Evolution and Human Behavior 35 (2014) 43–50

Contents lists available at ScienceDirect

Evolution and Human Behavior

journal homepage: www.ehbonline.org

Original Article

Evidence that accent is a dimension of social categorization, not a byproduct of perceptual salience, familiarity, or ease-of-processing☆

David Pietraszewski [a,*,1], Alex Schwartz [b]

[a] University of California, Santa Barbara & Center for Evolutionary Psychology, CA, USA
[b] University of California, Santa Barbara, CA, USA

ARTICLE INFO

Article history:
Initial receipt 18 June 2013
Final revision received 26 September 2013

Keywords:
Accent
Social categorization
Language
Cognition

ABSTRACT

Certain features of language, such as accent, are acquired from the local social environment during an acquisition period starting at infancy and ending roughly at puberty. If the scale of social interaction expands after puberty, humans will encounter variance in these linguistic features, and the linguistic similarity between people will indicate the degree to which they share a common early social history. There is reason to believe that such a dynamic is not a modern phenomenon, but has likely occurred throughout our species' evolutionary history since the advent of language. Consequently, the human mind may contain systems for categorizing others according to their linguistic repertoires, including their accents. Four studies demonstrate that (i) accent differences—both native versus non-native accents and also two different non-native accents, are the basis of spontaneous and implicit social categorization, and (ii) this is not driven by general acoustic-differences, low level sound differences, or differences in familiarity or ease-of-processing. Results are consistent with the hypothesis that human mind contains systems for categorizing others according to their accents.

1. Introduction

Language is powerful because it is referential, marking things out in the world in a relatively inexpensive, communicable format. Though the signal-to-referent relationship of a language must be shared to be useful, it is neither fixed nor universal (Pinker, 1994). Instead, different sequences of sounds pick out different features of the world in different times and places.

This apparent paradox – that only a common format affords communication, yet the format is free to vary – represents a solution to an otherwise intractable problem: ensuring coordination of the signal-to-referent relationship in the local environment (such that "car" does not mean [car] to me and [acorn] to you) while maintaining expansiveness and generativity in the signal (Komarova & Nowak, 2001; Kirby & Hurford, 2002). By designing a universal language acquisition device, natural selection is able to store the coordinated social signal in the local social environment, rather than in the germ line (Pinker, 1994; Kirby, 1998).

The solution of storing a shared communication format in the external social environment and not the germ line nevertheless strikes an uneasy balance. Common signal formats exist only so far as people frequently interact (Hill, 1978; Livingstone, 2002). Outside this boundary, the signal format becomes increasingly obsolete and loses its quick and cheap signaling function. For language systems to work well and frequently enough to be worth building, the scale of social interaction must correlate fairly strongly with the scale of signal format similarity (Nettle, 1999). This means that the psychological mechanisms underlying acquisition – the format-fixing devices – must be designed around the spatial and temporal regularities of human social interaction over evolutionary time (Bowern, 2010; Currie & Mace, 2012).

This represents an engineering problem for language acquisition. Aside from the "hard" problem of setting up fixed and open parameters of language in such a way as to make learning tractable (while at the same time affording extensive expression; e.g., Nowak, Krakauer, & Dress, 1999), language acquisition systems face the problem of choosing the right signal format. Given spatial and temporal variation, at what point should the acquisition system be satisfied that it has acquired the signal format of highest expected communicative utility? That is, how long should the acquisition system hold out before foreclosing on a particular language repertoire? If too early, the acquisition system would foreclose at a point in life prior to complete signal fixation or relocations requiring a change in language repertoire (such as leaving a natal language community when reaching adolescence; Bowern, 2010). If retained too late, the acquisition system will be maintained past the point at which the energetic and opportunity costs outweigh the probability of needing to change language repertoire (Hurford, 1991). The point in

☆ We wish to thank Annie E. Wertz for extensive comments on previous drafts of this paper.
* Corresponding author. Department of Psychology, Yale University, Box 208205, New Haven, CT 06520-8205, USA.
E-mail address: david.pietraszewski@yale.edu (D. Pietraszewski).
[1] David Pietraszewski is now at the Department of Psychology, Yale University.

1090-5138/$ – see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.evolhumbehav.2013.09.006

the lifespan at which acquisition closes is not arbitrary; natural selection has run through the culling process of trading off acquisition flexibility and costs, arriving at our modern phenotype in which language acquisition closes around adolescence (see Komarova & Nowak, 2001).

Studies of second language fluency suggest that achieving native-like competence requires exposure to the language by late adolescence, with critical windows for each component (e.g., grammar, syntax, phonemes; Long, 1990; Birdsong, 1999; Scovel, 1988; c.f. Herschensohn, 2007). Thus, language acquisition can occur into late adolescence and is not jettisoned earlier in life (immediately after primary language mastery). This suggests that over evolutionary time individuals changed language communities to some appreciable degree, causing the system to retain its flexibility beyond its initial calibration. This also suggests that even within the scale of evolutionarily-recurrent travel, different language formats were encountered within the first dozen years of life. By late adolescence, the acquisition system is jettisoned, suggesting that maintaining the acquisition machinery beyond puberty was not sufficiently beneficial. This life-stage timing corresponds well to what is known about the timing of residency shifts and exogamy, both in hunter–gatherer populations (e.g., Leacock & Lee, 1982; Kelly, 1995), and also in estimates of ancestral populations (Chapais, 2010).

The way the acquisition system works has a straightforward implication: the more two people share linguistic features, the more likely they spent their early lives in the same social community. In modern circumstances, these communities may include millions and extend across vast geographical distances. But in the context in which the acquisition system evolved – a context of hunting and gathering, traveling on foot, and living in fusion/fission bands – the social and spatial scale of language communities would be greatly diminished and tied to geographic range (Bowern, 2010; Currie & Mace, 2012). At such a scale, language communities would be communities of true face-to-face interaction. In such a world, knowing that two agents shared a community would be inference-rich. They would have many direct and indirect social relationships (either positive or negative), exposure to similar early social histories and similar environments, including pathogens, weather, flora and fauna, and so on (Kelly, 1995; Nettle, 1999; Chapais, 2010).

