidden layers, and output. Here we provide a brief overview. The neural network had three groups of units: Input (1,800 units), hidden (196 units), and output (40 units) layers were fully and bidirectionally connected (see Figure 1). Stimulus input to the network were bitmap images of schematic faces (from Lundqvist et al., 2004) and 52 Latin letters. Prior to presenting the bitmap images to the network, the images were convolved with a "difference of Gaussians" filter that approximates the contrast enhancement processes provided by the earliest stages of visual processing (Enroth-Cugell & Robson, 1966; Marr & Hildreth, 1980; Young, 1987; see Figure 1). The network was tasked with simultaneously producing two values associated with each image, each represented by a separate group of units (one positive or "happy," one negative or "angry," see Figure 1). Please see the online supplementary materials for an illustrative example of how the network computes these values and for details (including equations) on the network's use of a distributed, spatial representation of value.

Training and testing procedures were performed on 25 different sets of random initial weights in order to determine the consistency of identified effects. In order to train the network, the eight different schematic images were presented in a pseudorandom order, and the network was asked to generate a distributed pattern of activity corresponding to the rating of each schematic. The positive target value for each face corresponded to that face's normed positive rating, as identified by Lundqvist, Esteves, and Ohman (2004). Because there were not independent positive and negative ratings of each schematic face, the negative value during training was the positive value multiplied by −1.

Training procedure and performance. Each network started training with a different set of random weights, and these weights were changed after each schematic face in order to more closely approximate the target human ratings (see online supplementary materials for algorithm details). A full cycle through all eight schematics is termed a training *epoch*, and training consisted of 100 epochs. Training performance was measured as the correlation between the schematic face ratings produced by humans and the corresponding ratings produced by the model. Overall, the networks were able to recreate the normed ratings very well (mean $r = .94$; distribution of best training performance is summarized in Supplementary Table 2), and the networks learned the schematic ratings relatively quickly (as described in the online supplementary materials).

Testing procedure and performance. After every three epochs (i.e., 24 schematic face stimuli), learning was turned off, and the network was asked to generate positive and negative ratings to all 26 letters in both cases (52 different stimulus inputs). Testing performance across all letters was measured as the correlation between the scalar value produced by the network and the scalar

value produced by human perceivers (see prior section). This correlation was computed for both happy and angry values. The best performance for a given training run was recorded as the *test* with the highest correlation between the happy-face neural network scores (even if that best score occurred very early in training) and the happy-face judgments made by humans. This measure is subject to bias, because the null distribution of the maximum across epochs is not centered at 0. As such, we also computed an additional statistic that was the difference between the maximum correlation across the training epoch and the absolute value of the minimum correlation; if the null distribution of correlations is symmetric and centered at 0, then this difference statistic will be unbiased.[1]

Testing performance demonstrated a few critical phenomena. First, the distribution of best testing performance was clearly different from 0 for happy and angry letter ratings (at training k = 1: happy mean, $r = .32$, $t(24) = 28.30$, $p < .001$; angry mean, $r = .37$, $t(24) = 25.5$, $p < .001$; see Supplementary Figure 5). More importantly, this maximum correlation was larger than the absolute value of the minimum testing correlation across the expected range of thresholds (at training $k = 1$: happy difference = .09; $t(24) = 4.1$, $p < .001$; angry difference = .12; $t(24) = 5.7$, $p < .001$), demonstrating that bias cannot account for these results. Notably, the correlation between human and network ratings of letters was positively associated with the correlation between human and network ratings of faces ($\beta = .13$; $t = 6.6$, $p < .001$): As the neural network improved in its approximation of human ratings of affective faces, it improved in its approximation of human ratings of emotional resemblance in letters (but see Footnote 1). This was true even after controlling for amount of time spent on learning.

Aggregated testing performance. We generated neural-network activity aggregated across runs for use in Studies 1, 2, and 4. Network activity should be roughly equivalent across runs but each run requires noise that generates variance in network activity *across* runs. Differences among testing runs in their activity for letters is not unlike differences among participants in their ratings of letters: aggregating across runs should increase the reliability of network activity estimates just as aggregating across participants should increase the reliability of letter ratings. Consequently, optimal network activity for each letter was recorded on each testing run and subsequently averaged. We examined the correlation between these aggregated network values and perceptual judgments derived from human ratings. Happy network activity and human ratings of happiness were correlated at $r(50) = .48$, $p < .001$. Angry network activity and human ratings of anger were correlated at $r(50) = .50$, $p < .001$. These aggregated network activities were used to predict responses in Studies 1, 2, and 4.

Pretest Summary

Emotional resemblance can be described as the perceived similarity of specific letters to facial emotions. The validity of the emotional resemblance concept—as it applies to Latin letters—was demonstrated in several pretest studies. Participants agreed in their evaluations of facial joy and facial anger in 52 Latin letters (26 uppercase, 26 lowercase). Such agreement was observed within the U.S. and within Japan, and cross-cultural agreement equaled that observed for judgments of facial emotion in real faces (e.g., Matsumoto et al., 2000). These human judgments of emotional resemblance were correlated with the judgments of a neural network that was trained only to identify facial emotion, and was thus uncontaminated by the cultural meaning of individual letters. Agreement in judgments of emotional resemblance (both within and between cultures) could not be accounted for by the known affective components of letters associated with English-language frequency or with serial position in the alphabet. Moreover, these components could not account for agreement between human judges and a neural network, as neural network weights predicted human ratings residualized for letter-frequency and serial position (i.e., residualized ratings were the criterion).

Emotional Resemblance Scores for Studies 1–4

Pretests enabled us to obtain emotional resemblance scores for each of 52 Latin letters (26 letters, uppercase and lowercase), and these scores will be used in Studies 1–4. To limit extraneous influences on letter perception, we control for letter frequency and serial position in all analyses. Specifically, Studies 1–2 regarded responses to individual letters and in these studies, emotional resemblance scores of individual letters were statistically controlled for frequency and serial position. Studies 3–4 regarded responses to letter strings (including words) and in these studies, letter strings (including words) varying in emotional resemblance were matched for letter frequency and serial position. Accordingly,

[1] It may seem intuitive that more experience (training) with faces would lead to asymptotically more similarity between letter ratings and human ratings. However, there are two reasons why this argument does not hold for the current model. The first issue is that this argument would only hold if similarity to facial emotion were the only factor that impacts letter ratings, or if we could somehow perfectly statistically control for all other (potentially unknown) variables contributing to letter ratings. Otherwise, the asymptote for training on schematic faces would actually make the correlation between letters and their ratings lower, as the relative impact of the schematic face features would be exaggerated. The second issue is related to the first in that it is the general description of the computational phenomenon that would cause the issue of exaggerating the facial features: overtraining. In general, *overtraining* is the phenomenon where a model learns weights that emphasize and capitalize on very specific features that are tied to the training set that do not generalize to stimuli outside of that set. In addition to the problem mentioned in the preceding paragraph, overtraining also impacts the current situation by emphasizing the very specific spatial relationship and alignment between the schematic faces and letters (e.g., Amari, Murata, Müller, Finke, & Yang, 1996; Fukumizu, 1998). In this case, the training set is specifically the small ($N = 8$) set of schematized faces, and as the network becomes overtrained, the network would become less likely to respond well to stimuli that are not those specific faces. This means that the particular spatial relationship of the faces and letters (e.g. the "smiley face" and "U") becomes more important over training; in other words, the precise overlap of exact pixels would get more and more important over training. We are attempting to ignore these spatial relationships and focus on spatially invariant relationships (particularly through spatial sweeping attempts). As overtraining becomes a problem, the resolution of those sweeps would need to become higher in order to have a chance at capturing the right information. Because of these issues, we attempted to find a balance between training on the faces, but not overtraining. We couldn't set an arbitrary number of epochs to train because each network starts with random weights, and there is not a clear and consistent relationship between how much training to do and generalization performance. Some runs will naturally take long because the random initial position of the weights is farther away from the optimum. Finally, to clarify, overlearning is analogous to overfitting but is perhaps more like a specific case of overfitting.

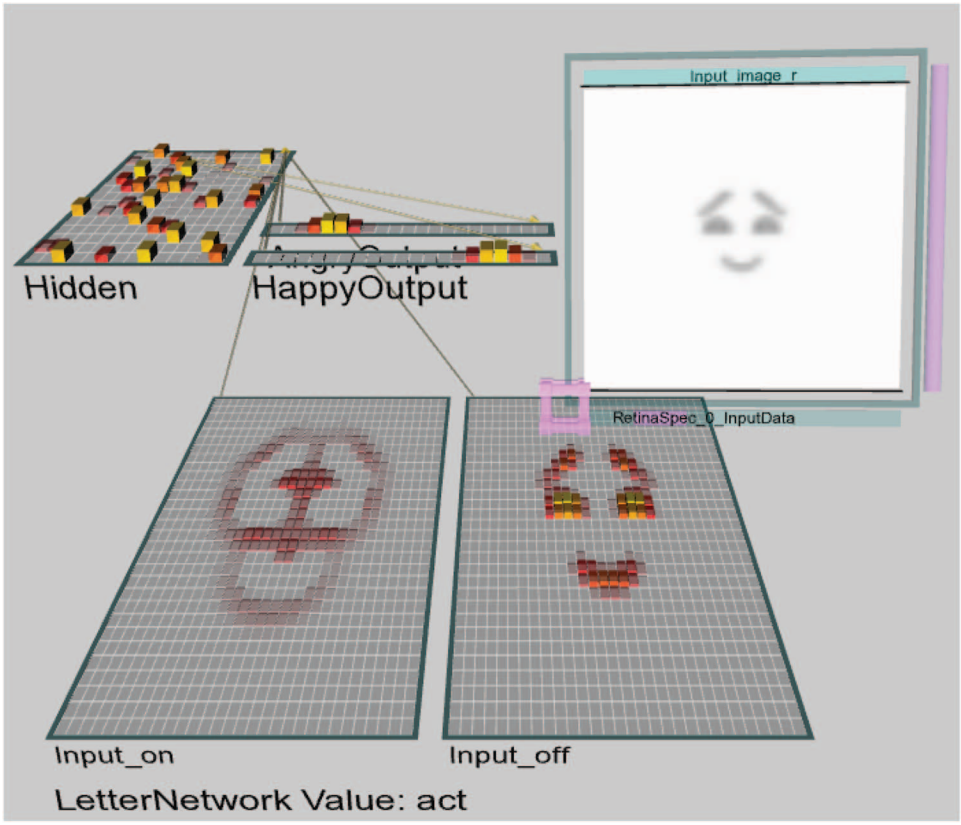


Figure 1. Network architecture: The network consisted of two input layers (each of which corresponds to a different filtered version of the bitmap image, see text), a hidden layer, and two output layers. The input layers (Input_on and Input_off) encoded contrast-enhanced versions of the actual bitmap image (presented in the top right). The Input_on layer represented a version of the image filtered by an on-center/off-surround filter, whereas the Input_off layer represented a version of the image filtered by an off-center/on-surround filter. The original image consisted of black lines on a white background (see top right), and so is more similar to the Input_off image. The output layers (HappyOutput and AngryOutput) used a spatial representation to code for value, such that a Gaussian bump farther to the right coded a larger value, and a Gaussian bump farther to the left coded a smaller value. The presented image is a "happy" face, and is associated with a large happy value and a small "angry" value (see output layers). See the online article for the color version of this figure.

results from Studies 1–4 aim to isolate the effect of perceived emotional resemblance on humans' affective responses to letters.

