Nonverbal Dialects and Accents in Facial Expressions of Emotion
Hillary Anger Elfenbein

The online version of this article can be found at:
http://emr.sagepub.com/content/5/1/90

Published by:
http://www.sagepub.com

On behalf of:
International Society for Research on Emotion

Additional services and information for Emotion Review can be found at:

Email Alerts: http://emr.sagepub.com/cgi/alerts

Subscriptions: http://emr.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

>> Version of Record - Jan 2, 2013

What is This?
Nonverbal Dialects and Accents in Facial Expressions of Emotion

Hillary Anger Elfenbein
Olin School of Business, Washington University in St. Louis, USA

Abstract

This article focuses on a theoretical account integrating classic and recent findings on the communication of emotions across cultures: a dialect theory of emotion. Dialect theory uses a linguistic metaphor to argue emotion is a universal language with subtly different dialects. As in verbal language, it is more challenging to understand someone speaking a different dialect—which fits with empirical support for an in-group advantage, whereby individuals are more accurate judging emotional expressions from their own cultural group versus foreign groups. Dialect theory has sparked controversy with its implications for dominant theories about cross-cultural differences in emotion. This article reviews the theory, its mounting body of evidence, evidence for alternative accounts, and practical implications for multicultural societies.

Keywords
emotion, expression, facial expression, nonverbal accents, nonverbal dialects, perception, recognition

Introductory psychology textbooks tell the story of Ekman (1972) and Izard (1971), who took black-and-white photographs of American facial expressions on round-the-world tours with the goal of demonstrating that emotions are universal. Many disparate observations have been made of these data. The first, from the original researchers, was that participants choosing among six multiple choices achieved far better performance than the 16.7% expected by chance—which demonstrates basic universality. The second, to be discussed later, was that some cultures did better than others (Matsumoto, 1989). The third observation is that the best performers were from the nation where the photographs originated, followed by those most culturally similar. This is the key observation behind dialect theory.

As early as 1964, Tomkins and McCarter articulated the metaphor that cultural differences in emotional expression are like “dialects” of the “more universal grammar of emotion” (p. 127). Dialect theory takes seriously this linguistic metaphor for communicating emotion through nonverbal cues, just as dialects of a language can differ subtly in accent, grammar, and vocabulary—for example, consider American versus British English. As in verbal language, it is more challenging to understand someone speaking a different dialect. To be sure, the dialects of a language are still mutually intelligible, but one must listen more carefully, and some of the meaning can get lost along the way.

It is important to distinguish the two interconnected processes within dialect theory: First, members of different cultural groups have different styles of generating nonverbal cues, which are systematic even if subtle. This process is also called expression, although that term implies intentionality whereas one can also generate cues without trying. Second, individuals tend to judge other people’s cues based on their own cultural style. This judgment process is also called recognition, although likewise this term implies intentionality yet can also be implicit. Dialect theory argues that accuracy breaks down through mirror-image cultural differences on both sides of the process. Ultimately, communication accuracy suffers to the extent that there is a mismatch between the style of display produced and the style expected by the perceiver (see Brunswick’s [1955] lens model). Even the term communication can be preemptive in that it implies a conscious goal—note that dialect theory is intended to describe both deliberate and spontaneous processes although,
as discussed later, the vast majority of existing evidence tests posed expressions.

In keeping with the distinction between expression and recognition, dialect theory distinguishes between nonverbal accents and dialects. Nonverbal accents are any differences across cultures in the appearance of an emotional expression. Nonverbal dialects are a subset of these accents—namely, the accents that also impede accurate recognition. In the linguistic metaphor, typically an accent is noticeable yet unchallenging, whereas a difference in dialect can create difficulty in understanding another person’s speech. Also consistent with the linguistic metaphor, one can speculate that the distinction between dialect and accent may become fluid in the face of cross-group contact, yet, among less acquainted groups, even small differences could create difficulty.

**Display and Decoding Rules**

In contrast with dialect theory, the most prominent alternative accounts for cultural differences in emotional expression and recognition focus on deception. Starting with the expression process, in the dialect model culturally specific cues can still appear even while trying to be as clear as possible. By contrast, Klimeberg (1938) and Ekman’s (1972) concept of display rules focuses on deliberately deceptive emotion regulation among individuals from cultures that emphasize social harmony. Ekman (1972) defined display rules as conscious management techniques to deintensify, intensify, neutralize, and mask displays with qualitatively different displays, and argued that members of each culture would express their emotions in exactly the same way if some groups were not constantly monitoring themselves and adjusting their displays to fit social norms. Ekman grounded this discussion in a summary of Friesen’s unpublished dissertation (described in Ekman, 1972), in which Japanese participants purportedly masked their facial displays in the presence of an observer, while American participants did not. Although these findings have been controversial due to incomplete reporting, including an additional experimental condition not mentioned in Ekman’s summary (Fridlund, 1994), this notion has been a powerful idea within cross-cultural research in emotion. By contrast, the dialect theory argues that no such deliberate deception is required to produce cultural differences in emotional expression style.