The diminished scale of ancestral language communities would not preclude exposure to variation in language; variation would be encountered in any context where the scale of social interaction is the same or larger for adults than for children (or more precisely, the same or larger for language-crystallized individuals than those still in their language acquisition years). Unlike exposure to certain phenotypic differences (i.e., 'race'), such circumstances do not appear dependent on modern transportation technology. Hunter–gatherer groups – even those with restricted home ranges – are typically aware of, interact with, and establish relationships with individuals with different linguistic repertoires (Bowern, 2010; Hill, 1978: Lee, 1972); a demographic pattern driven, in part, by the recurrence of exogamy (Chapais, 2010). Thus, over phylogenetic time since the advent of language, humans likely encountered variance in linguistic repertoires.

If a dimension of the social world is available over evolutionary time, and is sufficiently beneficial to track, then natural selection can design systems for tracking that dimension (e.g., Tooby & Cosmides, 1990). Linguistic repertoires are a good candidate: They would be present recurrently over evolutionary time and tracking them would provide a quick, easy, and reliable index of a collection of people, all of whom share a common social and environmental history. Tracking shared socio-environmental histories would afford immediate inferences relevant to valuation and decision-making (e.g., that this person has been exposed to different clines of skills, knowledge, cultural transmission, pathogens, and so on) and guide learning and motivation procedures towards a more accurate and flexible mapping of the local social environment (e.g., learning and expecting that people who talk like X tend be higher status, have a negative stance toward people who talk like Y, and so on, based on cues in the local social environment; Chapais, 2010; Kelly, 1995; McElreath, Boyd, & Richerson, 2003; Nettle, 1999; see also Moya, 2013 and Pietraszewski & Schwartz, this volume). Thus, natural selection may have designed the mind to track the language repertoires of others.

If the human mind is designed to track the acquisition-dependent language repertoires of others, the following should be true of proximate psychological mechanisms: (1) Language repertoires should be spontaneously tracked and some kind of repertoire summary variable should be attached to each agent. This attached variable should be computationally distinct from other vocal parameter representations (e.g., vocal features used to identify the speaker's age, sex, identity, emotional state, and so on), and should also be distinct from the content of what the person says. (2) The relative similarity between agents' language repertoires should be calculated such that agents should be clustered by similarity in some sort of representational format along this dimension. This clustering would in turn enable further inferences and down-stream representations (e.g., up- or down-regulating expectations of shared knowledge). These two processes should lead to behavioral outputs in which, everything else being equal, agents with similar language repertoires are treated as more similar to each other than agents with dissimilar repertoires. In other words, these processes should lead to categorization of agents by their language repertoires.

Further, there should be selectivity such that language repertoires should best satisfy the system's input conditions. Acoustic differences or similarities writ large should not be able to account for steps (1) and (2) above, and the mind should not treat all acoustic differences in the same way it treats linguistic repertoire differences. That is, there should be evidence of special design for tracking linguistic repertoire differences (Williams, 1966).

Previous research demonstrates that (i) the mind contains specialized and early-developing mental machinery that differentiates language from non-language and one language from another (DeCasper & Fifer, 1980; Mehler et al., 1988; Nazzi, Bertoncini, & Mehler, 1998; Kuhl, Tsao, & Liu, 2003; Sebastián-Gallés & Bosch, 2005), (ii) young children prefer native-language to non-native repertoires (Kinzler, Dupoux, & Spelke, 2007; Shutts, Kinzler, McKee, & Spelke, 2009; c.f. Cohen & Haun, 2013), (iii) discrimination between linguistic variants is subtle, sophisticated, and experience-dependent (Long & Preston, 2002; Cristia et al., 2012) and (iv) connections are forged between language repertoires and social inferences over the course of development (Giles & Powesland, 1975; Brown & Levinson, 1979; Hirschfeld & Gelman, 1997; Major, 2001; Robinson & Giles, 2001; Giles, Reid, & Harwood, 2010; Gluszek & Dovidio, 2010; Moya, 2013). However, research has not yet directly examined if language differences are a basis of social categorization (i.e., if the mind spontaneously categorizes people according to their language repertoires via the proximate processes described above, though see Online Supplementary Materials for suggestive evidence from Rakić, Steffans, and Mummendey (2011)). This is explored in the current paper. A series of four studies test if acquisition-dependent language differences are a spontaneous basis of social categorization. We furthermore test for evidence of special design by testing against multiple alternative hypothesis, including that categorization by language differences reflects nothing more than attention to sound differences, low-level acoustic features, or to stimuli that differ in familiarity or ease-of-processing.

2. Current studies

Each of the acquisition-dependent features of language – grammar, syntax, phonemes – have slightly different acquisition windows (Long, 1990; Birdsong, 1999; Scovel, 1988; c.f. Herschensohn, 2007).

In the current studies, we examined only one feature: accent (or more technically, global accent, which includes prosody as well as the repertoire of sound units and their combinations; Major, 2001). While the general hypothesis predicts that each acquisition-fixed feature should be tracked by the mind, the implications of the current studies can be conservatively restricted to accent. To test whether accents are a basis of social categorization, the current studies use a standard protocol from social psychology for measuring implicit and spontaneous social categorization: the memory confusion paradigm (Taylor, Fiske, Etcoff, & Ruderman, 1978).

2.1. The memory confusion paradigm

The memory confusion paradigm is a standard way of measuring social categorization implicitly, thus removing social desirability effects, and has been used extensively for over three decades (e.g., Taylor et al., 1978; Kurzban, Tooby, & Cosmides, 2001; Susskind, 2007). The logic of the paradigm is that if a particular feature – such as accent – is a basis of categorization, then people who are similar along that dimension should be more readily confused with one another in memory. That is, when trying to recall information, memories about people who share the same accent will be more likely to be confused with one another, even in the absence of conscious awareness that this is happening. Thus, patterns of memory confusions reveal fundamental categorization processes.

There are three phases of the paradigm (see Fig. 1). First is the initial presentation phase, in which participants are exposed to pairings of faces and statements. A sequence of faces is shown one at a time, and each face is accompanied by a statement (either in text or played aurally). Participants are simply told to form impressions of each person as they make each statement. Second is a distractor task, which is designed to prevent participants from explicitly rehearsing or thinking about the speakers and statements they had just seen. Third is a final surprise recall phase. Participants are shown an array of all of the faces they had seen previously, and are asked to try to remember which statement came from which speaker (i.e., who said what?). Statements appear in random order, and participants indicate their best guess as to who said that statement. This task is very difficult, error rates are very high, and participants experience their choices largely as random guessing (Taylor et al., 1978).