Because we have defined emotional resemblance as the *perceived* similarity of specific letters to facial emotions, we focus our discussion of results for Studies 1–4 on scores derived from human perceptual judgments. To avoid redundancy (Studies 1–4 yielded similar results for human perceptual judgments and neural network weights), we report results using neural network patterns in the online supplementary materials, and only summarize them in the main text for Studies 1, 2, and 4 (Study 3 did not use neural network estimates). In the studies that follow, we use these scores to examine how emotional resemblance shapes affective responses to written language.

Study 1

In Study 1, we sought to examine the extent to which people incidentally perceive emotional resemblance in Latin letters. Readers do not typically try to see facial emotion in Latin letters, so it was important to test whether people process letters' emotional resemblance in the absence of an intention to do so.

To examine such incidental perception, we used a flanker task (Eriksen & Eriksen, 1974). Participants were instructed to categorize schematic happy and angry faces as "happy" or "angry" on each of a large number of trials. Each trial consisted of a centered schematic face, flanked on the left and right by a Latin letter. To the extent that letters are *incidentally* processed with respect to facial emotion, letter flankers should modulate the speed with which participants identify emotion in the centered face. Angry faces should be identified more quickly when flanked by letters that resemble angry faces than when flanked by letters that resemble happy faces. Happy faces should be identified more quickly when flanked by letters that resemble happy faces than by letters that resemble angry faces.

The predicted effects could accrue for at least two reasons. First, target stimuli may be identified more quickly when they do (vs. do not) share visual features with flanking stimuli because such visual

redundancy sometimes makes target stimuli easier to see (e.g., Pomerantz, Pristach, & Carson, 1988; Treisman & Gelade, 1980). Indeed, the emotional resemblance score of any flanker letter regards its perceived similarity to facial emotion and may thus be visually redundant with target facial emotions. Second, target stimuli may be identified more quickly when they do (vs. do not) share visual features with flanking stimuli because both target and flanker activate the same response (either positive or negative affect). Either way, observation of the predicted effects would indicate that participants *incidentally* perceived similarity between facial emotions (the targets) and letters (the flankers). Accordingly, we expected a congruency effect; congruent flankers should facilitate processing relative to incongruent flankers. Data analyses thus examined response times to angry faces and happy faces as a function of the emotional resemblance scores of flanking letters.

Method

Design. There were two within-subjects factors and two between-subjects factors. The within-subjects factors were (a) angry versus happy target face and (b) flanker letters (ranging from angry to happy). The between-subjects factors were whether letters were (a) uppercase or lowercase and (b) the location of the happy and angry response keys.

Participants. The Denver Craigslist website and a University of Denver website were used to recruit participants. The use of both websites enabled us to recruit a sample with a diverse educational background and with a wide range of ages. We estimated effect-size on the basis of prior emotional flanker studies and on this basis, sought 112 participants to achieve 75% power in the current study.[2] The initial sample included 110 participants but one participant stopped shortly after starting the study, two participants reported confusion and used incorrect response keys, and data for two participants were not saved. The final sample of 105 participants (45 male) had an average age of 31 years with quartiles of 20, 26, and 42 years of age. Of these participants, 68% were non-Hispanic Caucasian, 11% were Hispanic, 9% were Asian, 5% were African American, 3% were Native American, and 4% were mixed-race. About 13% of the sample never attended college, 47% had completed some college but had not (yet) graduated, 9% had an associate's degree and were no longer in school, 22% had completed a bachelor's degree, and 10% had gone beyond a bachelor's to receive a master's or PhD. The diversity of this sample is likely to reflect a broad range of English reading and writing abilities.

Materials. We created 104 black-and-white images for this study. Each image included either a happy (friendly) or angry (threatening) schematic face presented at center screen and flanked by two identical letters. Stimuli were black and the remainder of the computer screen had a white background. The two schematic faces were chosen from among a large group of schematic faces created by Lundqvist and Ohman (2005). This larger group included schematic faces that were quite similar except for a few slightly different features (e.g., different brow angle). We selected a happy and an angry face rated especially positively and negatively, respectively.

Each flanker screen was created in Microsoft PowerPoint© as a single slide. Each slide included one of the two aforementioned faces, centered and resized to be 18 mm in height. This face was flanked to the left and to the right by an English letter, 10 mm in height. The letter was placed 5 mm from the edge of the face (on either side) and was vertically centered relative to the face (for an example, see Figure 2). This procedure was repeated for each of 52 (uppercase and lowercase) Latin letters and two schematic faces, resulting in 104 flanker screens.

Each participant saw only 52 of these screens, however, as letter-case was a between-subjects variable. We included letter-case as a between-subjects factor because of a systematic difference in letter-height between UPPER and lower case but also because the form of some letters differs by case whereas case does not influence the form of other letters (e.g., "a" and "A" vs. "c" and "C").

Procedure. Participants completed the study in individual cubicles and were informed that they would take part in a study with a large number of trials. Participants were instructed that, for each trial, their task was to indicate whether a centered face looked happy or angry. They were instructed to do so as quickly and accurately as possible, while ignoring the symbols on either side of the face. A happy face sticker was placed on the "h" key for half of the participants and on the spacebar for the other half of participants. An angry face sticker occupied the other key in both conditions. We selected upper and lower keyboard positions rather than right-left keyboard positions in an effort to limit any motor contingencies that might direct attention to one or the other side of the central face. Participants were instructed to place the index finger and thumb of their dominant hand on the upper and lower keyboard stickers, respectively, and were instructed to use these to respond. After confirming that they understood the instructions, participants began the flanker task. Stimuli in this study (and Studies 2–4) were presented on a 17-in. CRT monitor with an 85 Hz refresh rate, about 40 cm in front of each participant.

There were three blocks of 52 trials each. Each block included every possible combination of letter and happy/angry face. After each of the first two blocks, participants were required to take a 1 min break. In total there were 156 trials. After the conclusion of these trials, participants were debriefed, thanked, and reimbursed.

Results

Analytic strategy. Data from two participants were excluded because their average response times were at least four standard

[2] We located five publications with a total of seven studies that included facial emotion as both target and flanker (Barratt & Bundesen, 2012; Fenske & Eastwood, 2003, Study 1a and Study 2; Horstmann et al., 2006, Study 1; Schmidt & Schmidt, 2013; Watson et al., 2012). We also located two publications with four studies that included stimuli resembling facial emotion as targets or flankers (Horstmann et al., 2006, Studies 2–3; Watson et al., 2012). All designs were repeated measures and we therefore estimated the mean and standard deviation for each experimental group (cf. Dunlap, Cortina, Vaslow, & Burke, 1996) from figures and text in order to calculate Cohen's *d* for congruent-incongruent comparisons. Some studies report different values for positive and negative targets—in those cases, we calculated effect sizes for positive and negative targets separately and took an average as our design focused on congruent-incongruent comparisons. Effect-sizes ranged from $d = .03$ to $d = .41$ with an average $d = .18$ (median = .19). Studies with only face stimuli and studies with face-resembling stimuli had similar effect sizes (mean *d*s = .18 and .17, respectively). These effect-sizes correspond to what Cohen (1992) regards as small. To achieve 75% power in a repeated-measures design with a small effect and two within-subjects factors, we sought a sample size of 112 (computed with the software program G-power; Faul, Erdfelder, Lang, & Buchner, 2007).

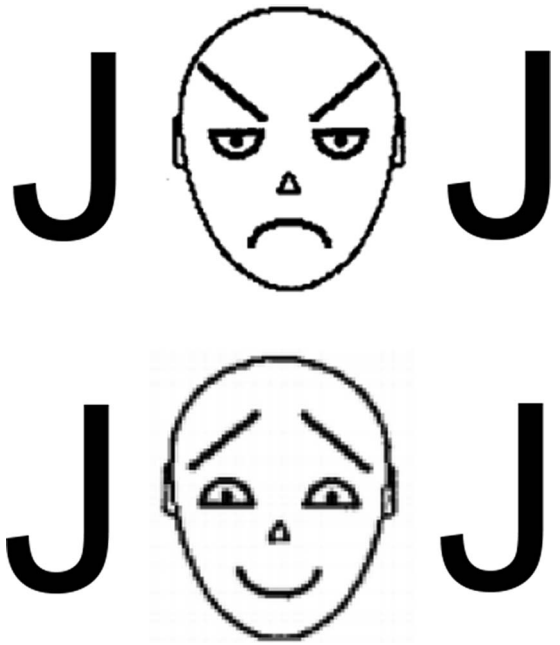


Figure 2. Examples of flanker screens in Study 1. Please see text for size and spacing parameters.

deviations (*SD*) above the sample mean (no other average response times were three *SD* above the sample mean). For the remaining participants, incorrect responses (4%) were excluded as were individual response-times that were more than 3 *SD*'s above a participant's personal mean response-time (3%).

Data analyses were conducted using generalized estimating equations (GEE; Zeger & Liang, 1986). In the current context, GEE not only accounts for a participant's average response-time for a particular face/letter pairing (e.g., happy face + letter "m") but also for the variance *among* that participant's response-times per pairing. In this way, GEE accounts for within-participant variance unrelated to treatments, within-participant variance due to treatments, and between-participants variance due to treatments. GEE generates regression weights and in many respects is similar to traditional multiple regression but accounts for within-subject dependencies.

As in hierarchical linear regression, effects are interpreted at the step at which they are entered in the equation—hence, main effects are interpreted before adding two-way interaction effects which are, in turn, interpreted prior to adding three-way interaction effects and so on. For this analysis, mean response time was predicted by two categorical between-subjects factors: letter-case (lower vs. upper) and keyboard (happy on "h" vs. spacebar), one categorical within-subjects factor: target face emotion (angry vs. happy), and one continuous within-subject factor: flanker letter emotional resemblance.

Letter emotional resemblance was calculated as difference scores: residualized happy-face ratings minus residualized angry face ratings. These difference scores thus reflect the degree to which each letter looked like a happy face more than an angry face, independent of the degree to which each letter occurred early in the alphabet and occurred frequently in English-language texts in the Pretest Summary section (for more details, see online supplementary materials). An alternative index of emotional resemblance is reported in the online supplementary materials, and is based on neural network weights for each letter (see online supplementary materials for Study 1 analyses in which neural network weights were used to index emotional resemblance). Finally, alternative analyses including age and/or education as factors did not meaningfully change the results reported below so these analyses are not reported further.

Flanker effects of emotional letters. We focus this section on the predicted two-way interaction between central face emotion and flanker emotional resemblance but we first report main effects. We observed only one main effect, in which participants responded faster to angry-face targets ($M = 584$) than happy-face targets ($M = 594$), $B = 9.51$, $SE = 4.13$, Wald $\chi^2 = 5.29$, $p = .02$. Our primary predictions regarded the two-way interaction between central face emotion and flanker emotional resemblance. This interaction was indeed significant, $B = -13.17$, $SE = 3.64$, Wald $\chi^2 = 13.09$, $p < .001$ (see Figure 3). Participants were faster to respond to angry-faces to the extent that flanker letters resembled angry (vs. happy) faces, $B = 7.08$, $SE = 2.64$, Wald $\chi^2 = 7.21$,

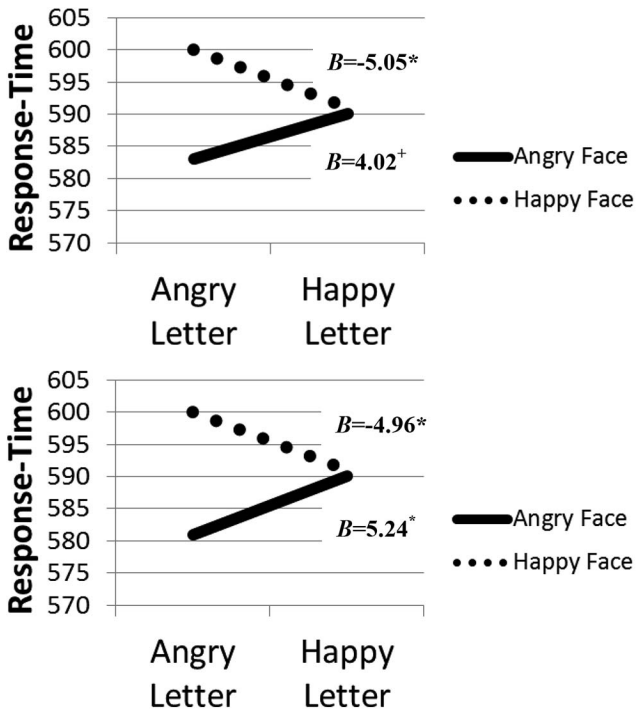


Figure 3. Study 1 response-time to identify facial anger is indexed via solid line and response-time to identify facial happiness is indexed via dotted line. In the top panel, human-rated emotional resemblance is plotted at one standard deviation below (angry letter) and above (happy letter) the emotional resemblance mean. In the bottom panel, neural-network emotional resemblance is plotted at one standard deviation below (angry letter) and above (happy letter) the emotional resemblance mean.