Likewise, in the recognition side of the process, individuals can stumble over roadblocks in judging others’ emotions even while trying to be as perceptive as possible. By contrast, Matsumoto’s (1989) extension of display rules, known as decoding rules, focuses on deliberately deceptive regulation. In attempting to explain the cultural differences in Ekman’s (1972) and Izard’s (1971) studies, he argued that Americans are simply more effective at recognizing emotions because Americans do not suppress their true understanding of emotional displays out of concern for group harmony. By contrast, according to dialect theory, cultural differences in recognition can still emerge when people are trying to be as perceptive as possible. No one needs to deliberately ignore others’ emotions, but instead they can stumble over differences in the styles of expressions from cultural out-groups.

To be sure, differences in dialects can exist alongside the deliberate emotion regulation strategies represented by display and decoding rules. The reason why these have often been considered in opposition is the inaccurate claim that display and decoding rules alone can explain the body of findings on cultural differences in emotion recognition (Matsumoto, 2002). As we see in the following sections, they cannot—if not dialect theory, something else is necessary to close the gap.

**Empirical Evidence**

Initial evidence for dialect theory comes from the in-group advantage, whereby individuals are more accurate when judging emotional expressions from their own cultural group versus foreign cultural groups. We demonstrated this in a meta-analysis across 182 independent samples in 87 articles, with the majority of studies examining facial expressions (Elfenbein & Ambady, 2002b). Many samples came from the very same classic articles that were intended to demonstrate universality, but many others were unintentionally cross-cultural, with investigators borrowing research protocols from international colleagues without hypotheses about the cultural differences that might result. Interestingly, in-group advantage did not differ significantly across research team, nor did it vary along methodological lines (which frequently coincided with research teams). Issues of language could not explain away in-group advantage, because it existed across cultural groups sharing the same native language. Issues of racial bias could not explain away in-group advantage either, because it existed among all-Caucasian groups. Indeed, the only significant moderator was cross-cultural exposure—that is, in-group advantage was smaller when judging more familiar cultural groups.

The earliest research establishing accents used a different design that was particularly stringent (Marsh, Elfenbein, & Ambady, 2003). While conducting the meta-analysis, I noticed that the brochure for Matsumoto and Ekman’s (1988) collection of Japanese and Caucasian facial expressions included a combination of Japanese and Japanese Americans. In every other way, these stimuli were perfectly controlled—the same lighting, clothing, and so forth. Indeed, the developers instructed participants exactly how to move their facial muscles, so that the resulting expressions should have been the same in every way other than the apparent ethnicity of the face. Even so, collaborator Abby Marsh and I found informally that we could tell the Japanese apart from the Japanese Americans and, in the resulting experiment, participants could too. Participants were no more accurate than chance when attempting to distinguish the nationality of neutral photographs taken of the same actors—which rules out nuisance explanations such as hair style or the possible effects on facial appearance of diet, climate, and so on. However, when these same targets attempted to pose emotional expressions, their nationality became visible to participants.
We interpreted this as strong support for nonverbal accents—that is, even in a set of facial expressions for which researchers attempted to dampen every possible cultural difference in appearance, these cultural differences still leaked through. Note that these accents were not dialects, because recognition was not impaired.

In the decade since the initial research, evidence has been mounting and increasingly direct for the dialect theory. First, my colleagues and I linked the in-group advantage directly to differences in the appearance of expressions using a novel methodology (Elfenbein, Mandal, Ambady, Harizuka, & Kumar, 2004). We created composite facial expressions based on the left and right hemispheres of a face—that is, taking one photograph and turning it into two pictures, one that showed the left side twice and one that showed the right side twice. In-group advantage was greater when participants judged the left hemisphere, which is more intense and mobile, compared with the right hemisphere, which is more prototypical. Expression style was the only plausible explanation, given the within-subjects design for both posers and judges. Even more directly, in a study with Québécois and Gabonese participants (Elfenbein, Beaupré, Lévesque, & Hess, 2007), we documented accents by identifying specific muscle movements—that is, action units (AUs)—that varied across the groups’ posed facial expressions, as seen in Figure 1. Further, greater cultural differences in judgment accuracy were found for the emotions that had greater cultural differences in expression style. This strongly supports dialect theory.