Unbeknownst to participants, these errors reveal non-conscious categorization processes. For instance, if a participant had categorized speakers by their accents during the initial presentation phase, then during the recall phase – even if they forget which specific speaker said a particular statement – they will be more likely to misattribute the statement to someone else who also had same accent as the original speaker, as opposed to someone who had a different accent. If however they did not categorize the speakers by their accents during the initial presentation phase, then their attribution errors will be random with respect to same-accent or different-accent individuals during this recall phase. Thus, the occurrence of more within-accent than between-accent attribution errors reveals that accent categorization occurred, and the magnitude of the difference between these two types of errors quantifies the strength of this categorization.

Studies suggest that the categorization processes measured by the memory confusion paradigm are largely independent of explicit intuitions and beliefs, and immune in some cases to explicit instruction to attend or not attend to certain dimensions (e.g., Stangor, Lynch, Duan, & Glass, 1992). Furthermore, categorization does not occur along any and all noticeable dimensions. Shared shirt color, for example, while large and visible, is not a substantial dimension of categorization unless given additional meaning (Stangor et al., 1992; Brewer, Weber, & Carini, 1995; Weeks & Lupfer, 2004; Sack, 2005; Pietraszewski, 2009). Thus, this method does not simply capture the ability to discriminate between superficial differences, but rather more extensive encoding, representation, and retrieval processes (e.g., Cosmides, Tooby, & Kurzban, 2003; Stangor et al., 1992). This highlights the important theoretical point that social categorization is a distinctly different process than simply noticing the differences between people.

Finally, because only attribution errors are used to calculate categorization, this paradigm has the added virtue of allowing for the realistic presentation of multiple different speakers while at the same time dissociating the affects of speaker idiosyncrasies from the dependent measure. To the degree that participants notice and remember a speaker's idiosyncratic vocal features, they will make a correct attribution, and this will not factor into the categorization dependent measure. Moreover, even if participants do notice and remember vocal idiosyncrasies, but not well enough to remember the exact speaker, this information cannot help them in any systematic manner with respect to accent when choosing among the incorrect speakers (the exception would be if speaker idiosyncrasies and accent were inadvertently confounded in the stimuli, which will be tested by reversing the accent stimuli in Study 3). If the only thing that different speakers share is their accent, then the error of attributing one such speaker for another can only be the result of having categorized the speakers by accent. Thus, this memory confusion paradigm is particularly well-suited for exploring accent categorization.

2.2. Overview of the four studies

A series of four studies (i) test if acquisition-dependent language differences are a spontaneous basis of social categorization and (ii) test against multiple alternative hypotheses for the design of the underlying proximate psychology, including that categorization by

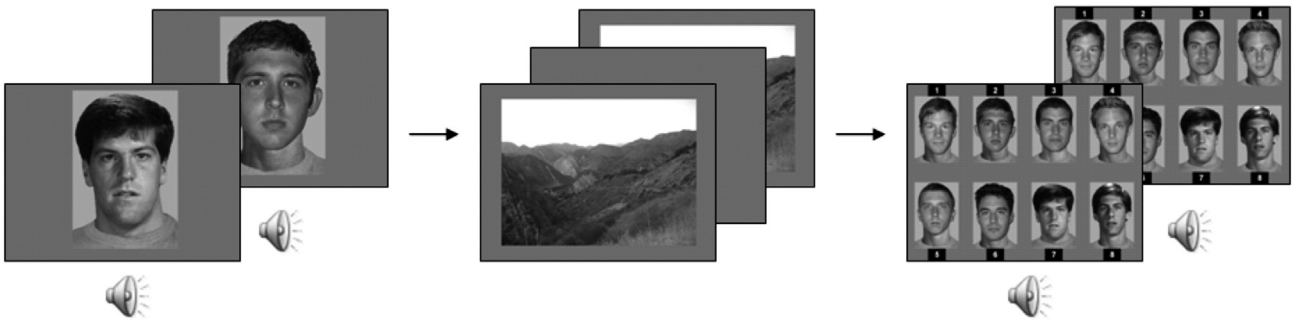

Fig. 1. Schematic of the memory confusion paradigm (from left to right). 1) Initial presentation phase: acoustic stimuli are presented concurrently with faces; these pairings are presented in sequential order. 2) One minute distracter task: in the current studies a change-detection paradigm; to prevent recency and rehearsal effects). 3) Surprise recall phase: all eight targets are presented at once, statements are presented sequentially in a randomized order, participants select who said each statement.

language differences reflects nothing more than attention to sound differences, low-level acoustic features, or to stimuli that differ in familiarity or ease-of-processing. In Study 1 we examined if participants would categorize targets according to whether they had a native versus a non-native accent. In Study 2 we tested against the first alternative hypothesis: that accent categorization reflects a more general process of categorizing speakers by any obvious and noticeable sound difference. Study 3 tested against a second alternative hypothesis: that accent categorization was driven by inadvertent confounds in our stimuli between accent categories and speaker vocal idiosyncrasies (such as pitch, loudness, speaking duration, etc.), and that categorization inadvertently reflected the encoding of these and other low-level non-linguistic sound differences. In Study 4 we examined if participants would categorize by targets by two different non-native accents. This allowed us to (i) test if the underlying proximate psychology makes more precise distinctions than just "ingroup" versus "outgroup", and (ii) test against two final alternate hypotheses: that accent categorization is a byproduct of categorization by vocal features that differ in familiarity or in ease-of-processing. If the mind is designed to track the accents of others, then categorization by accent should occur in the context of a native versus non-native contrast, and also when both accents are non-native. Additionally, none of the alternative hypotheses should be able to account for these categorization effects.

3. Study 1: Native versus non-native accent

Study 1 featured speakers with accents that were, from our participants' perspectives, either native or non-native. If accent is a basis of social categorization, speakers who have a native accent should be categorized distinctly from those who have a non-native accent.

3.1. Method

Speakers representing two different accent groups—native English language speakers from the United States (California) and native English language speakers from England (southern counties) were recorded while reading a list of twenty-four statements (hereafter we will refer to these accent groups as US English and British English). Statements were then normalized for volume and stored uncompressed in a digital format. From these, four male US English voices and four male British English voices were randomly assigned to eight photos of Caucasian males with neutral expressions (a subset of which consists of modified images from the NimStim face stimuli set; Tottenham et al., 2009). Which individual voice was assigned to which individual face varied in two different between-subjects versions.

3.2. Participants

Thirty students from the University of California, Santa Barbara, USA (13 male, 17 female; mean age $\pm$ SD $=$ 19.43 $\pm$ 1.72 years) participated for course credit.