$p = .007$.[3] Conversely, participants were faster to respond to happy faces to the extent that flanker letters resembled happy (vs. angry) faces, $B = -5.90$, $SE = 2.27$, Wald $\chi^2 = 6.73$, $p = .009$. These effects remain significant when the keyboard factor is removed from analyses (e.g., Flanker Emotional Resemblance × Target Face interaction $p < .001$). No other significant interactions emerged, so these results were not qualified by other factors.[4] Finally, analyses with neural network ratings produced identical patterns of significance and means (both simple effects reached significance with the neural network estimates).

Discussion

The results of Study 1 suggest that emotional resemblance in Latin letters is incidentally processed. Specifically, participants were unable to ignore the emotional resemblance of flanking letters when attempting to identify the emotion on a centrally located face. To the extent that Latin letter flankers resembled facial anger and not facial happiness, response-times to angry target faces were reduced and response-times to happy target faces increased. This pattern held when emotional resemblance in letters was indexed by human ratings (main text) as well as when emotional resemblance was indexed by neural network activities (online supplemental materials).

Participants incidentally processed the visual resemblance between letters and emotional faces, but there are at least two routes through which such visual resemblance could have led to congruence effects in this Flanker paradigm. First, participants may have incidentally processed visual similarity between flanker letters and the target face (e.g., Pomerantz et al., 1988; Treisman & Gelade, 1980), thus making the target easier (if similar) or harder (if dissimilar) to see. Hence, the observed congruency effects may not owe to affective processing per se, but rather visual similarity. However, visual resemblance between targets and flankers often yields *incongruence* effects (Bjork & Murray, 1977; Egeth & Santee, 1981) and such incongruence is especially likely for letter stimuli and especially in tasks requiring judgments based on internal features (vs. whole shape) of the stimulus (e.g., van Leeuwen & Lachmann, 2004). A second explanation is that participants incidentally processed the affective meaning of the flanker letters, which either facilitated or interfered with the affective judgment of the target. In Study 2 we more specifically test the idea that people unintentionally process the affective meaning of letters, according to letters' emotional resemblance.

Study 2

To more directly examine the extent to which participants unintentionally activate affective responses to letters' emotional resemblance, we used an affect misattribution task AMP (Payne et al., 2005). This changes the judgment task from one in which participants try to identify the emotion of a face (Study 1) to one in which participants try to evaluate their affective responses to a neutral, nonface stimulus (Study 2).

The AMP is an empirically validated instrument which allows for measurement of unintentional affective responses to stimuli. On each trial of the AMP, a letter "prime" is presented at a brief (75 ms) stimulus duration just prior to a neutral target stimulus. Immediately following stimulus presentation, participants evaluate the target stimulus (as "pleasant" or "unpleasant") as quickly as possible. We predicted letters would elicit more positive affect to the extent that they resembled happy facial emotion more than angry facial emotion.

Method

Participants. Two-hundred and 21 undergraduates (151 female) at the University of Denver (DU) participated in exchange for credit in a psychology course. We assumed a small effect size, consistent with Study 1, but sought higher power than Study 1 given that this study was a conceptual replication. We thus sought 210 participants to achieve 90% power in the current study but oversampled in anticipation of excluding native Chinese readers (for whom Chinese characters are meaningful). The final sample included 211 student participants, after excluding 10 native Chinese readers.[5]

Materials. Target stimuli were 200 Chinese ideographs drawn from those used in Payne, Cheng, Govorun, and Stewart (2005). Prime letters were screen-centered in 96-point Arial font.

Procedure. Participants completed the study in individual cubicles and were informed that they would take part in a study with a large number of trials. Participants were instructed that, for each trial, their task was to identify the pleasantness of Chinese character drawing. They were informed that on each trial, they would only see the drawing briefly and that it would be preceded by a letter. They were instructed to ignore the letter and evaluate the drawing only ("e" keypress = unpleasant; "i" keypress = pleasant). Specifically, and following Payne et al. (2005), they were told:

> It is important to note that the letter can sometimes bias people's judgments of the drawings (Chinese characters). Because we are interested in how people can avoid being biased, please try your absolute best to not let the image of the letter bias your judgment of the drawings! Give us an honest assessment of the drawings (Chinese characters), regardless of the letters that precede them.

Each trial began with a letter prime presented for 75 ms, followed by a blank screen for 125 ms. The target drawing then appeared for 100 ms, and was followed by a backward mask until the participant responded. Participants completed 208 trials in total, including eight trials for each of 26 letter primes. Target images were randomly paired with letter primes. After the conclusion of these trials, participants were debriefed, thanked, and given credit.

[3] This effect becomes marginally-significant when letter frequency and serial position are not accounted for, $B = 4.02$, $SE = 2.43$, Wald $\chi^2 = 2.73$, $p = .10$.

[4] However, an interaction between letter case and keyboard positioning approached significance, $B = 89.40$, $SE = 50.19$, Wald $\chi^2 = 3.17$, $p = .08$. For the sake of brevity and because this effect does not involve letter valence or emotional resemblance, we do not interpret this effect here. No other interactive effects approached significance (all other ps $> .15$).

[5] All analyses reported in this study remained statistically significant when including these participants ($p < .05$). DU recruits an unusually large (for the school's size) student population from China, and introductory psychology courses routinely include 2–10% native Chinese students.

Results

As in Study 1, data analyses were conducted using GEE. Analyses for Study 2 used a logistic model given the binary dependent variable. This GEE thus generated regression weights and in many respects is similar to traditional logistic regression but accounts for within-subject dependencies. The only predictor was letter emotional resemblance weights, calculated identically to Study 1. As predicted, participants evaluated targets more positively when they were primed by letters with higher emotional resemblance scores, $B = 0.03$, $SE = 0.01$, Wald $\chi^2 = 4.41$, $p = .036$ (see online supplemental materials for neural network analyses).

Discussion

The results of Study 2 suggest that perceivers unintentionally generate affective responses to Latin letters on the basis of those letters' emotional resemblance. Specifically, participants were unable to ignore the emotional resemblance of prime letters when attempting to evaluate a target stimulus. To the extent that Latin letter flankers resembled facial happiness and not facial anger, target stimuli were more likely to be evaluated as pleasant than as unpleasant. Hence, participants *did* seem to unintentionally generate affective responses to emotional resemblance in Latin letters.

Study 3

The preceding pretests and studies provide evidence consistent with the idea that: that people perceive facial emotion in Latin letters, that such perception occurs in cultures that utilize Latin letters in writing as well as in cultures that do not, and that such emotional resemblance is perceived incidentally and elicits unintentional affective responses from perceivers. Yet it is possible that emotional resemblance effects are inconsequential in the context of letter strings. We conducted Studies 3–4 in an effort to examine effects of emotional resemblance in letter strings without meaning or obvious pronunciation (Study 3) and in meaningful letter strings (i.e., words) with obvious pronunciation (Study 4). Study 3 enables us to design the experiment in such a way to eliminate a variety of factors as explanations for the effects of emotional resemblance (see next paragraph). However, Study 3 suffers with respect to generalizability to reading real words (hence, Study 4 follows).

Several factors might interfere with letter-based emotional resemblance effects when those letters are combined into strings. Research on reading demonstrates that letters are processed quite differently when those letters are presented alone as compared with when those letters are presented in the context of other letters—even when the resulting string is not a word. This is plainly illustrated by the well-known illusion in which the two outer lines of "A" fail to connect at the top and are placed between other letters T- -E or C- -T (with perceptions of the middle letter as "H" or "A", respectively). Although this sort of effect is especially strong in the context of a word, surrounding letters have an influence on letter identification even within nonwords. Specifically, the speed and accuracy with which a given letter is identified depends on the frequency of its co-occurrence with the surrounding letters (e.g., Rumelhart & McClelland, 1982). This is one example of how letter processing differs in the context of other letters but there are other examples (Gomez, Ratcliff, & Perea, 2008; Healy, 1994; Richman & Simon, 1989), many of which suggest that processing of a target letter often has less to do with the visual features of that letter than the surrounding context letters. Thus, we here examined the effects of emotional resemblance when letters appear in the context of letter *strings*.

Overview

In Study 3 we examined whether the presence of individual "angry" or "happy" letters in letter strings would influence perceivers' affective evaluations of those strings. To do so, we inserted letters with angry or happy emotional resemblance scores into nonword letter strings and asked participants to quickly indicate their affective response (good/bad) to each letter string. We elected to explore the influence of emotional resemblance in nonwords before evaluating those effects in real words, in order to protect against several possible alternative explanations possible with real words. By using nonwords, we were able to preclude the possibility that words' conceptual meaning account for emotional resemblance effects or that orthographic neighborhood accounts for such effects. We also strongly limited the influence of acoustic properties of word pronunciation. Thus in Study 3, all letter strings were conceptually meaningless, had no one-letter orthographic neighbors, and were nonpronounceable.

Each trial in Study 3 required participants to briefly view a string of letters and to indicate whether they thought this string seemed "good" or "bad." We created pairs of letter strings which were identical with regard to the "neutral" letters they contained, but varied with respect to the valence of the target letters (i.e., "angry" or "happy"). For example, one pair of letter strings was "JGQL" and "KFQL"—the target letters are underlined here for presentational purposes–they were not underlined in the experiment itself. In this example, two of four letters were *critical* but the number of target letters included in a string was also manipulated (i.e., one, two, or four). We predicted that letter strings would be more likely to be called "good" when they included happy (vs. angry) critical letters.

Method

Participants. One-hundred and one undergraduates (19 male) at the University of Denver participated in exchange for payment or extra credit in a psychology course. There were 73 Caucasian, 9 Asian, five Hispanic, two Native American, three African American, one Middle-Eastern, one Hawaiian, and six multiracial participants.[6]

Materials. We created seven matched pairs of target letters, such that each pair included one letter that was rated as resembling facial happiness more (and facial anger less) than the other letter. We matched uppercase letters on emotional resemblance extremity (absolute difference from 0), on frequency, and for whether the letter was a vowel or consonant (all paired letters were consonants). We restricted letter pairs to consonants because vowel pairs did not meet other matching criteria. Finally, the low frequencies

[6] Based on the results of Study 2, we expected a small effect size in Study 3. To achieve 75% power for an effect of letter emotion with 2 (Letter Emotion) × 3 (Number Of "Emotional" Letters) repeated measures, we sought 98 participants (sample size estimate calculated with G*Power).

of the letters "x" and "z" could not be matched with any "happy" letters so these two letters were excluded from pairs. These criteria resulted in the following pairs (happy letter first): J-K, S-T, G-F, P-N, D-H, C-M, B-W, constituting over half of the Latin alphabet. These paired letters are listed with relevant statistics (e.g., frequency) in Table 1.

