

The ethological basis of certain signals of affect and emotion¹

John J Ohala, University of California, Berkeley

Abstract

A key assumption of ethology is that characteristic behavior of organisms is that which brings some benefit, or in the jargon, “increased fitness,” to them or their kin. Innate (genetically-determined) behavior has this characteristic, e.g., the broken-wing feint of different species of ground-nesting birds which draws prey away from the nest of chicks. (This is not to deny that characteristic behavior can also be socially-determined, e.g., obeying traffic laws, especially after the first penalty for not doing so). Morton (1977) has documented presumably innately determined factors which shape the vocalizations of several bird and mammal species’ in agonistic encounters (face-to-face competition). His generalizations apply equally well to certain behaviors that the human mammal also produces in agonistic situations but also, perhaps by some socially-determined extrapolation to characteristic intonations for question vs. declaration and to sound symbolic vocabulary designating small (and thus non-threatening and endearing) vs. large (threatening and impressive) entities (Ohala 1984). In this paper I review the evidence and theories underlying these connections between the sound shape and the meanings or communicative intentions conveyed by them and extend these principles to attempt to give a novel, if speculative, account of apparently similar cross-cultural use of the eyes and eyebrows in non-verbal communication where, again, the “messages” can be construed as conveying threat vs. non-threat. Two central element in this account are (1) the biological fact that the ratio of eye diameter to head diameter varies markedly with age: infants are said to have “large eyes”, in reality just a large ratio of eye diameter to head diameter, whereas in mature individuals this ratio is much smaller and (2) humans (and many other species) have the capacity to vary the apparent eye size and thus can exploit the apparent ratio of eye to head diameter for signaling emotion and affect.

Introduction

When someone speaks, several messages are transmitted. Acoustically the signal conveys information about the linguistic elements in the message (phonemes, syllables, words, phrases, and so on) – these are the elements that could be included in written transcript of the utterance (even if the full complexity of this component required IPA transcription). But the signal also conveys information about the speaker’s dialect (which may be a reflection of their geographical origin), their first language background or the degree of competence they have in the language they are speaking, their sex or sexual orientation, their approximate age, their state of health, possibly their level of education or “social refinement”, and their personal identity. In addition, the speech signal may convey something about the speaker’s attitude or emotional state (i.e.,

¹ This is a revision of Ohala 1996.

ironic, sincere, confident, subservient, etc.). In face-to-face communication some of the same “messages” are conveyed via the visual channel via facial expressions and other bodily movements and postures – the latter commonly known as ‘kinesics’. Recognizing and parsing these various “messages” is of considerable interest currently in the domain of human-machine interaction via spoken language. There is particular interest in identifying those aspects of the signal conveying attitude and emotion. In automatic speech recognition (ASR) identifying these elements could assist in indicating where in the utterance the speaker is putting emphasis, whether the speaker is cooperative or not or, perhaps, the probability that the speaker may be lying. For automatic synthesis of speech, including especially that form of synthesis that includes an image of a talking face, knowledge of how these so-called non-linguistic or extra-linguistic messages are conveyed could help to create more naturalistic synthesis incorporating some natural human emotions.

In this paper I hope to demonstrate that ethology – broadly, the study of the factors that contribute to species-specific and cross-species behavior – can provide us data, methods, and theories that can help us to understand how humans signal affect and emotion. The goal is to arrive at a theory of such signals that is as general as possible, encompassing both vocal and gestural signals in humans and non-humans.

Today most ethological explanations are expressed in evolutionary terms. Although all species exhibit a whole spectrum of behaviors, ethology focuses its energies on behaviors that are stereotyped – exhibited in more or less the same form given the same or similar circumstances – and behaviors which are species-specific or, even better, common to a variety of species – and, behaviors which are readily determined as contributing to the survival or “fitness” of the species. An example is the ‘broken-wing’ feint exhibited by several birds when a predator gets too close to the nest of young. This is found in such diverse species as the Little Ringed Plover, the Common Gallinule, Kildeer, the White-Crowned Sparrow, and the Great Horned Owl. It is presumably designed to lead the predator away from the nest when a more easily obtainable prey presents itself. Obviously to the extent that this trick works (the nestlings remain safe and the adult bird perpetrating the feint flies away, avoiding attack) this behavior increases the survivability or “fitness” of the species.

Not coincidentally, Charles Darwin (1872) was among the first to speculate in a systematic way about the genetic (inheritable) basis of certain behaviors, especially the expression of emotion in man and animals, e.g., he even pondered the possible relationship between characteristic mouth shapes exhibited in expressions of emotion and the nature of the sound produced – citing Helmholtz (1868). There were other many other earlier, less rigorous, speculations on the evolution of behavior, including some by Darwin’s grandfather, Erasmus Darwin (1803). But the accumulated literature on comparative behavior is a precious resource for any speculations on the origin of the expression of emotion and affect – including those presented here.