3.3. Procedure

Participants were told they would be seeing samples of statements made by a random sampling of people who had participated in surveys, and to form impressions of these individuals as they make their statements. The statements – the content of which have been validated as neutral stimuli for establishing baseline levels of categorization by race and sex in previous memory confusion studies (Pietraszewski, 2009) – were then played aurally along with a photo of the "speaker." Statements were designed to be irrelevant to any social category and not form any kind of narrative or conversation (see Online Supplementary Materials). Eight speakers each made three statements, for a total of twenty-four statements. Each speaker's photo was presented for the entire duration of the statement, plus two additional seconds thereafter. The first two statements always came from native British English voices and the third and fourth from native US English voices. Accent order was unsystematic thereafter, within the constraint that each speaker spoke once during statements 1–8, once again in statements 9–16, and once in statements 16–24. No relationship existed between accent category and the content of the statements or features in the photos (i.e., in the parlance of the social categorization literature, there was no structural fit in the stimuli; Blanz, 1999; Oakes, 1987; Oakes, Turner, & Halsam, 1991).

After the standard one minute distracter task to prevent recency and rehearsal effects (in this case, we used a change-detection paradigm featuring four outdoor and indoor scenes without people) participants were asked to determine which speaker made each statement. An array of all eight speakers appeared on screen, and the previous twenty-four statements were then presented aurally in a random order. Each of the speakers in the array was assigned a number (1–8). Participants indicated their response by pressing the number of the speaker that they believed said the statement (see Fig. 1).

Categorization was measured by calculating the difference in error rates between same-accent errors and different-accent errors. A standard correction for the different base rate probabilities for the two error types was applied (i.e., because there is one correct speaker, there is one less same-accent error than different-accent error possible, thus different-accent errors were multiplied by .75). If participants categorize targets according to their accents, then same-accent errors will exceed different-accent errors.

3.4. Results

Participants made a substantial number of errors (out of 24 responses, there were 13.57 errors on average (SD $=$ 4.52), more than half of the responses). Critically, of these there were substantially more same accent errors (M $\pm$ SD $=$ 8.77 $\pm$ 2.71) than different accent errors (M $\pm$ SD $=$ 3.60 $\pm$ 3.00), $t_{29} = 6.62$, $p < .001$, $r = .78$ (see 'Study 1' in Fig. 2), indicating that participants were strongly categorizing targets by accent.

3.5. Discussion

These results demonstrate that accent – in this case, a native versus non-native accent distinction – was indeed the basis of strong social categorization. When participants incorrectly attributed statements to speakers, they were much more likely to choose a speaker who shared the same accent as the original speaker. There was nothing in the photos to indicate the speakers' native versus non-native accent status (they were all Caucasian men in grey t-shirts), therefore for categorization to occur, participants must have encoded the speakers' native versus non-native accent information during the initial presentation phase and retained this category information during the recall task.

4. Studies 2–4: Testing against alternative hypotheses

Although participants categorized targets by native versus non-native accents in Study 1, this does not warrant the inference that this represents categorization by accent per se. This is because a native versus non-native accent difference is also confounded with other stimulus differences, including obvious sound differences, non-obvious low-level sound differences, differences in familiarity, and differences in ease-of-processing. These other stimulus differences may have been driving categorization, not accent. Studies 2–4 test against these alternative hypotheses.

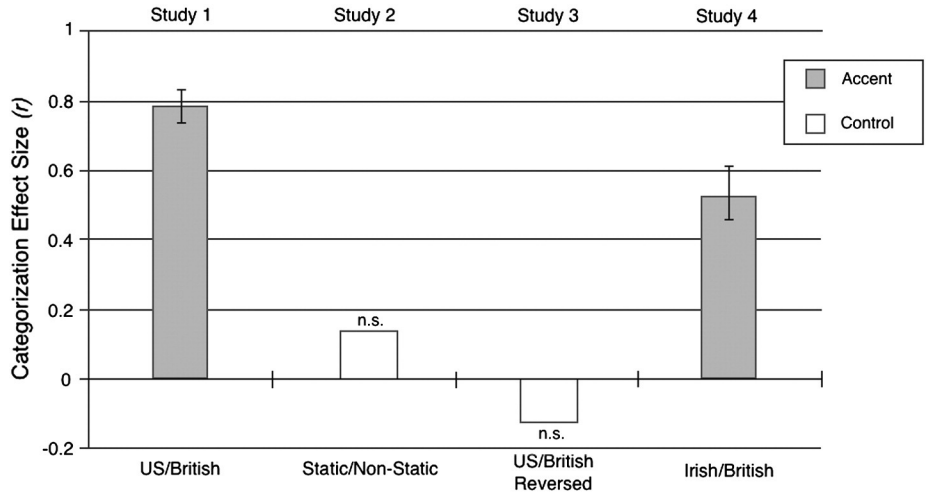

Fig. 2. Categorization effect sizes for each study (i.e., the degree to which participants made more within-category attribution errors than between-category attribution errors). The X axis indicates the stimulus distinction used in each study. "US", "British", and "Irish" denote accent groups. Effect sizes are significantly different from zero, unless labeled n.s. ($p > .05$). (Note: The negative effect size for the US/British reversed stimuli in Study 3 indicates that between-category errors exceeded within-category errors—the "opposite" of categorization). Error bars: ± 1 S.E.

5. Study 2: Alternative hypothesis: general categorization by obvious sound differences

Accent differences are also sound differences. Perhaps there is something about the acoustic modality such than any obvious and readily-perceivable acoustic difference becomes the basis of categorization in the memory confusion paradigm. Previous research establishes that obvious and easily discriminable visible features do not usually elicit categorization, unless they have some relevance or meaning (e.g., Stangor et al., 1992; Brewer et al., 1995). Yet no such acoustic control has ever been conducted with this paradigm. Showing that a non-meaningful but readily-perceivable acoustic difference does not elicit categorization is important because it excludes the possibility that any readily-perceivable acoustic difference will also become the basis of categorization. This was accomplished by presenting a non-linguistic but easily perceivable acoustic difference —the presence or absence of static in speakers' statements. If speakers are not categorized along this dimension, then this will demonstrate that categorization by accent in Study 1 was not driven simply by the perception of superficial acoustic differences between speakers.