These target letter pairs were then used to generate pairs of letter *strings*. Because each target letter pair consisted of two letters roughly equated on extremity, frequency, and letter type (i.e., consonant), each pair of letter strings consisted of two strings roughly equated on these characteristics. A pair of letter strings was created by first generating a base of "neutral" letters—letters with emotional resemblance scores close to 0. For example, one base was _ _ Q L. The base was identical for both members of a letter-string pair. Target letters were then added to the base. For example, J and G both had highly positive emotional resemblance scores and were used to create the string JGQL. This string was paired with KFQL—K and F were matched with J and G, respectively. Hence, the "happy" letter string JGQL was paired with the "angry" letter string KFQL. In this example, the two target letters are the first two letters in the string but target letter location varied between pairs (and always identical within a pair). We created 24 pairs of letter strings.

An important consideration in generating these nonwords was to ensure that verbal pronunciation of nonwords did not generate affect independent of the visual features of the letter strings. Consequently, all letter strings were unpronounceable (no vowels) and both exposure time and response time were limited in order to discourage participants from sounding-out the string. An additional requirement for each pair was that the "happy" letter string have an equal number of orthographic neighbors as the "angry" letter string. Consequently, all letter strings had zero orthographic neighbors (checked with the orthographic word database, *MCWord*; Medler & Binder, 2005)—that is, none of the letter strings could become a word by replacing a single letter. A third constraint concerned letter position in the string. Position of a letter within a string may have an influence on how much processing that letter receives (cf. Grainger & Holcomb, 2010; Rayner, 2009). Consequently, for each happy/angry pair of letter strings, members of a letter-pair simply replaced one another in the string and thereby maintained the same position. Finally, because double-letters may have a different impact than single letters, we only allowed a letter to appear once in a string.

These constraints yielded a limited number of possible letter strings, and this was especially true for letter strings with fewer neutral letters. That is, neutral letters were identical in both members of a pair and hence did not require pairing. For that reason, there was more flexibility in generating letter strings that had more neutral letters. In total, 24 pairs of four-letter strings were created. Of these 24 pairs, 14 included a single emotional letter, eight included two emotional letters, and two included all (four) emotional letters. We expected people to evaluate nonwords more positively to the extent that they included letters that resembled facial joy rather than facial anger.

Table 1
Study 3 "Happy" and "Angry" Letters

Letter emotion	Angry *Z*	Happy *Z*	Diff	Frequency	Serial
Happy Letters					
J[1]	−.32	.51	.83	78,706	10
S[2]	−.14	.64	.88	304,971	19
G[3]	.00	.39	.39	93,212	7
P[4]	−.15	.21	.36	144,239	16
D[5]	−.23	.13	.36	129,632	4
C[6]	−.47	.52	.99	229,363	3
B[7]	−.35	.20	.55	169,474	2
Average	−.24	.37	.61	164,228	8.7
Angry Letters					
K[1]	.26	−.21	−.47	46,580	11
T[2]	.26	−.27	−.53	325,462	20
F[3]	.36	−.60	−.96	100,751	6
N[4]	.35	−.26	−.61	205,409	14
H[5]	.12	−.26	−.38	123,632	8
M[6]	.62	−.35	−.97	259,474	13
W[7]	1.00	.08	−.92	107,195	23
Average	.42	−.27	−.69	166,929	13.6

Note. Matched letter pairs are indicated by superscripts. Angry and happy columns are z-scored average ratings of facial anger and facial happiness, respectively, from North American perceptual judgments. Frequency is of uppercase letters in English texts (Jones & Mewhort, 2004) and serial position is of alphabetic order.

Procedure. Prior to their arrival, participants were randomly assigned to one of two counterbalancing conditions. Each participant only viewed one member of each letter-string pair—thus, each participant completed a total of 24 trials (counterbalanced across participants). For any one participant, half of the trials included a letter string with "angry" target letters and half of these trials included a nonword with "happy" target letters. Thus, each participant viewed 12 letter strings with "angry" critical letters and 12 letter strings with "happy" critical letters.

Upon arrival participants were told that they would be participating in a study on psycholinguistics. They were told that they would see strings of letters that were *not* words. For each letter-string, they were asked to give their "gut response" to whether the string seemed "good" or "bad." Participants were also told that they would only be given one second in order to encourage intuitive responses. They were instructed to focus their attention, on each trial, on a series of asterisks that would appear at the beginning of each trial. After reading these instructions participants began the study. Each trial began with a series of asterisks presented center screen for 500 ms. These asterisks were immediately replaced by the letter-string, which remained on the screen until the participant's response. Participants pressed a green button if the string seemed "good" and with a red button if the string seemed "bad." If response time was longer than 1 s, participants were informed that the response time was too long and were encouraged to respond faster immediately after the trial. After completing this study, participants completed several unrelated studies and then were debriefed, paid, and thanked.

Results

Binary logistic GEE was used for data analytic purposes, though in this study only categorical predictors were included. Predictors included *condition* (i.e., counterbalancing), *letter emotion* (happy, angry), and two dummy-coded *letter number* variables to index whether one, two, or four emotional letters appeared in a letter string (two letters served as the reference group). The outcome variable was whether participants responded good (coded as "1")

or bad (coded as "0") and because the outcome variable was binary, GEE analyses were based on a logistic distribution.

As predicted, the main effect of letter emotion was significant, $B = .18$, $SE = .09$, Wald $\chi^2 = 4.05$, $p = .04$, such that letter strings were regarded as "good" more often when they contained happy letters ($M = .58$) than when they contained angry letters ($M = .53$). An unanticipated main effect (of a dummy-coded variable) emerged such that letter strings with four emotional letters were more likely to be regarded as "good" than letter strings with two emotional letters, $B = .33$, $SE = .15$, Wald $\chi^2 = 4.82$, $p = .03$. However, both main effects were qualified by an interaction between target letter emotion and whether one or two letters were emotional, $B = -.33$, $SE = .17$, Wald $\chi^2 = 3.84$, $p = .05$. No such interaction emerged for whether two or four letters were emotional, $B = -.01$, $SE = .07$, Wald $\chi^2 = .02$, $p = .88$ (see Figure 4).

We conducted follow-up analysis within each number of emotional letters. For letter strings with only one emotional letter, there was no difference in "good" responses to happy letter-strings compared with angry letter strings, $p = .79$. Conversely, for strings with two emotional letters, "good" responses were more likely to letter strings that contained happy letters than to letter strings that contained angry letters, $B = .36$, $SE = .12$, Wald $\chi^2 = 8.95$, $p = .003$. The same pattern emerged with letter strings in which all four letters were emotional, such that participants were nonsignificantly more likely to respond "good" to strings containing happy letters than to letter strings containing angry letters, $B = .35$, $SE = .28$, Wald $\chi^2 = 1.53$, $p = .22$.

One explanation of the lack of significance for nonwords with four emotional letters is that there were fewer trials with four (vs. two) emotional letters and standard error (*SE*) increases with fewer trials. Indeed, *SE* was lower for nonwords with two emotional letters (eight strings per participant) than for nonwords with four emotional letters (two strings per participant). Equally as important, the approximate effect size for emotional resemblance was equivalent for strings with two emotional letters as for strings with four emotional letters. As illustrated in Figure 4, strings with two positive letters were more likely to be evaluated as positive ($M = .59$) than were strings with two negative letters (M = .50). Also as in Figure 4, strings with four positive letters were more likely to be evaluated as positive ($M = .66$) than were strings with four negative letters ($M = .58$). Statistical power was thus comparatively low for pairs with four (vs. one or two) emotional letters. Accordingly, the size of the emotional resemblance effect was roughly equivalent in the two-letter and four-letter conditions. No other significant effects emerged in these analyses.[7]

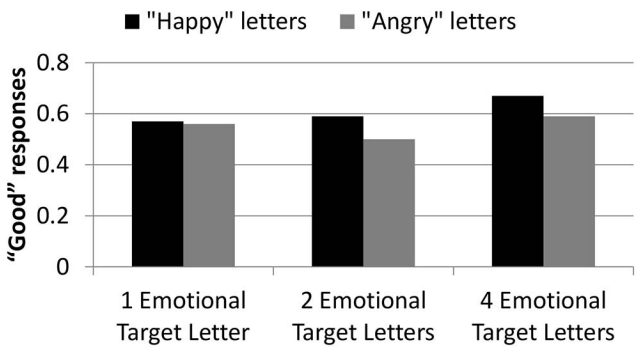


Figure 4. Study 3 proportion of "good" responses as a function of number of emotional letters in the string, and whether those emotional letters were "angry" or "happy."

Discussion

Much as people evaluate faces more positively when they include happy versus angry features, people evaluated nonword letter strings more positively when they contained letters that resembled happy faces than when they contained letters that resembled angry faces. Notably, this effect was more likely when at least half of the letters were emotional. This interaction may reflect a few different processes. For example, with more emotional letters in a string, there is a greater chance of attending to an emotional letter and perhaps that one letter drives evaluations of the letter string. Alternatively, perceivers may attend to all letters and compute a sum or mean of facial happiness and this value should be higher with more happy letters.

Finally, emotional resemblance effects were not statistically *significant* for strings with four emotional letters (or strings with one emotional letter). One possible interpretation of these results is that there is something anomalous about the specific strings we used with two emotional letters. This interpretation cannot be completely ruled out but there are two reasons it is likely to be incorrect. First, the emotional resemblance effect size is nearly identical for the two-letter condition and the four-letter condition (9% and 8%, respectively, more positive responses for strings with "happy" vs. "angry" letters). A straightforward explanation for the lack of statistical significance is the larger *SE* associated with the much smaller sample of strings with four (vs. two) emotional letters. More generally, the same emotional letters were used in the one-letter, two-letter, and four-letter conditions such that the same letter pairs were used to compare "happy" versus "angry" strings (see Table 3). We thus believe that a more likely explanation owes to the computation process used by perceivers. This computation may not be one of simple averaging: Strings with four emotional letters were no more influential than those with two emotional letters. Consequently, a more complex computational procedure may be involved than simple averaging. For example, participants may each have had an affective threshold for regarding a nonword as "good" and for many participants two "happy" letters may have been sufficient to reach that threshold. The important point for our purposes, however, is that evaluations of letter strings could be predicted from whether component letters resembled facial happiness or facial anger.

In Study 3, we used nonwords in an effort to limit confounds that are difficult to meaningfully limit with real words. First, and most importantly, real words have semantic meanings that can impact perceivers' affective evaluations or interact with emotional resemblance to shape those evaluations. Conversely, nonwords do not carry semantic meaning. Second, real words typically have a

[7] We did not have predictions for response time, given the inherent ambiguity in the letter strings. Indeed, response times were not influenced by any of the manipulations ($p > .12$), whether or not we only analyze responses occurring within the response window (1,000 ms or less). We also conducted post-hoc analyses, in which we added the participants' affective judgment (good or bad) as a predictor. This factor did not interact with emotional resemblance to predict response times ($p = .38$).

large number of orthographic neighbors and the meaning of those orthographic neighbors could influence affective evaluations. Conversely, the nonwords used in Study 3 had no orthographic neighbors. Third, real words can be pronounced and their acoustic properties may influence affective evaluations. Conversely, the letters strings in Study 3 were nonpronounceable and "sounding out" was discouraged by limiting exposure and response times. Several other controls were included in Study 3 that can also be applied to real words: We eliminated the possibility of several confounds between letter emotion and letter frequency by matching nonword pairs on letter frequency. We also ensured that the emotional letters in any given pair were approximately equal in their extremity of emotional resemblance. Finally, the current study does not simply use a single well-matched pair of letter strings but rather 24 such pairs, reducing the likelihood that any observed results are due to the idiosyncratic characteristics of a single pair of letter strings. Overall, we observed an influence of emotional resemblance on affective evaluations of 24 pairs of letter strings that were extremely similar aside from whether component letters resembled facial anger or facial happiness.