Evidence has also been mounting in work conducted by other labs. This evidence uses facial expressions and other nonverbal channels such as the voice and body language, for which the core tenets of dialect theory are similarly hypothesized. Balanced 2 x 2 designs showed in-group advantage among Americans and Japanese viewing facial expressions (Dailey et al., 2010), Americans and Namibian villagers judging nonlinguistic vocalizations (Sauter, Eisinger, Ekman, & Scott, 2010) and European and Asian Americans viewing full-channel videos of spontaneous emotions (Kang & Lau, 2012). Other studies showed in-group advantage for African students in the United States (US) judging facial expressions and vocal tones (Wickline, Bailey, & Nowicki, 2009), English, German, Arabic, and Spanish speakers judging nonsense syllables from Spain (Pell, Monetta, Paulmann, & Kotz, 2009), speakers of English, German, Chinese, Japanese, and Tagalog judging voices from the US (Thompson & Balkwill, 2006), and Japanese, Sri Lankans, and Americans judging Japanese postures (Kleinsmith, De Silva, & Bianchi-Berthouze, 2006).

Some articles have provided intriguing evidence for the basic propositions of dialect theory, namely that the lower recognition of out-groups’ emotions results from subtle differences in expression style. Kleinsmith et al. (2006) found that perceivers judging static postures in Japan, Sri Lanka, and the US used different cues. Dailey et al. (2010) modeled the conditions that reproduce in-group advantage using a neural network that imitated the receptive fields in the visual cortex that “learn” how to represent objects visually. When trained with sample stimuli that were culturally normative for the US versus Japan, the neural network developed slightly different visual representations.

Finally, some articles have made applied use of insights from dialect theory. Previous findings that there is greater emotional impairment for African American versus Caucasian schizophrenics no longer held when using stimuli from both ethnic groups (Pinkham et al., 2008).

As mentioned earlier, the majority of this work uses posed expressions, leading some critics to speculate that in-group advantage exists only for poses (Matsumoto, Olide, & Willingham, 2009). Two notable exceptions demonstrated in-group advantage for judging spontaneous full-channel videos (Kang & Lau, 2012) and spontaneous anxiety during interracial interactions (Gray, Mendes, & Denny-Brown, 2008). As mentioned earlier, Elfenbein et al. (2004) found greater dialects in the more spontaneous versus posed side of the face. Work in progress with Ursula Hess demonstrates that dialects emerge similarly under cognitive load versus control conditions, with the assumption that cognitive load increases spontaneity by limiting cognitive resources for effortless control.

Critical Accounts

The in-group advantage and resulting dialect theory sparked controversy with its implications for dominant theories about cross-cultural differences in emotion, namely display and decoding rules. Evidence for in-group advantage could not be explained by these factors alone—that is, participants suppressing their displays and perceptions, respectively, for the sake of harmony. Notably, Japanese participants bested Americans when the tasks originated in Japan (Elfenbein & Ambady, 2002b).

Matsumoto (2002) wrote a commentary on this work, in which he asserted a set of methodological requirements that he would require before believing the evidence. Two of these—balanced designs and equally clear signals—were included in the original analysis or controlled for, respectively (Elfenbein & Ambady, 2002a). However, Matsumoto’s last purportedly methodological concern was actually a difference in perspective that gets to the heart of dialect theory. He argued that in-group advantage would disappear if members of each culture expressed their emotions in precisely the same way—in the case of facial expressions, moving precisely the same muscles in the same combinations. This is a matter of violent agreement. In keeping with dialect theory, if there are no differences in the style of emotional expression, then there should be no in-group advantage. If British people spoke in exactly the same manner as Americans, then there would be no linguistic misunderstanding. Accordingly, we referred to Matsumoto’s recommendation as a “cultural eraser” (Elfenbein & Ambady, 2002a, p. 244). It is an oxymoron to require all cross-cultural studies first to pass their stimuli through a filter designed to erase all cross-cultural differences.