5.1. Method

Study 2 was identical to Study 1, with the exception of the acoustic stimuli. Rather than speakers with two different accents, Study 2 involved speakers who all shared the same native accent as the participants (i.e., they had no perceived accent). These were eight native US English speakers: the four used in Study 1, plus four others. Using Adobe Audition software (Adobe Systems, Inc.) static was applied to four of the speakers' statements as heavily as possible, with the constraint that the speakers' statements could still be heard and understood.

5.2. Participants

Fifty-two students from the University of California, Santa Barbara, USA (25 male, 27 female; mean age $\pm$ SD $=$ 19.06 $\pm$ 1.27 years) participated for experimental credit.

5.3. Results

Participants made 13.27 attribution errors on average (SD $=$ 4.42), which are nearly identical to the accuracy rate found in Study 1, $t_{80} = .29$, $p = .772$, suggesting that the static stimuli were comparable to the accent stimuli in terms of attribution difficulty. However, participants did not make more attributions to targets who were similar along the static present/absent dimension (M $\pm$ SD $=$ 5.98 $\pm$ 2.47) than attributions to targets who were dissimilar along the static present/absent dimension (M $\pm$ SD $=$ 5.47 $\pm$ 2.70), $t_{51} =$ 1.03, $p = .309$, $r = .14$ (see 'Study 2' in Figure 2), indicating that participants did not categorize the targets according to whether they spoke with or without static.

5.4. Discussion

Failure to find significant categorization by static demonstrates that categorization by accent in Study 1 could not have been driven by a more general categorization process that picks up any acoustic difference between stimuli. Therefore, the alternative hypothesis of general categorization by obvious sound differences can be eliminated. However, this result still leaves open the possibility that participants may have instead picked up on non-obvious and subtle low-level acoustic differences between speakers in Study 1, and that this was driving categorization by accent. We test against this alternate hypothesis in Study 3.

6. Study 3: Alternative hypothesis: categorization by low-level sound differences

Study 3 examined whether there were vocal features inadvertently confounded with accent that might have driven categorization in Study 1. There is substantial evidence that subtle vocal features can have profound impacts on person perception. For example, particular vocal features are associated with age (Helfrich, 1979), sex (Smith, 1979), size, height, and strength in men (Sell et al., 2010), and many temporary states such as ovulation (Bryant & Haselton, 2009), health, fatigue, sexual arousal, pregnancy, menstruation, and many others (see Laver & Trudgill, 1979 for a review). Because different speakers were used for each accent category in Study 1, it is possible that inadvertent and systematic vocal differences occurred between the two sets of differently-accented speakers. For example, perhaps the speakers from England had lower fundamental frequencies on average, or spoke at a systematically different rate.

To ensure that the categorization found in Study 1 was not driven by detection of any of these lower-level features, and was instead driven by accent differences, the stimuli from Study 1 were played in reverse to a new set of participants in Study 3. Playing in reverse

preserves all of the low-level sounds within the stimuli, while destroying the linguistically-derived information. If inadvertently confounded low level acoustic features were driving categorization in Study 1, then categorization would still occur even when the US English and British English speakers' statements are played in reverse.

6.1. Method

Study 3 was identical to Study 1, except that the stimuli were played in reverse. Participants were also told that the stimuli that they would be hearing might be degraded. This was done to ensure that participants would not be confused or surprised that the statements could not be linguistically understood, avoiding the perception that this reflected some kind of experimental error.

6.2. Participants

Sixty-six students from the University of California, Santa Barbara, USA (10 males, 50 females, 6 unknown; mean age ± SD = 20.09 ± 1.30) participated for course credit.

6.3. Results

Participants made 18.70 error attributions on average (SD = 2.57), substantially more than the number in the previous two studies, $t_{146} = 8.65$, $p < .001$, $r = .58$, suggesting that participants found the task of attributing reversed stimuli more difficult. Participants did not make more same accent errors (M ± SD = 7.70 ± 2.55) than different accent errors (M ± SD = 8.25 ± 2.15), $t_{55} = -1.08$, $p = .283$, $r = -.13$ (see 'Study 3' in Figure 2), indicating that there was no categorization of these stimuli. This means that there were no inadvertently confounded low-level sound differences causing categorization in Study 1.

6.4. Discussion

There was no categorization of the speakers when the accent stimuli were presented in reverse. This demonstrates that the categorization by accent found in Study 1 was not driven by categorization of inadvertently confounded low-level vocal features within the stimuli. If these features had been driving categorization to any significant degree at all in Study 1, then they would have been revealed through categorization in Study 3. There was none. Therefore, studies 2 and 3 exclude alternative hypotheses related to obvious and non-obvious sound differences. However, aside from sound differences, there are also familiarity and ease-of-processing differences between native versus non-native accents. These are not yet excluded. Study 4 tests against these two final alternate hypotheses.

7. Study 4: Two non-native accents

The goals of study 4 were twofold. First was to establish a second accent-related phenomenon: examining if two different non-native accents would also be the basis of social categorization, just as native versus non-native accents were categorized in Study 1. Second was to test against two final alternate hypotheses: that categorization by accent is driven by either familiarity or ease-of-processing differences.

From our participants' perspective, targets in Study 1 spoke with either a native or a non-native accent. Thus, accent differences were also confounded with familiarity differences. Therefore Study 1 may have reflected a more general process of categorizing the world into the familiar and unfamiliar, rather than by accent differences. Similarly, native versus non-native accent differences also differ in how easy they are to process, with foreign-accented speech being more difficult (Cristia et al., 2012). If participants had more difficultly in understanding one set of speakers more than the other, then this might lead to categorization by listening effort, rather than categorization by accent. If only a native versus non-native accent contrast is used, there is no way to tease apart the effects of accent differences from that of differences in familiarity and ease-of-processing.

Study 4 therefore provides the critical test between these possibilities, by presenting two different non-native accents (British and Irish) to participants, and examining whether categorization still occurs. If robust categorization is still found, then this will demonstrate that differences in familiarity and ease-of-processing cannot account for accent categorization.

7.1. Method

Study 4 was identical to Study 1, except that the four native US English voices were replaced with four native English language speakers from Ireland (hereafter "Irish") voices (recorded in the same manner as Study 1).

7.2. Participants

Fifty-three students from the University of California, Santa Barbara, USA (12 male, 41 female; mean age ± SD = 19.43 ± 2.28) participated for course credit.