Study 4

The preceding studies demonstrate that people perceive emotional resemblance in Latin letters, that such emotional resemblance is incidentally perceived, and that it influences affective evaluations of letter strings. It is thus plausible that affective responses to real English words may derive, in part, from emotional resemblance in the letters that define those words. Study 4 examines the extent to which emotional resemblance in Latin letters influences affective responses to English words. Can words generate affect in perceivers by virtue of their resemblance to facial emotion?

Overview and Design

We purposefully selected 51 clearly positive and clearly negative words from the Affectively Normed English Word database (ANEW; Bradley & Lang, 1999). These words naturally varied in the extent to which they contained "happy" or "angry" letters. In Study 3, each trial consisted of the presentation of one word and participants' affective response to that word. Within 800 ms of exposure to a word, participants indicated whether the word made them feel "good" or "bad" on each of 153 trials.[8]

Our primary hypothesis was that—above and beyond effects of semantic valence— words would be more likely to evoke positive affect to the extent that those words were comprised of letters that resembled facial joy (vs. facial anger). We expected these effects regardless of whether the words themselves were semantically positive or negative, regardless of whether emotional resemblance of letters was indexed by human judgments or neural network ratings, and regardless of whether positive affect was measured as introspective judgments or RTs to make those judgments.

Method

Participants. Participants were recruited in introductory psychology courses at the University of Denver and given partial course credit in exchange for their participation. The sample of 122 participants (20 male) included participants between 18 and 23 years of age, of whom 66% were non-Hispanic Caucasian, 13% were Hispanic, 13% were Asian, 6% were mixed-race, 1% were African American, and the remaining participant did not report ethnicity.[9]

Words. We selected 51 words from the ANEW database, including 27 words with an average rating of 7 or higher (on a 1–9 scale ranging from negative to positive) and 24 with an average rating of 3.25 or below, constituting our sets of *positive* and *negative* words.[10] For each word, emotional resemblance was calculated as the average of the component letters' emotional resemblance scores (see Study 1 for letter calculations). We purposefully selected words so that neither word valence nor emotional resemblance scores were confounded with other possible predictors of affective responding, including word length, word arousal (ANEW ratings of arousal), word frequency (included in the ANEW database), and composite letter frequency (average frequency of lowercase letters in each word; Jones & Mewhort, 2004). Perhaps the most important criterion was that word valence be orthogonal to emotional resemblance scores.

To confirm that neither categorical word valence nor emotional resemblance was confounded with word arousal, word length, word frequency, or composite frequency of lowercase letters, we analyzed the relationship of word valence and emotional resemblance to those other factors and to each other. For word valence, we conducted *t tests* using positive versus negative word categories (correlations using continuous ANEW scores are reported after *t* tests). Word valence was not predictive of word length, $t(46) = .71$, $p = .48$; $r = .11$, $p = .45$; word arousal, $t(46) = .99$, $p = .33$; $r = -.15$, $p = .32$; word frequency, $t(46) = 1.55$, $p = .13$; $r = .22$, $p = .13$; or composite letter frequency, $t(46) = .43$, $p = .67$; $r = .02$, $p = .88$. Importantly, word valence was also not predictive of emotional resemblance scores, whether indexed via human ratings, $t(46) = -.89$, $p = .38$; $r = .13$, $p = .37$; or neural network activities, $t(46) = .38$, $p = .71$; $r = .06$, $p = .70$. Emotional resemblance scores based on human ratings were not significantly correlated with word length, $r = .04$, $p = .77$; word arousal, $r = .01$, $p = .98$; word frequency, $r = .03$, $p = .82$; nor composite letter frequency, $r = .16$, $p = .28$. Emotional resemblance scores based on neural network ratings were not significantly correlated with word length, $r = .12$, $p = .41$; word arousal,

[8] Much later in the session, participants rated the extent to which neutral faces caused them to feel good or bad—some faces had features that subtly resembled anger expressions and some faces had features that subtly resembled joy expressions. This manipulation included several confounded variables and experimenter errors. Hence, although results supported hypotheses, we do not present this data here.

[9] As in Studies 1–2, we expected a small effect size in Study 5. To achieve 75% power for an effect of letter emotion across two within-subjects levels of word valence (positive/negative) and with a continuous within-subjects factor of emotional resemblance, we sought 107 participants (sample size estimate calculated with G*Power). However, we continued collecting data until the end of the academic quarter in noticing that the first participants did not complete the study before leaving.

[10] Three of these words (lazy, heal, thrill) were spelled incorrectly and elicited responses that were quite different from the remainder in their category. For example, nearly 30% of responses to "heal" were *bad*. These words were therefore excluded from data analysis. The analyses reported in this section regarded the final 48 words but inferential statistics were nearly identical to those for the full 51 words.

$r = .05$, $p = .70$; word frequency, $r = -.09$, $p = .55$; nor composite letter frequency, $r = .19$, $p = .18$. All 51 words, along with their various ratings and frequencies, are listed in Table 2. Each word was rendered in black 24-point Arial font, centered on the computer screen and presented against a plain white background (see Figure 5).

Procedure. Participants completed the study in individual cubicles and were informed that they would take part in a study with a large number of trials. Participants were instructed that, for each trial, their task was to indicate whether the word made them feel good or bad. They were instructed to do so as quickly and accurately as possible, with the caveat that they render their judgment in less than a second (the actual response window was 200 ms to 800 ms). Participants received a warning if their response time was outside of the response window. All participants completed three blocks of 51 trials each such that words were presented in a

Table 2
Study 4 Words

Word	Word meaning	Letters	Affect	Arousal	Word frequency	Avg. letter frequency	H-emotion resemblance	N-emotion resemblance
Abuse	Negative	5	1.80	6.83	18	3,934,262	.4620	.2222
Afraid	Negative	6	2.00	6.67	57	3,809,931	−.0267	.0571
Alive	Positive	5	7.25	5.50	57	4,147,895	−.1220	−.1715
Bloody	Negative	6	2.90	6.41	8	2,718,283	.1833	.0353
Cancer	Negative	6	1.50	6.42	25	4,266,657	.1783	.0429
Caress	Positive	6	7.84	5.14	1	4,579,400	.3567	.6972
Cheer	Positive	5	8.10	6.12	8	4,907,581	.2200	−.1759
Crime	Negative	5	2.89	5.41	34	3,966,982	.0040	−.2853
Cute	Positive	4	7.62	5.53	5	4,205,817	.3250	.1271
Damage	Negative	6	3.05	5.57	33	3,885,557	.2883	−.2251
Dead	Negative	4	1.94	5.73	174	4,436,315	.2825	−.0473
Devil	Negative	5	2.21	6.07	25	3,569,103	−.2200	−.1500
Dirty	Negative	5	3.08	4.88	36	3,520,967	−.1400	−.1110
Dump	Negative	4	3.21	4.12	4	1,676,525	.0750	−.5490
Easy	Positive	4	7.10	4.48	125	4,563,468	.4550	.2750
Evil	Negative	4	3.23	6.39	72	3,868,924	−.2725	−.2222
Family	Positive	6	7.65	4.80	331	2,695,101	−.1867	−.0871
Fear	Negative	4	2.76	6.96	127	4,610,124	.1150	.0129
Friendly	Positive	8	8.43	5.11	61	3,528,076	−.1825	−.0792
Gift	Positive	4	7.77	6.14	33	3,134,674	−.1775	.2502
Hate	Negative	4	2.12	6.95	42	5,367,293	.1175	−.0546
Heal	Positive	4	7.09	4.77	2	4,628,658	.0325	−.0542
Hell	Negative	4	2.24	5.38	95	3,951,001	−.2450	.0780
Honest	Positive	6	7.70	5.32	47	4,942,736	.0667	.2683
Hope	Positive	4	7.05	5.44	178	4,170,636	.2525	−.4032
Humor	Positive	5	8.56	5.50	47	2,980,754	.0420	−.3657
Jewel	Positive	5	7.00	5.38	1	3,823,670	−.0600	−.0909
Joyful	Positive	6	8.22	5.98	1	1,886,760	.0917	.3157
Killer	Negative	6	1.89	7.86	21	3,662,369	−.3450	−.0285
King	Positive	4	7.26	5.51	88	2,682,603	−.3475	−.2425
Kiss	Positive	4	8.26	7.32	17	3,340,135	−.0975	.9319
Kitten	Positive	6	6.86	5.08	5	4,713,482	−.2983	−.0458
Lazy	Negative	4	4.38	2.65	9	2,236,349	−.3425	−.3111
Leader	Positive	6	7.63	6.27	74	4,968,064	.1533	−.0915
Love	Positive	4	8.72	6.44	232	3,919,408	−.0050	−.0948
Menace	Negative	6	2.88	5.52	9	4,785,133	.2000	−.2720
Merry	Positive	5	7.90	5.90	8	3,709,431	.0220	−.5458
Palace	Positive	6	7.19	5.10	38	4,006,424	.2483	−.0166
Party	Positive	5	7.86	6.69	216	3,445,408	.0680	−.2598
Poison	Negative	6	1.98	6.05	10	3,993,866	.1150	.1209
Puppy	Positive	5	7.56	5.85	2	1,288,420	.2280	−.7672
Scorn	Negative	5	2.84	5.48	4	3,909,876	.1160	.4512
Sick	Negative	4	1.90	4.29	51	2,783,686	−.1025	.5858
Stink	Negative	5	3.00	4.26	3	3,843,513	−.3460	.3489
Success	Positive	7	8.29	6.11	93	3,690,660	.4643	1.0014
Talent	Positive	6	7.56	6.27	40	5,184,950	−.1183	.1353
Thrill	Positive	6	8.05	8.02	5	3,705,856	−.4333	.1128
Toxic	Negative	5	2.10	6.40	3	3,369,656	−.5600	−.0927
Ugly	Negative	4	2.43	5.38	21	1,608,816	.1625	−.0925
Victim	Negative	6	2.18	6.06	27	3,107,252	−.3000	−.1755
Wise	Positive	4	7.52	3.91	36	4,367,760	−.1350	.0024

Note. Word frequencies, affect ratings and arousal ratings are from the ANEW database (Bradley & Lang, 1999). Human ("H") and neural ("N") emotion resemblance scores are mean-centered.