Consistent with the notion that forcing stimuli to have exactly the same appearance is a cultural eraser, numerous researchers have replicated the lack of in-group advantage under these conditions (Beaupré & Hess, 2005, 2006; Kang & Lau, 2012; Lee, Chiu, & Chan, 2005; Matsumoto et al., 2009; Tracy & Robins, 2008). Indeed, Elfenbein et al. (2007) replicated this in a separate condition using culturally erased stimuli...
<table>
<thead>
<tr>
<th>Prototype</th>
<th>Québécois</th>
<th>Gabonese</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>AU4+5b+23</td>
<td>AU4+5b+25</td>
</tr>
<tr>
<td>Contempt</td>
<td>AU4+7+23</td>
<td>AU10+17</td>
</tr>
<tr>
<td>Happiness</td>
<td>AU14-unilateral</td>
<td>AU6+12D+26</td>
</tr>
<tr>
<td>Sadness</td>
<td>AU1+4+15b</td>
<td>AU4</td>
</tr>
<tr>
<td>Serenity</td>
<td>None</td>
<td>AU12b</td>
</tr>
<tr>
<td>Shame</td>
<td>AU32+54</td>
<td>AU4+14</td>
</tr>
</tbody>
</table>

Figure 1. Facial expression dialects from Québec and Gabon in Elfenbein, Beaupré, Lévesque, and Hess (2007).
A Theoretical Account of Origins: Taking the Linguistic Metaphor Seriously

Whereas Kurt Lewin argued that “there is nothing so practical as a good theory” (1951, p. 169), in the present case, there has been nothing so theoretical as hard data. The dominant models of cultural differences in emotion have been subjected to relatively minimal testing and, by contrast, dialect theory arose primarily from an attempt to explain within a single framework the diverse body of evidence. It has been important to distinguish theoretical perspectives from each other by way of competing predictions rather than mere argumentation. However, now that the empirical base is filling—even if by no means completely—it is time to develop a deeper theoretical account for the origin and meaning of accents and dialects.

Crafting this theoretical account forces us to wrestle with the question of origin: Why should cultures have accents and dialects in their nonverbal communication of emotion? In attempting to answer this question, we benefit from taking the linguistic metaphor seriously, and drawing from the base of knowledge in linguistics that asks: Why does verbal language have accents and dialects? Scholars argue that spoken language continually evolves, and that it tends to diverge across groups of people who are separated by geographic or social boundaries (O’Grady, Archibald, Aronoff, & Rees-Miller, 2001). With smaller separation these accents are noticeable yet harmless. With increasing separation—that is, social stratification—these dialects create challenges to comprehension that can largely be overcome. With greater separation still, distinct languages emerge that cannot be mutually understood.

As such, the concept of social stratification looms large. However, what are the underlying psychological mechanisms that result from the sociological construct of stratification? It is worth theorizing about two distinct processes likely to act separately and in tandem. First, some changes in verbal—and presumably nonverbal—language occur merely through random drift. Particularly in the absence of formal records, passing down language from one generation to the next has evolution through no deliberate effort. Ordinary change involves constant mutations and even errors, such as “an apkin” becoming “a napkin,” or “a napon” becoming “an apron” (Palmer, 1882). In linguistic drift, social stratification does not create dialects directly, but rather indirectly when these drifts become shared among some speakers but not others (O’Grady et al., 2001). By contrast, in the second psychological mechanism, some changes occur through motivated processes of asserting a distinct social identity. For example, jargon and slang can create a marker or even deliberate barrier that defines group membership.

It is worth noting that the exact form of an accent does not necessarily need a functional goal. For example, there may be no reason why Bostonians drop the retroflex r at the end of a word instead of the dental l. This may or may not be the case for nonverbal accents. Recent work shows that perceivers from Eastern versus Western groups tend to focus more on the eyes than the mouth, perhaps because the eyes provide greater diagnostic cues to hidden meaning (Yuki, Maddux, & Masuda, 2007). Likewise, internal representations show greater consensus in East Asian perceivers for eye-related cues versus Westerners for mouth and eyebrow-related cues (Jack, Blais-Scheepers, Schyns, & Caldara, 2009).

What Does the Linguistic Metaphor Buy Us?