7.3. Results

Participants made 15.99 attribution errors on average (SD = 3.54), more than in Studies 1 and 2, $t_{133} = 3.60$, $p < .001$, $r = .30$, suggesting that the attribution task featuring all non-native accents was more difficult than the tasks featuring half non-native accents or featuring half static-present statements. Participants made more same accent errors (M ± SD = 7.98 ± 1.97) than different accent errors (M ± SD = 6.00 ± 2.39), $t_{52} = 4.43$, $p < .001$, $r = .52$ (see 'Study 4' in Figure 2), indicating that participants did indeed categorize speakers by the two non-native accents. This error difference was less than that found in Study 1 ($t_{48.3} = 3.54$, $p = .001$, $r = .37$), indicating that the level of categorization by two different non-native accents was somewhat lower than the level of categorization by a native versus non-native accent.

7.4. Discussion

Study 4 demonstrates that differences in familiarity or ease-of-processing alone cannot be driving categorization by accent; when presented with two non-native accents – which were equally unfamiliar and equally difficult to understand (cf. to US English) – participants still categorized speakers by accent at high levels. Thus, categorization by accent cannot be a consequence of a more general categorization process of grouping people according to how familiar versus unfamiliar their features are, or by how difficult or easy it is to process and understand what they are saying. However, because categorization levels were lower when both accents were non-native (compared to when one was native), these factors do seem to have a small additional effect on overall levels of categorization.

8. General discussion

The current studies explored whether accents are a basis of social categorization using a standard measure of spontaneous and implicit social categorization. Based on considerations of likely phylogenetic invariances – that linguistic variability was a recurrent and informative feature of ancestral environments – we hypothesized that natural selection may have designed the mind to track the language repertoires of others, and in particular, their accents.

In fact, categorization by accent occurred readily and spontaneously. When native versus non-native accented speakers were

presented, categorization was significant and robust (Study 1). A general tendency to categorize targets along any perceivable sound difference cannot explain this result because when statements featured the presence or absence of static, no categorization occurred (Study 2). This failure to categorize targets by static adds to an already sizable set of findings demonstrating that perceptual salience does not explain categorization in the memory confusion paradigm (e.g., Brewer et al., 1995; Sack, 2005: Stangor et al., 1992). This result also contradicts the intuitively-appealing idea that people are initially categorized by their accents simply because they sound different from one another.

Prior to making conclusions about the significance of accent it was further necessary to verify there were no inadvertent acoustic confounds lurking within the stimuli. Whenever presenting multiple voices to participants, one concern is that vocal idiosyncrasies or other subtle acoustic features may be confounded with the dimension of interest. It was therefore critical to demonstrate that vocal features other than accent were not driving these effects. This was accomplished in Study 3 by presenting the same acoustic stimuli from Study 1, but played in reverse. This preserved the sounds within the stimuli, but destroyed the ability to linguistically-process them. Categorization of targets did not occur in this case, demonstrating that categorization by accent in Study 1 was not driven by any acoustic confounds lurking within the stimuli.

Study 4 demonstrated that categorization by two different non-native accents also occurs readily and spontaneously. This was crucial to demonstrate because it allowed us tease apart the effects of accent differences from that of familiarity and ease-of-processing differences. Categorization by two different non-native accents was substantial (although somewhat reduced from levels seen in Study 1), occurring at the magnitude typically found for categorization by race (e.g., Taylor et al., 1978; Hewstone, Hantzi, & Johnston, 1991; Stangor et al., 1992). This robust categorization by two non-native accents demonstrates that categorization by accent is not simply driven by differences in familiarity, nor by differences in ease-of-processing. This result also demonstrates that the design of the language repertoire tracking system retains distinctions within the "foreign accent" category, rather than lumping all non-native accents into a single "outgroup" category. This design feature is necessary for extending language-community-specific inferences (e.g. a person speaking repertoire A attacks you, leading you to up-regulate your expectation that other speakers of A may do the same) and suggests that the function of tracking language includes collecting repertoire-specific information and generating repertoire-specific inferences and expectations, rather than simply categorizing the world into ingroups versus outgroups (see Pietraszewski & Schwartz, this volume).

In sum, the results of each study were consistent with the hypothesis that accent is a dimension of social categorization. Categorization occurred by both native versus non-native accent and also by two different non-native accents. None of the alternative hypotheses were able to account for any of these categorization effects. To our knowledge, these are the first studies to use both non-social and low-level acoustic controls in a study of language perception, categorization, or evaluation. These are also the first studies to dissociate accent differences from familiarity and ease-of-processing differences (c.f. Hirschfeld & Gelman, 1997; Kinzler et al., 2007; Rakić et al., 2011; Shutts et al., 2009). Thus, although previous studies have shown that native versus non-native language contrasts impact dependent measures (Kinzler et al., 2007; Rakić et al., 2011; Shutts et al., 2009), the present studies provide the strongest evidence to date of the selective importance of accent per se.

The effect sizes for categorization by accent were large and robust, similar to the effect sizes for sex and race categorization found in previous studies (Hewstone et al., 1991; Sani, Bennett, & Soutar, 2005; Stangor et al., 1992; Taylor et al., 1978; Susskind, 2007; Taylor & Falcone, 1982; Van Twyver & Van Knippenberg, 1998; cf. Kurzban et al., 2001). Transiently-activated social categories (i.e., those activated based on additional external cues or transient goals) typically are not of this magnitude. This suggests accent differences are robustly encoded and stored, and may be a default dimension of person perception and representation, underwritten by dedicated cognitive machinery, similar to sex and age (Messick & Mackie, 1989; Hamilton, Stroessner, & Driscoll, 1994; Lieberman, Oum, & Kurzban, 2008)—a possibility that will be directly addressed in the paper that follows.

Although the current studies establish accent categorization, much remains to be investigated. First, we are treating 'accent' as a unitary construct. Linguists distinguish between four classes of differences between accents: 1) 'Systemic' features: the overall number of word-differentiating sound distinctions within each repertoire (phonemes), 2) 'Structural' features: the different rules about how phonemes can be joined together (phonotaxis), 3) 'Selectional' features: the mapping of particular sounds to particular words, and 4) 'Realizational' features: the particular ways in which phonemes are pronounced (Laver & Trudgill, 1979). Each feature in the progression represents a more subtle difference between accents. The first two classes can be thought of as describing the phonological resources of an accent; the latter two, the way in which those phonological resources are deployed (see also Kerswill & Shockey, 2007). The accents used in the current studies differ along the broadest category—systemically different (or vary in global accent; Major, 2001). It is an open empirical question whether similar categorization results would hold for more subtle accent differences. The same question also extends to other types of language variants that are not accents. One is diglossia, the 'language fashion' that occurs within a single language community for the purposes of social demarcation (Ferguson, 1959). Some of these other variants, including diglossias, signal volitional alliances, affiliation, and patterns of interaction that can shift and occur over much shorter time. These are therefore likely represented differently by the mind than acquisition-dependent language features such as accent.