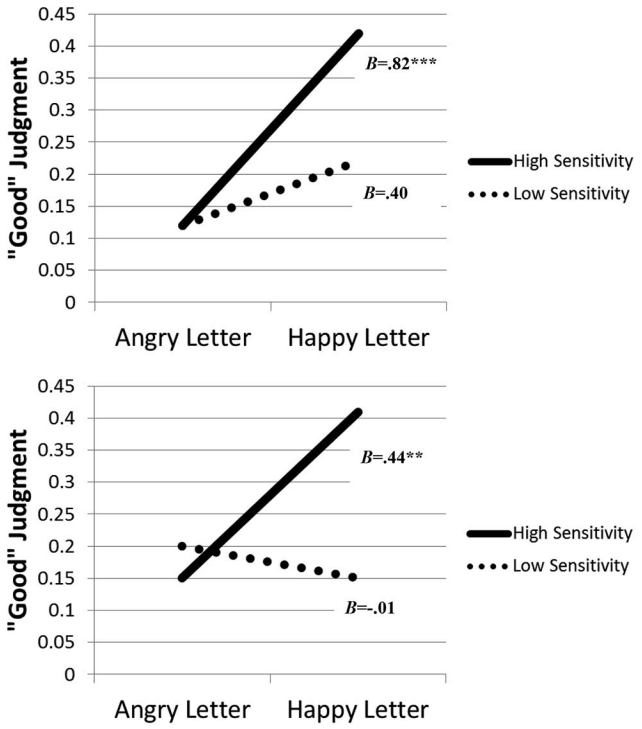


Figure 5. Log odds that a word would be identified as "positive" in Study 4, with unstandardized coefficients. Solid and dashed lines are plotted, respectively, at 1 *SD* above the facial index mean (high sensitivity) and 1 *SD* below the facial index mean (low sensitivity). The *x*-axis in the top panel indexes *human-rated* emotional resemblance plotted at 1 *SD* below (angry letter) and above (happy letter) the emotional resemblance mean. The *x*-axis in the bottom panel indexes *neural-network* emotional resemblance plotted at 1 *SD* below (angry letter) and above (happy letter) the emotional resemblance mean.

random order within block. Hence, each word was judged three times.

Results

Analytic strategy. Just over 4% of responses fell outside of the response window—these responses were either slower than 800 ms (4.3%) or faster than 200 ms (less than .1%). Hence, over 95% of responses fell within the response window. Data analyses reported below were restricted to responses that were within our response window. In what follows, we describe "good" responses as indicative of positive affective responding and "bad" responses as indicative of negative affective responding. Because good-bad responses were binary, more positive affective responding necessarily meant less negative affective responding. For the sake of clarity and word economy, we restrict our language to be in terms of "positive affect."

Data analyses were conducted using GEE (Zeger & Liang, 1986). In the current context, GEE not only accounts for a participant's average response for a particular word but also for the variance *among* that participant's responses to that particular word (each word was presented three times). We used binary logistic GEE to analyze affective responses to each word as a function of words' affective meaning (positive or negative), emotional resemblance score, and the interaction between these two variables. Because of the absence of neutral words, we elected to treat word valence as a categorical variable.

We instituted a response window in this study to generate sufficient variance for analyzing bivalent responses to words. Thus, our primary interest was in affective judgments made by participants. However, for exploratory purposes we also examined the relationship between emotional resemblance and response-times to make those judgments. To the extent that response-times in this paradigm also reflect emotional resemblance, we expected two-way interactions such that increasingly positive (happy) emotional resemblance scores predicted fast response-times to positive-meaning words and slow response-times to negative-meaning words.

Affective judgments. An initial GEE analysis was conducted on affective responses with word affective meaning as a categorical variable and emotional resemblance as a continuous predictor. Unsurprisingly, positive words ($M = .95$) elicited more positive judgments than did negative words ($M = .08$), $B = -5.37$, $SE = .13$, Wald $\chi^2 = 1606.47$, $p < .001$. More importantly, and as predicted, words with higher emotional resemblance scores elicited more positive affect than did words with lower emotional resemblance scores, $B = .38$, $SE = .12$, Wald $\chi^2 = 9.17$, $p = .002$, indicating that participants exhibited more positive affective responses when the letters in words resembled facial joy (vs. facial anger; see Figure 5). These findings were not qualified by a Word Valence × Emotional Resemblance interaction, $B = -.35$, $SE = .29$, Wald $\chi^2 = 1.39$, $p = .24$. All of these findings were replicated in analyses using neural network estimates of emotional resemblance (see online supplemental materials). Finally, treating word valence as continuous (according to ANEW scores) rather than categorical did not change the pattern of results observed here (e.g., $p < .001$ for word valence main effect; $p = .004$ for emotional resemblance main effect; $p = .81$ for interaction).

Response times. Our hypothesis for response-times was for an interaction between emotional resemblance and word valence. Specifically, we expected that increasingly "happy" emotional resemblance scores should speed responses to positive words but slow responses to negative words. To test this hypothesis, we restricted analyses to trials in which a participant's affective response was congruent with the affective meaning of the word (e.g., responding *good* to "wise"). Word valence and emotional resemblance were entered as Step 1 predictors and the Word Valence × Emotional Resemblance interaction was entered at Step 2 in GEE analyses.

Participants exhibited faster response-times to positive words ($M = 518$) than to negative words ($M = 538$), $B = 19.73$, $SE = 2.18$, Wald $\chi^2 = 81.67$, $p < .001$, as in prior research (Unkelbach, Fiedler, Bayer, Stegmüller, & Danner, 2008). There was also a marginal main effect of emotional resemblance, such that response-times were slower to the extent that words contained letters resembling facial joy, $B = 5.05$, $SE = 2.87$, Wald $\chi^2 = 3.09$, $p = .08$. Critically and as predicted, these effects were qualified by the interaction between word valence and emotional resemblance, $B = 26.72$, $SE = 5.69$, Wald $\chi^2 = 22.05$, $p < .001$. Finally, treating word valence as continuous (according to ANEW scores) rather than categorical did not change the pattern of results observed here (e.g., $p < .001$ for word valence main effect; $p =$

.06 for emotional resemblance main effect; $p = .001$ for interaction).

We decomposed the Emotional Resemblance × Categorical Word Valence interaction by analyzing response-times to positive words separately from response-times to negative words. Participants were faster to respond to positive words to the extent those words included letters resembling facial joy (vs. facial anger), $B = -9.24$, $SE = 4.36$, Wald $\chi^2 = 4.49$, $p = .03$. Conversely, participants were *slower* to respond to negative words to the extent those words included letters that resembled facial joy (vs. facial anger), $B = 17.47$, $SE = 3.74$, Wald $\chi^2 = 21.83$, $p < .001$. We replicated the two-way interaction with neural network estimates of emotional resemblance.[11]

Discussion

Affective responses to words depended on the extent to which component letters resembled facial joy or facial anger. On a speeded response task, words composed of letters resembling facial joy evoked more positive responses than did words composed of letters resembling facial anger. This was true whether letters' emotional resemblance was indexed by human ratings or neural network activities. Importantly, these effects emerged even though we selected words with emotional resemblance scores that were *not* associated with word frequency, letter frequency, word length, or arousal ratings. Moreover, we ensured that positive words were approximately equal to negative words with respect to emotional resemblance scores, word frequencies, letter frequencies, word length, and arousal ratings. In summary, emotional resemblance in words evoked reliable affective responses.

In an effort to encourage immediate affective responses, as opposed to more deliberate evaluations, we used a response window. Our analyses were thus focused on affective judgments rather than the amount of time it took to make those judgments. That is, response-windows limit variability in response-times and we were not confident that this limited variability could be explained by emotional resemblance. Nonetheless, response-times also depended on emotional resemblance in words. An interaction between word valence and emotional resemblance was observed for both human-rated emotional resemblance and neutral-network emotional resemblance. To the extent that words contained letters resembling facial joy but not facial anger, participants were *slower* to indicate that they felt "bad" in response to negative words and (in some analyses) *faster* to indicate that they felt "good" in response to positive words.

In general, Study 4 was designed to test whether the influence of emotional resemblance in letters extended from letter strings to real words. We observed robust support for the role of emotional resemblance in affective responses to words. Speeded affective judgments of words depended on emotional resemblance, whether such resemblance was indexed by human ratings or neural-network activities. Similarly, the speed of those judgments—even under a response window—depended on emotional resemblance, whether such resemblance was indexed by human ratings or neural-network activities.

General Discussion

The research described here is consistent with our theory that letter forms differ in the degree to which they resemble affective stimuli, and that people have positive responses to words to the extent letters resemble facial joy more than facial anger. Emotional resemblance weights were established empirically, through cross-cultural pretesting and by a neural network trained only to identify facial emotion. Using these weights, we found that emotional resemblance in written language was incidentally processed by participants in a flanker paradigm (Study 1), elicited unintentional affective responses from participants (Study 2), accounted for participants' affective responses to orthographically controlled letter strings (Study 3), and shaped participants' affective responses to real words (Study 4). Collectively, these findings are consistent with the theory that affective responses to written language depends on words' conceptual meaning as well as the resemblance of their visual form to affective stimuli.

Implications for Perceptual Resemblance, Attitudes, and Symbolic Communication

When people have affective responses to what they've read in novels, product labels, or attitude experiments, it is typically assumed that those responses owe to the conceptual meaning of words. In contrast, the evidence presented here is consistent with the view that people have affective responses to the visual signature of any given word, due to the emotional resemblance of that signature. Accordingly, the current work may contribute to scientific understanding of symbolic communication, human judgment and decision-making, as well as research on implicit attitudes. Perhaps the most direct implications, however, are for effects of perceptual similarity. In this section, we describe how the current work may inform related phenomena.

Perceptual resemblance. We are not the first to demonstrate that objects with features resembling facial emotion can evoke responses similar to those evoked by facial emotions themselves. There is evidence that simple geometric shapes that resemble features of happy or angry faces can generate similarly speedy attentional and affective responses (e.g., Watson et al., 2012; see below). Moreover, there is evidence that people exhibit consensus in judging the resemblance of everyday objects to facial emotion and then evaluate those objects accordingly (e.g., objects with a happy expression are evaluated positively; Aggarwal & McGill, 2007; Ichikawa et al., 2011; Landwehr et al., 2011; Windhager et al., 2008). To date, this research has focused on objects that include a prototypical face-configuration: two eye-like features centered above or around a mouth-like feature, creating an inverted pyramid-like shape (e.g., electrical outlets; Ichikawa et al., 2011). However, the current work demonstrates that emotional resemblance effects can extend to objects that have *partial* overlap with facial emotion features. People are able to locate facial emotion characteristics in Latin letters despite the obvious physical differences between faces and letters, and those subjective percepts of emotional resemblance very quickly influence responses to those letters (Studies 1–2). Although emotional resemblance effects for

[11] We also replicated the simple effect of response-times to negative words, in that participants were slower to respond to negative words to the extent those words included letters that resembled facial joy. However, unlike for human ratings, response times to positive words were not influenced by emotional resemblance, as estimated by neural network ratings. See online supplementary materials for discussion.

triangles and electrical outlets may have a narrow influence in daily life, written language is critical to optimal functioning in many civilizations and its emotional resemblance effects should not be considered trivial by any metric. Emotional resemblance in written words influenced participants' affective responses, leaving open the possibility that the resemblance of letters to facial emotion influences evaluations of everything from novels and webpages to political candidates and products at the grocery store (see next section).

Beyond emotional resemblance per se, the current work extends research on resemblance in social perception. A healthy and burgeoning literature on face perception suggests that resemblance can play an important role in social judgment. For example, the presence of male-typical features on an Asian face can interfere with accuracy in race judgment (Freeman & Ambady, 2011; Johnson, Freeman, & Pauker, 2012). The presence of anger on a female face can interfere with accuracy in gender judgment (Hess, Adams, Grammer, & Kleck, 2009). The presence of Afrocentric features on a "White" face can cause American perceivers to evaluate that face negatively and in terms African American stereotypes (Blair, Judd, & Fallman, 2004) and there is evidence that trait impressions of female faces are positive to the extent that those faces have features that resemble the faces of human babies (for a review, see Zebrowitz, 1997).

The current research links resemblance effects in face perception with visual word perception, contributing to a burgeoning literature on the similarities and differences in these two perceptual domains. In general, the "fusiform face area" (FFA) is especially active during face perception and the "visual word form area" (VWFA; cf. Dehaene & Cohen, 2011) is especially active during word identification, yet these two brain areas are located in different brain hemispheres. Hence, the overlap between affective processing of faces and letters might suggest moderation by factors known to moderate hemispheric specialization, such as handedness.