Theorists grappling with culture and emotion benefit from taking the linguistic metaphor seriously for several reasons. Importantly, it allows work to proceed without resolving the question of whether emotions are universal. Clearly the existence of dialects does not deny the possibility of a universal language—indeed, nonverbal dialects should be expected, given that every verbal language with geographic range has regional variations (O’Grady et al., 2001). However, neither does the existence of dialects speak in favor of universality. The basic similarity that one sees in emotional expression style across cultures may result from a biologically based affect program, as posited by Ekman’s (1972) neurocultural theory. However, alternatively it could result from convergent evolution, whereby similar concerns and selection pressures led to similar solutions—for example, the penguin and puffin birds, which came to look alike but are related only distantly. In linguistic terms, most languages use m in the word for mother, presumably because labial sounds are the first consonants babies develop. Likewise, disgust may appear similarly because it adapted from crinkling one’s nose in response to strong smells. Modern evolutionary accounts emphasize individuals’ adaptation and the potential functions of emotional expressions, beyond leftover vestiges that read out our internal states (Owren & Rendall, 2001).

Other benefits of the linguistic metaphor include the ability to borrow theories from the allied social science of linguistics. Further, the linguistic metaphor emphasizes that cultural differences in the expression and perception of emotion are mirror-image processes, so that unlocking the nature of one allows us also to unlock the other. By contrast, the theories of display and decoding rules were developed separately and these rules do not necessarily correspond to each other. Finally, the linguistic metaphor emphasizes the potential for automaticity in cultural differences in emotional expression and recognition. Whereas display rules and decoding rules are a matter of conscious management techniques, as discussed earlier, no such tactics are required by...
dialect theory. After all, Americans do not adjust their speech with each utterance to avoid sounding British.

However, the value of the linguistic metaphor is not to imply that it must become a new set of handcuffs. Indeed, there may be explanations for in-group advantage that do not necessarily follow linguistic principles. For example, an appraisal view of nonverbal dialects has the potential to preserve the notion that people across cultures have a universal mapping from their internal feeling states to their outward displays. The idea here is that emotions exist within broader families—such as irritation, rage, and anger—and cultures may differ in their modal experience within these emotion families. If there is a one-to-one mapping from experiences to the appearance of facial expressions, then this difference in modal experience could lead to dialects that are better recognized by in-group members (Hess & Thibault, 2009). This account appears to have promising preliminary evidence (Hess, Thibault, Lévesque, & Elfenbein, 2008), but a great deal more empirical support would be necessary. Whatever the account for in-group advantage, it needs to be capable of explaining the large body of existing findings which display and decoding rules alone do not.

Another area for further theoretical development is to reconcile the notion of in-group advantage with the three distinct functions for emotional expressions posited by Bühler’s (1934/1990) Organon model (Schêrer, 1988). There has been increasing scrutiny of the idea that faces directly express internal feelings, and even Darwin himself argued that the term “expression” may be preemptive (Parkinson, 2005). First, the “push” function in the Organon model—namely that expressions are symptoms of internal states—has received the most attention, and is at least implicitly the focus of most research reviewed earlier. Second, the “pull” function—namely that expressions are used as signals to produce a reaction in others (see also Fridlund, 1994; Owren & Rendall, 2001)—has received relatively less empirical attention. Third, the “symbolic” function—namely that expressions represent objects or events, similar to linguistic expressions—has been studied the least of all. These different functions are not mutually exclusive and can even reinforce each other over time, as simple reflexes that produce reliable signals can become used deliberately (Russell, Bachorowski, & Fernández-Dols, 2003). It is worth speculating whether in-group advantage is greatest when emotional expressions serve the second and third functions, which correspond more naturally to the linguistic metaphor for interpersonal communication. In this case, various dialects may be better preserved in expressions serving “push” functions, even if the empirical evidence still suggests there is in-group advantage in spontaneous or semi-spontaneous settings.

Bridging the Gap

Despite the somewhat gloomy finding of a cross-cultural barrier in communicating emotions, there are also reassuring data that this barrier can be overcome. Similar to linguistic dialects, in-group advantage is lower across cultural groups enjoying greater physical proximity or cross-group communication (Elfenbein & Ambady, 2002). Also like linguistic dialects, individuals learn from exposure to a new host culture (Elfenbein & Ambady, 2003). Indeed, in-group advantage in recognizing facial expressions appeared to disappear in as little as 10 minutes of practice with feedback (Elfenbein, 2006). As such, dialect theory provides guidance for overcoming cross-cultural challenges. Because in-group advantage results from familiarity with culturally specific elements of nonverbal expression, training and intervention programs can increase familiarity with these elements, thus eliminating in-group advantage. Such training is already starting to take place, for example in work commissioned by the US Army Research Institute for soldiers going overseas. By contrast, if in-group advantage resulted solely from motivation or bias—rather than knowledge and information—the phenomenon would be harder to correct. As such, these findings suggest optimism for our increasingly global and multicultural societies.

References