Second, participants were only exposed to known accents differences, in that they likely had some exposure to each accent prior to the study (i.e., British and Irish accents are frequently heard either in person or in movies and television). It is therefore an open question if novel accents, or accents within an unknown language, are categorized or even perceived. The field of perceptual dialectology suggests that the perception of linguistic variation is to some degree experience-dependent, such that differences between local, frequently-encountered linguistic repertories are richly differentiated, whereas differences among more distant and less frequently-encountered repertories are not perceived at all (e.g., Long & Preston, 2002). Future studies will therefore be necessary to establish both how sensitively accent boundaries are perceived and whether social categorization occurs with more subtle or unfamiliar language repertoire differences (e.g., Cristia et al., 2012), and can also tease apart to what degree the effects of familiarity are independent from differences in ease-of-processing.

Finally, in the paper that follows, we examine a final alternate hypothesis to the proposal that mind may be designed to categorize by accent: if accent categorization is instead a result of coalitional psychology operating over the course of ontogeny, picking up on and categorizing people by any features that happen to predict their coalitional affiliation patterns (as has been recently been found to be true of categorization by race; Kurzban et al., 2001; Cosmides et al., 2003). That is, we explore whether accent categorization is a byproduct of coalitional categorization (like race), or is instead a dedicated dimension of social categorization.

Supplementary Materials

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.evolhumbehav.2013.09.006.

References

Birdsong, D. (Ed.). (1999). *Second language acquisition and the critical period hypothesis*. New Jersey: Lawrence Erlbaum.

Blanz, M. (1999). Accessibility and fit as determinants of the salience of social categorizations. *European Journal of Social Psychology*, *29*, 43–74.

Bowern, C. (2010). Correlates of language change in hunter–gatherer and other 'small' languages. *Language and Linguistics*, *4*, 665–679.

Brewer, M. B., Weber, J. G., & Carini, B. (1995). Person memory in intergroup contexts: Categorization versus individuation. *Journal of Personality and Social Psychology*, *69*, 29–40.

Brown, P., & Levinson, S. (1979). Social structure, groups, and interaction. In K. R. Scherer, & H. Giles (Eds.), *Social markers in speech* (pp. 109–146). Cambridge: Cambridge University Press.

Bryant, G., & Haselton, M. G. (2009). Vocal cues of ovulation in human females. *Biology Letters*, *5*, 12–15.

Chapais, B. (2010). The deep structure of human society: Primate origins and evolution. In P. M. Kappeler, & J. B. Silk (Eds.), *Mind the gap: Tracing the origins of human universals* (pp. 19–51). New York: Springer.

Cohen, E., & Haun, D. (2013). The development of tag-based cooperation via a socially acquired trait. *Evolution and Human Behavior*, *34*, 230–235.

Cosmides, L., Tooby, J., & Kurzban, R. (2003). Perceptions of race. *Trends in Cognitive Sciences*, *7*, 517–521.

Cristia, A., Seidl, A., Vaughn, C., Schmale, R., Bradlow, A., & Floccia, A. (2012). Linguistic processing of accented speech across the lifespan. *Frontiers in Psychology*, *3*, 1–15.

Currie, T. E., & Mace, R. (2012). The evolution of ethnolinguistic diversity. *Advances in Complex Systems*, *15*, 1–20.

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: newborns prefer their mothers' voices. *Science*, *208*, 1174–1176.

Ferguson, C. A. (1959). 'Diglossia'. *Word*, *15*, 325–340.

Giles, H., & Powesland, P. F. (1975). *Speech style and social evaluation*. New York: Academic Press.

Giles, H., Reid, S., & Harwood, J. (Eds.). (2010). *The dynamics of intergroup communication*. New York: Peter Lang.

Gluszek, A., & Dovidio, J. F. (2010). The way they speak: A social psychological perspective on the stigma of nonnative accents in communication. *Personality and Social Psychology Review*, *14*, 214–237.

Hamilton, D. L., Stroessner, S. J., & Driscoll, D. M. (1994). Social cognition and the study of stereotyping. In P. G. Devine, D. L. Hamilton, & T. M. Ostrom (Eds.), *Social cognition: Impact on social psychology* (pp. 291–346). San Diego: Academic Press.

Helfrich, H. (1979). Age markers in speech. In K. R. Scherer, & H. Giles (Eds.), *Social markers in speech* (pp. 63–107). Cambridge: Cambridge University Press.

Herschensohn, J. (2007). *Language development and age*. Cambridge: Cambridge University Press.

Hewstone, M., Hantzi, A., & Johnston, L. (1991). Social categorization and person memory: The pervasiveness of race as an organizing principle. *European Journal of Social Psychology*, *21*, 517–528.

Hill, J. H. (1978). Language contact systems and human adaptations. *Journal of Anthropological Research*, *34*, 1–26.

Hirschfeld, L. A., & Gelman, S. A. (1997). What young children think about the relationship between language variation and social difference. *Cognitive Development*, *12*, 213–238.

Hurford, J. R. (1991). The evolution of the critical period for language acquisition. *Cognition*, *40*, 159–201.

Kelly, R. L. (1995). *The foraging spectrum: Diversity in hunter–gatherer lifeways*. Washington, DC: Smithsonian Institution Press.

Kerswill, P., & Shockey, L. (2007). The description and acquisition of variable phonological patterns: Phonology and sociolinguistics. In M. C. Pennington (Ed.), *Phonology in context* (pp. 51–75). New York: Palgrave Macmillan.

Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences*, *104*, 12577–12580.

Kirby, S. (1998). Fitness and the selective adaptation of language. In J. R. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.), *Approaches to the evolution of language: Social and cognitive bases* (pp. 1–19). New York: Cambridge University Press.

Kirby, S., & Hurford, J. R. (2002). The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi, & D. Parisi (Eds.), *Simulating the evolution of language* (pp. 121–147). London: Springer.

Komarova, N. L., & Nowak, M. A. (2001). Natural selection of the critical period for language acquisition. *Proceedings of the Royal Society London (Biological Sciences)*, *268*, 1189–1196.