Conversely, the results also support emerging perspectives on interactions between face perception and visual word perception. Behrmann and Plaut (2013), for example, explain that both face recognition and word recognition require high-acuity vision of overlearned stimuli. Indeed, neuroscientific evidence suggests that faces, perhaps more than any other visual object, compete with linguistic characters for processing resources. Specifically, these two types of stimuli seem to compete for processing space in or near the visual word form area (VWFA). In fact, one study demonstrated that increases in literacy accompanied the displacement of face processing from the VWFA (left fusiform) to the right fusiform gyrus (Dehaene et al., 2010). Other studies demonstrate that increased childhood performance in digit or letter identification is accompanied by decreased activation to faces in the left hemisphere (Dundas, Plaut, & Behrmann, 2013). Finally, compared with matched controls, prosopagnosics have difficulty in processing words and pure alexics have difficulty processing faces (Behrmann & Plaut, 2013). Although debate remains about the domain-specificity of face and word perception, the current set of studies builds on evidence for the interaction of neural mechanisms involved in the visual perception of language and those involved in the visual perception of faces to demonstrate a role for face perception (facial emotion perception) in the evaluation of written language.

Judgments and attitudes. The ability to accurately perceive written language is essential to making choices in Western culture: Voters see candidate names before selecting one on a voting ballot, shoppers see brand names on medicine packages before selecting one for purchase, and employers read applications before selecting interview candidates. Accordingly, political candidates, products, and job applications may receive more positive evaluations to the extent that the written letters in candidate names, product names, and job applications resemble facial joy more than facial anger. The weights generated in pretesting (see online supplementary materials) can be used in future research to explore how emotional resemblance in written language shapes judgment and decision-making.

Additionally, it is noteworthy that research on attitudes has been revolutionized by implicit measures that require speedy evaluative responses to words and images, often briefly presented words and images. By limiting exposure times and/or encouraging speedy responses, researchers limit deliberative responses to stimuli and instead aim to capture participants' basic cognitive associations between concepts. Yet with limited exposure times, words presented in priming tasks may elicit affective responses to words' visual features rather than (or in addition to) words' conceptual meaning (e.g., Abrams & Greenwald, 2000). Similarly, responses on implicit association tasks may reflect the visual features of word stimuli as much as conceptual meaning. In both cases, measures of implicit attitudes may include variance due to emotional resemblance.

Symbolic communication. The history of written language demonstrates that humans are sensitive to visual imagery in symbolic communication. From ancient Egypt and Mesopotamia to many other regions and times, the development of most written languages has been preceded by systems of symbolic communication that were built on visual imagery. These proto-writing systems may seem primitive yet the current work suggest that humans are sensitive to visual imagery in written communication, exhibiting speedy affective processing of such imagery. The current research provides evidence that people very quickly process affective imagery in written language, and to the extent people prioritize speed in text messages and emails, they may convey affect by more frequently selecting words with the corresponding emotional resemblance.

Limitations and Future Directions: Specificity and Reading Processes

The current work represents the first investigation—to our knowledge—of emotional resemblance in written language. Accordingly, several unanswered questions remain and we highlight those here.

Specificity to facial emotion. Emotional resemblance effects may be limited to facial emotion or may extend to other emotional stimuli. Extant research suggests that People are exceptionally sensitive to facial emotions, reliably perceive facial emotions early in life, exhibit affective responses to facial emotions even before they know what they see, and see facial emotions in shapes such as triangles and household products (Aggarwal & McGill, 2007; Ichikawa et al., 2011; Landwehr et al., 2011; Watson et al., 2012; Windhager et al., 2008). Accordingly, facial emotions may represent a unique visual basis for emotional resemblance effects in

written language. Yet people do exhibit speedy affective responses to other stimuli that have a consistent shape, such as guns and flowers. There are at least two ways in which the effects observed here may generalize to other affective stimuli. First, the key features of facial emotion may include basic perceptual features common to many affective stimuli, such that visual features of angry faces may be common to negative stimuli ranging from snakes and spiders to guns and dead bodies. If so, the phenomenon we have discovered here may not be specific to facial emotions but rather features common to many affective stimuli. Second, it is possible that people become highly sensitive to the visual features of other affective stimuli, especially if those stimuli are frequently encountered. However, at present we *can* conclude that resemblance to facial emotion influenced participants' affective/evaluative responses to written language.

Reading. In the current work we explored the influence of emotional resemblance on responses to Latin letters, well-controlled letter strings, and full words. The external validity of these findings is not trivial, as people read only one or two words at a time when identifying political candidates, product names, or when participating in priming studies. Given the speed with which people perceive facial emotions (Dimberg et al., 2000; McAndrew, 1986; Murphy & Zajonc, 1993), even in written words (Study 3), we expect these findings to extend to evaluations of novels, job applications, and persuasive messages. However, the processes involved in reading sentences and paragraphs are not identical to reading isolated words, so it would be premature to draw the conclusion that emotional resemblance effects operate in the reading of longer texts. As but one example, readers eyes' often skip over short words (e.g., as, if) and extract the meaning of a sentence or word without close analysis of the letters in those words (see Rayner, 2009). In this context, visual attention may moderate the influence of emotional resemblance (only attended words may be influential) or other reading processes may interfere with the impact of strictly visual information on affective responses.

Moreover, when people read English outside of the laboratory they often do not focus on their evaluative responses to each word they read. Accordingly, the evaluative tasks used in the current research may overestimate the degree to which emotional forms shape affective responses during informal reading. Typical reading is often focused upon understanding the text, and in these circumstances, it is possible that emotional resemblance would fail to influence affective responses to words. However, there are many contexts in which people *are* focused on their evaluations of written words, as noted above (e.g., candidate names in voting booths, product names in grocery stores, text of a written opinion article), and for which emotional resemblance is likely to be relevant. This limitation can of course be applied to other studies in which evaluation and attitudes are measured, and seems especially relevant in any study in which realistic stimuli—such as faces or words—are used. Hence, although the current results may not generalize to all reading contexts, they do provide evidence consistent with our theory that affective processes are sensitive to pictorial meaning in written language. Nonetheless, it will be important to explore how the visual signature of words impacts readers' affective experiences of reading larger texts in nonevaluative contexts.

Moderators. Factors that might increase the influence of emotional resemblance on daily decision-making include factors that are thought to increase the role of intuition and affect in decision-making. Hence, uncertainty or ambivalence about the evaluated object (e.g., Tversky & Kahneman, 1974), limited cognitive resources (Bargh, 1994), and an explicit goal to evaluate the object (Klauer & Musch, 2002; Klinger et al., 2000) might all favor a role for the emotional resemblance in affective responses to written language. Moreover, recency and/or frequency of exposure to faces or facial emotion might increase the activity of face (emotion) processing mechanisms (e.g., via procedural priming; Forster, Liberman, & Friedman, 2009) and thus increase the likelihood of emotional resemblance effects in reading.

The hypothesized moderators reflect the idea that certain contexts increase the likelihood that emotional resemblance influences choices outside of the laboratory. For example, crowds require perceivers to engage in repeated face processing such that emotional resemblance may be especially likely to influence choices in crowded shopping markets, voting venues, and so on. These proposed moderators are also informative with respect to many other phenomena studied by psychologists. One particularly interesting application of the proposed moderation is with respect to persuasive messages. When people read such messages they often engage in evaluative processing of the message thesis, yet such evaluation can often occur when people lack the motivation or ability to thoroughly process the message. Such low-elaboration processing (also known as heuristic processing) makes people especially susceptible to superficial influences (Chen & Chaiken, 1999; Petty & Cacioppo, 1986). We suspect that emotional resemblance can influence affective responses to persuasive messages, though we suspect that such influence is limited to people who are engaging in heuristic processing while reading those messages.

Conclusion

Extensive pretesting and three studies provide converging evidence consistent with the theory that the visual perception and evaluation of Latin letters, letter-strings, and English words depends on the perceived resemblance of Latin letters to facial emotion.

References

Abrams, R. L., & Greenwald, A. G. (2000). Parts outweigh the whole (word) in unconscious analysis of meaning. *Psychological Science, 11,* 118–124. http://dx.doi.org/10.1111/1467-9280.00226

Adams, W. J., Gray, K. L. H., Garner, M., & Graf, E. W. (2010). High-level face adaptation without awareness. *Psychological Science, 21,* 205–210. http://dx.doi.org/10.1177/0956797609359508

Aggarwal, P., & McGill, A. L. (2007). Is that car smiling at me? Schema congruity as a basis for evaluating anthropomorphized products. *The Journal of Consumer Research, 34,* 468–479. http://dx.doi.org/10.1086/518544

Alpers, G. W., & Gerdes, A. B. M. (2007). Here is looking at you: Emotional faces predominate in binocular rivalry. *Emotion, 7,* 495–506. http://dx.doi.org/10.1037/1528-3542.7.3.495

Amari, S. I., Murata, N., Müller, K. R., Finke, M., & Yang, H. H. (1996). In D. S. Touretzky, M. C. Mozer, & M. E. Hasselmo (Eds.), *Advances in neural information processing systems* (pp. 176–182). Cambridge, MA: MIT Press.

Aronoff, J., Woike, B. A., & Human, L. M. (1992). Which are the stimuli in facial displays of anger and happiness? Configural bases of emotion recognition. *Journal of Personality and Social Psychology, 62,* 1050–1066. http://dx.doi.org/10.1037/0022-3514.62.6.1050

Bar, M., & Neta, M. (2006). Humans prefer curved visual objects. *Psychological Science, 17,* 645–648. http://dx.doi.org/10.1111/j.1467-9280.2006.01759.x

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (2nd ed., pp. 1–40). Hillsdale, NJ: Erlbaum.

Bargh, J. A., Chaiken, S., Raymond, P., & Hymes, C. (1996). The automatic evaluation effect: Unconditional automatic attitude activation with a pronunciation task. *Journal of Experimental Social Psychology, 32,* 104–128.

Barratt, D., & Bundesen, C. (2012). Attentional capture by emotional faces is contingent on attentional control settings. *Cognition and Emotion, 26,* 1223–1237. http://dx.doi.org/10.1080/02699931.2011.645279

Behrmann, M., & Plaut, D. C. (2013). Distributed circuits, not circumscribed centers, mediate visual recognition. *Trends in Cognitive Sciences, 17,* 210–219. http://dx.doi.org/10.1016/j.tics.2013.03.007

Bjork, E. L., & Murray, J. T. (1977). On the nature of input channels in visual processing. *Psychological Review, 84,* 472–484. http://dx.doi.org/10.1037/0033-295X.84.5.472

Blair, I. V., Judd, C. M., & Fallman, J. L. (2004). The automaticity of race and Afrocentric facial features in social judgments. *Journal of Personality and Social Psychology, 87,* 763–778. http://dx.doi.org/10.1037/0022-3514.87.6.763

Bradley, M. M., & Lang, P. P. J. (1999). *Affective Norms for English Words (ANEW): Instruction manual and affective ratings*. Gainesville, FL: The Center for Research in Psychophysiology, University of Florida.

Calvo, M. G., & Nummenmaa, L. (2007). Processing of unattended emotional visual scenes. *Journal of Experimental Psychology: General, 136,* 347–369.

Chen, S., & Chaiken, S. (1999). The heuristic-systematic model in its broader context. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 73–96). New York, NY: Guilford Press.

Codispoti, M., Bradley, M. M., & Lang, P. J. (2001). Affective reactions to briefly presented pictures. *Psychophysiology, 38,* 474–478.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159. http://dx.doi.org/10.1037/0033-2909.112.1.155

Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences, 15,* 254–262. http://dx.doi.org/10.1016/j.tics.2011.04.003

Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Nunes Filho, G., Jobert, A., . . . Cohen, L. (2010). How learning to read changes the cortical networks for vision and language. *Science, 330,* 1359–1364. http://dx.doi.org/10.1126/science.1194140

Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychological Science, 11,* 86–89. http://dx.doi.org/10.1111/1467-9280.00221

Duckworth, K. L., Bargh, J. A., Garcia, M., & Chaiken, S. (2002). The automatic evaluation of novel stimuli. *Psychological Science, 13,* 513–519.

Dundas, E. M., Plaut, D. C., & Behrmann, M. (2013). The joint development of hemispheric lateralization for words and faces. *Journal of Experimental Psychology: General, 142,* 348–358.

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods, 1,* 170–177. http://dx.doi.org/10.1037/1082-989X.1.2.170

Egeth, H. E., & Santee, J. L. (1981). Conceptual and perceptual components in interletter inhibition. *Journal of Experimental Psychology, 7,* 506–517.

Enroth-Cugell, C., & Robson, J. G. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *The Journal of Physiology, 187,* 517–552. http://dx.doi.org/10.1113/jphysiol.1966.sp008107

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon identification of a target letter in a non-search task. *Perception & Psychophysics, 16,* 143–149. http://dx.doi.org/10.3758/BF03203267

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39,* 175–191. http://dx.doi.org/10.3758/BF03193146

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology, 50,* 229–238.

Fenske, M. J., & Eastwood, J. D. (2003). Modulation of focused attention by faces expressing emotion: Evidence from flanker tasks. *Emotion, 3,* 327–343. http://dx.doi.org/10.1037/1528-3542.3.4.327

Fischer, S. R. (2001). *A history of writing*. London, UK: Reaktion Books.

Fischer, S. R. (2003). *A history of reading*. London, UK: Reaktion Books.

Forster, J., Liberman, N., & Friedman, R. (2009). What do we prime? On distinguishing between semantic priming, procedural priming, and goal priming. In E. Morsella, J. A. Bargh, & P. M. Gollwitzer (Eds.), *Oxford handbook of human action* (pp. 173–193). New York, NY: Oxford University Press.

Fox, E., Russo, R., & Dutton, K. (2002). Attentional bias for threat: Evidence for delayed disengagement from emotional faces. *Cognition and Emotion, 16,* 355–379. http://dx.doi.org/10.1080/02699930143000527

Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review, 118,* 247–279. http://dx.doi.org/10.1037/a0022327

Fukumizu, K. (1998). Effect of batch learning in multilayer neural networks. In S. Usui & T. Omori (Eds.), *ICONIP' 98 Proceedings*. Paper presented at the 5th International Conference on Neural Information Processing, Kitakyushu, Japan, October 21–23. Burke, VA: IOS Press.

Gelb, I. J. (1963). *A study of writing*. Chicago: University of Chicago Press.

Globisch, J., Hamm, A. O., Esteves, F., & Öhman, A. (1999). Fear appears fast: Temporal course of startle reflex potentiation in animal fearful subjects. *Psychophysiology, 36,* 66–75.

Gomez, P., Ratcliff, R., & Perea, M. (2008). The overlap model: A model of letter position coding. *Psychological Review, 115,* 577–600. http://dx.doi.org/10.1037/a0012667

Grainger, J., & Holcomb, P. J. (2010). Neural constraints on a functional architecture for word recognition. In P. L. Cornelissen, P. C. Hansen, M. L. Kringlebach, & K. Pugh (Eds.), *The neural basis of reading* (pp. 3–32). New York: Oxford University Press.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74,* 1464–1480.

Healy, A. F. (1994). Letter detection: A window to unitization and other cognitive processes in reading text. *Psychonomic Bulletin & Review, 1,* 333–344. http://dx.doi.org/10.3758/BF03213975

Hess, U., Adams, R. B., Grammer, K., & Kleck, R. E. (2009). Face gender and emotion expression: Are angry women more like men? *Journal of Vision, 9*(12), 19.

Horstmann, G., Borgstedt, K., & Heumann, M. (2006). Flanker effects with faces may depend on perceptual as well as emotional differences. *Emotion, 6,* 28–39. http://dx.doi.org/10.1037/1528-3542.6.1.28

Ichikawa, H., Kanazawa, S., & Yamaguchi, M. K. (2011). Finding a face in a face-like object. *Perception, 40,* 500–502. http://dx.doi.org/10.1068/p6926

Johnson, K. L., Freeman, J. B., & Pauker, K. (2012). Race is gendered: How covarying phenotypes and stereotypes bias sex categorization. *Journal of Personality and Social Psychology, 102,* 116–131. http://dx.doi.org/10.1037/a0025335

Jones, M. N., & Mewhort, D. J. K. (2004). Case-sensitive letter and bigram frequency counts from large-scale English corpora. *Behavior Research Methods, Instruments, & Computers, 36,* 388–396. http://dx.doi.org/10.3758/BF03195586

Klauer, K. C., & Musch, J. (2002). Goal-dependent and goal-independent effects of irrelevant evaluations. *Personality and Social Psychology Bulletin, 28,* 802–814. http://dx.doi.org/10.1177/0146167202289009

Klinger, M. R., Burton, P. C., & Pitts, G. S. (2000). Mechanisms of unconscious priming: I. Response competition, not spreading activation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 441–455. http://dx.doi.org/10.1037/0278-7393.26.2.441

Kouider, S., & Dehaene, S. (2007). Levels of processing during non-conscious perception: A critical review of visual masking. *Philosophical Transactions of the Royal Society B: Biological Sciences, 362,* 857–875.

Landwehr, J. R., McGill, A. L., & Herrmann, A. (2011). It's got the look: The effect of friendly and aggressive "facial" expressions on product liking and sales. *Journal of Marketing, 75,* 132–146. http://dx.doi.org/10.1509/jmkg.75.3.132

Lundqvist, D., Esteves, F., & Ohman, A. (2004). The face of wrath: The role of features and configurations in conveying social threat. *Cognition and Emotion, 18,* 161–182. http://dx.doi.org/10.1080/02699930244000453

Lundqvist, D., & Ohman, A. (2005). Emotion regulates attention: The relation between facial configurations, facial emotion, and visual attention. *Visual Cognition, 12,* 51–84. http://dx.doi.org/10.1080/13506280444000085

Maratos, F. A., Mogg, K., & Bradley, B. P. (2008). Identification of angry faces in the attentional blink. *Cognition and Emotion, 22,* 1340–1352.

Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London Series B*. Royal Society, *207,* 187–217.

Matsumoto, D., LeRoux, J., Wilson-Cohn, C., Raroque, J., Kooken, K., Ekman, P., . . . Goh, A. (2000). A new test to measure emotion recognition ability: Matsumoto and Ekman's Japanese and Caucasian brief affect recognition test (JACBART). *Journal of Nonverbal Behavior, 24,* 179–209. http://dx.doi.org/10.1023/A:1006668120583

McAndrew, F. T. (1986). A cross-cultural study of recognition thresholds for facial expressions of emotion. *Journal of Cross-Cultural Psychology, 17,* 211–224. http://dx.doi.org/10.1177/0022002186017002005

Medler, D. A., & Binder, J. R. (2005). *MCWord: An on-line orthographic database of the English language*. Retrieved from http://www.neuro.mcw.edu/mcword/

Murphy, S. T., & Zajonc, R. B. (1993). Affect, cognition, and awareness: Affective priming with optimal and suboptimal stimulus exposures. *Journal of Personality and Social Psychology, 64,* 723–739. http://dx.doi.org/10.1037/0022-3514.64.5.723

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology, 89,* 277–293. http://dx.doi.org/10.1037/0022-3514.89.3.277

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology, 19,* 123–205. http://dx.doi.org/10.1016/S0065-2601(08)60214-2

Pflughaupt, L. (2007). *Letter by letter: An alphabetical miscellany*. New York, NY: Princeton Architectural Press.

Pomerantz, J. R., Pristach, E. A., & Carson, C. E. (1988). Attention and object perception. In: B. Shepp B & S. Ballesteros (Eds) *Object perception: Structure and process* (pp. 53–89). Hillsdale, NJ: Erlbaum.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 62,* 1457–1506. http://dx.doi.org/10.1080/17470210902816461

Richman, H. B., & Simon, H. A. (1989). Context effects in letter perception: Comparison of two theories. *Psychological Review, 96,* 417–423. http://dx.doi.org/10.1037/0033-295X.96.3.417

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review, 89,* 60–94. http://dx.doi.org/10.1037/0033-295X.89.1.60

Sacks, D. (2003). *Letter perfect: The marvelous history of our alphabet from A to Z*. New York, NY: Broadway Books.

Schmidt, F., & Schmidt, T. (2013). No difference in flanker effects for sad and happy schematic faces: A parametric study of temporal parameters. *Visual Cognition, 21,* 382–398. http://dx.doi.org/10.1080/13506285.2013.793221

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86,* 420–428. http://dx.doi.org/10.1037/0033-2909.86.2.420

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97–136. http://dx.doi.org/10.1016/0010-0285(80)90005-5

Trigger, B. G. (1998). Writing systems: A case study in cultural evolution. *Norwegian Archaeological Review, 31,* 39–62.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131. http://dx.doi.org/10.1126/science.185.4157.1124

Unkelbach, C., Fiedler, K., Bayer, M., Stegmüller, M., & Danner, D. (2008). Why positive information is processed faster: The density hypothesis. *Journal of Personality and Social Psychology, 95,* 36–49. http://dx.doi.org/10.1037/0022-3514.95.1.36

van Leeuwen, C., & Lachmann, T. (2004). Negative and positive congruence effects in letters and shapes. *Perception & Psychophysics, 66,* 908–925. http://dx.doi.org/10.3758/BF03194984

Watson, D. G., Blagrove, E., Evans, C., & Moore, L. (2012). Negative triangles: Simple geometric shapes convey emotional valence. *Emotion, 12,* 18–22. http://dx.doi.org/10.1037/a0024495

Weisbuch, M., & Ambady, N. (2008). Affective divergence: Automatic responses to others' emotions depend on group membership. *Journal of Personality and Social Psychology, 95,* 1063–1079. http://dx.doi.org/10.1037/a0011993

Windhager, S., Slice, D. E., Schaefer, K., Oberzaucher, E., Thorstensen, T., & Grammer, K. (2008). Face to face: The perception of automotive design. *Human Nature, 19,* 331–346. http://dx.doi.org/10.1007/s12110-008-9047-z

Yap, M. J., Balota, D. A., Sibley, D. E., & Ratcliff, R. (2012). Individual differences in visual word recognition: Insights from the English Lexicon Project. *Journal of Experimental Psychology: Human Perception and Performance, 38,* 53–79.

Yoon, K. L., Hong, S. W., Joormann, J., & Kang, P. (2009). Perception of facial expressions of emotion during binocular rivalry. *Emotion, 9,* 172–182. http://dx.doi.org/10.1037/a0014714

Young, R. A. (1987). The Gaussian derivative model for spatial vision: I. Retinal mechanisms. *Spatial Vision, 2,* 273–293. http://dx.doi.org/10.1163/156856887X00222

Zebrowitz, L. A. (1997). *Reading faces: Window to the soul?* Boulder, CO: Westview Press.

Zeger, S. L., & Liang, K. Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics, 42,* 121–130. http://dx.doi.org/10.2307/2531248

Received June 20, 2016
Revision received November 5, 2018
Accepted April 16, 2019 ■