Kuhl, P. K., Tsao, F. -M., & Liu, H. -M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences*, *100*, 9096–9101.

Kurzban, R., Tooby, J., & Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. *Proceedings of the National Academy of Sciences*, *98*, 15387–15392.

Laver, J., & Trudgill, P. (1979). Phonetic and linguistic markers in speech. In K. R. Scherer, & H. Giles (Eds.), *Social markers in speech* (pp. 1–32). Cambridge: Cambridge University Press.

Leacock, E., & Lee, R. (Eds.). (1982). *Politics and history in band societies*. Cambridge: Cambridge University Press.

Lee, R. B. (1972). !Kung spatial organization: An ecological and historical perspective. *Human Ecology*, *1*, 125–147.

Lieberman, D., Oum, R. E., & Kurzban, R. (2008). The family of fundamental social categories includes kinship: Evidence from the memory confusion paradigm. *European Journal of Social Psychology*, *38*, 998–1012.

Livingstone, D. (2002). The evolution of dialect diversity. In A. Cangelosi, & D. Parisi (Eds.), *Simulating the evolution of language* (pp. 99–117). London: Springer.

Long, M. H. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition*, *12*, 251–285.

Long, D., & Preston, D. R. (2002). *Handbook of perceptual dialectology*. , *2*, . Philadelphia: John Benjamins North America.

Major, R. C. (2001). Foreign accent: The ontogeny and phylogeny of second language phonology. New Jersey: Lawrence Erlbaum Associates.

McElreath, R., Boyd, R., & Richerson, P. J. (2003). Shared norms and the evolution of ethnic markers. *Current Anthropology*, *44*, 122–129.

Mehler, J., Jusczyk, P., Lambertz, G., Halstead, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*, 143–178.

Messick, D. M., & Mackie, D. M. (1989). Intergroup relations. *Annual Review of Psychology*, *40*, 45–81.

Moya, C. (2013). Evolved priors for ethnolinguistic categorization: A case study from the Quechua-Aymara boundary in the Peruvian Altiplano. *Evolution and Human Behavior*, *34*, 265–272.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Towards an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 756–766.

Nettle, D. (1999). *Linguistic diversity*. New York: Oxford University Press.

Nowak, M. A., Krakauer, D. C., & Dress, A. (1999). An error limit for the evolution of language. *Proceedings of the Royal Society London (Biological Sciences)*, *266*, 2131–2136.

Oakes, P. J. (1987). The salience of social categories. In J. Turner, M. Hogg, P. Oakes, S. Reicher, & M. Wetherell (Eds.), *Rediscovering the social group: A self categorization theory* (pp. 117–141). Oxford, UK: Blackwell.

Oakes, P. J., Turner, J. C., & Halsam, S. A. (1991). Perceiving people as group members: The role of fit in the salience of social categorizations. *British Journal of Social Psychology*, *30*, 125–144.

Pietraszewski, D. (2009). Erasing race with cooperation: Evidence that race is a consequence of coalitional inferences. (Doctoral dissertation, UC Santa Barbara, 2009).

Pietraszewski, D., & Schwartz, A. (this volume). Evidence that accent is a dedicated dimension of social categorization, not a byproduct of coalitional categorization. Evolution and Human Behavior.

Pinker, S. (1994). *The language instinct*. New York: William Morrow & Co.

Rakić, T., Steffans, M. C., & Mummendey, A. (2011). Blinded by the accent! The minor role of looks in ethnic categorization. *Journal of Personality and Social Psychology*, *100*, 16–29.

Robinson, W. P., & Giles, H. (Eds.). (2001). *The new handbook of language and social psychology*. Chichester & New York: Wiley.

Sack, J.D. (2005). The effect of social essentialism on categorization and induction. (Doctoral dissertation, UC Santa Barbara, 2005). Dissertation Abstracts International, 66, 3469.

Sani, F., Bennett, M., & Soutar, A. U. (2005). The ecological validity of the "who said what?" technique: An examination of the role of self-involvement, cognitive interference and acquaintanceship. *Scandinavian Journal of Psychology*, *46*, 83–90.

Scovel, T. (1988). *A time to speak: A psycholinguistic inquiry into the critical period for human speech*. New York: Newbury House.

Sebastián-Gallés, N., & Bosch, L. (2005). Phonology and bilingualism. In J. K. Kroll, & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 68–87). New York: Oxford University Press.

Sell, A., Bryant, G. A., Cosmides, L., Tooby, J., Sznycer, D., von Rueden, C., et al. (2010). Adaptations in humans for assessing physical strength from the voice. *Proceedings of the Royal Society London (Biological Sciences)*, *276*, 575–584.

Shutts, K., Kinzler, K. D., McKee, C. B., & Spelke, E. S. (2009). Social information guides infants' selection of foods. *Journal of Cognition and Development*, *10*, 1–17.

Smith, P. M. (1979). Sex markers in speech. In K. R. Scherer, & H. Giles (Eds.), *Social markers in speech* (pp. 109–146). Cambridge: Cambridge University Press.

Stangor, C., Lynch, L., Duan, C., & Glass, B. (1992). Categorization of individuals on the basis of multiple social features. *Journal of Personality and Social Psychology*, *62*, 207–218.

Susskind, J. E. (2007). Preadolescents' categorization of gender and ethnicity at the subgroup level in memory. *British Journal of Developmental Psychology*, *25*, 213–225.

Taylor, S. E., & Falcone, H. T. (1982). Cognitive bases of stereotyping: The relationship between categorization and prejudice. *Personality and Social Psychology Bulletin*, *8*, 426–432.

Taylor, S. E., Fiske, S. T., Etcoff, N. L., & Ruderman, A. J. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology*, *36*, 778–793.

Tooby, J., & Cosmides, L. (1990). The past explains the present: Emotional adaptations and the structure of ancestral environments. *Ethology and Sociobiology*, *11*, 375–424.

Tottenham, N., Tanaka, J., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., et al. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, *168*, 242–249.

Van Twyver, M., & Van Knippenberg, A. (1998). Effects of group membership and identification on categorization and subtyping in memory. *European Journal of Social Psychology*, *28*, 531–553.

Weeks, M., & Lupfer, M. B. (2004). Complicating race: The relationship between prejudice, race, and social class categorizations. *Personality and Social Psychology Bulletin*, *30*, 972–984.

Williams, G. (1966). Adaptation and natural selection. Princeton, NJ: Princeton University Press.