The fact is, there are remarkable similarities – in both macro- and micro-patterns – in the expression of affect or emotion in humans and certain non-human species, especially those non-human species anatomically and physiologically capable of using some of the same signaling modalities as humans, i.e., the vocal-auditory channel and modification of a plastic facial covering, i.e., “facial expressions”. Specifically, when it comes to facial expressions, only certain mammalian species have a “plastic” facial covering that can be modified for signaling purposes. This leaves out birds and most reptiles. But it includes other mammalian classes such as primates and canids. When it comes to vocal signals, even avian and amphibian classes may show patterns of behavior similar to humans and other mammalian species. As a result we humans believe we can “understand” the facial and vocal aspects of dogs’ and monkeys’ displays because at some basic level they resemble or are essentially the same as ours.

Identifying the “Function” of the Displays of Emotion and Affect.

An important theoretical and practical issue in the study of displays of emotion and affect is whether they arise directly from or at least mirror the inner physiological and psycho-physiological state of the signaler or whether they might constitute contrived or even deceitful signals whose purpose is not so much to reflect the state of the signaler but rather to influence the behavior of the receiver of the signal. In this latter case, given the default Darwinian dictum that the result of evolution is a something that increases the survivability or fitness of the individual, we would be asking if the stereotyped signaling behavior benefits the signaler. To begin to address this issue we have to acknowledge that certain vocal and non-vocal gestures are closely linked to the signaler’s inner physiological state even in situations where it might not be to the signaler’s benefit to allow this information to “leak out”, e.g., the slight tremolo evident in the fundamental frequency (F_0) in the voice of nervous or frightened individuals as well as the conceivably related tremor or shaking of hands and arms. Similarly, excessive perspiration in humans that is easily detected by others is commonly displayed by frightened individuals. Possibly in the same category of easily detectable signs of the inner psycho-physiological state of the individual is the reduced F_0 range of people who are depressed – often accompanied by lethargy and slow movements of limbs, head turning, etc. Even so, we can still entertain the hypothesis that some of the common emotional signals might better be viewed as designed to produce a response from those receiving the signal that is favorable to the signaler. Ethological studies suggest that this is, indeed, the case.

Voice F_0

An important breakthrough in this area originated with Morton (1977) who documented an impressive cross-species homology in the shape of the acoustic component of agonistic displays (those produced in face-to-face competitive encounters) by both mammals and birds (28 species in both orders): the confident aggressor emits a vocalization with a low F_0 (within the range that they are capable of) and which may be rough and aperiodic. A submissive vocalization, in

contrast, has a high F_0 and is tone-like (i.e., without any aperiodicity). The dog's aggressive growl and submissive yelp are familiar examples of these patterns. As it happens, Morton showed that the same pattern, labeled the "frequency code" by Ohala (1984), recurs in species ranging from the rhinoceros to the shrew. (It is also found in some species of toads (Davis & Halliday, T. R., 1978).) As Morton argued, this pattern can be explained as a kind of bluff (deceit) by the animal and the animals' exploitation of physical principles. The relative size of the animal is a good predictor of the outcome of a competition if, after all posturing and 'persuasions' have been exhausted, actual combat ensues. Therefore, to appear as large as possible in order to intimidate its adversary and thus forestall such a fight, the animal erects its hair or feathers, elevates its tail or tail feathers, and, in general, manipulates all the plastic features under its control to convey the signal "I am large and a threat to you". Since the F_0 of the "voice" of an animal is inversely related to its bodily dimensions, a low F_0 also enhances the gestalt impression of the vocalizer being large, mature, and potentially more threatening. The roughness or aperiodicity contributes to this impression, too, since larger vocal cords or syringeal membranes are likely to have secondary modes of vibrations which, when they interact with the primary vibration, create some aperiodicity. Conversely, an animal that submits to an aggressor in order to avoid a fight does the opposite to convey an impression that it is small and non-threatening. Some aspects of submissive displays quite clearly employ mimicry of infants and thus probably exploit the powerful genetically-dictated inhibitions against harming infants. Submissive dogs (or, indeed, most canids) that roll onto their back exposing their stomach are said to be imitating a routine common to puppies inviting maternal licking. Ohala (1984, 1994) presented evidence that the frequency code operates in human vocalizations, too: low F_0 to convey threat or self-confidence, high F_0 to convey non-threat or deference. He used natural utterances that were re-synthesized in a way to eliminate all spectral parameters of speech but retaining F_0 which had either the original F_0 or with an upshifted F_0 . Listeners judged samples with the lower F_0 as 'more dominant, more self-confident'. In addition in the case of two utterances similarly stripped of spectral details but with one have a falling F_0 at the end and the other a rising F_0 , listeners judged the one with falling F_0 as 'more dominant'. He also invoked the frequency code to account for the near-universal pattern where statements are coded by a low or falling F_0 but questions (that are not otherwise coded by grammatical or syntactic means) have a high or rising F_0 (Bolinger 1978).

Some important principles are manifested in such ethological accounts of the correlation between signal shape and signal function.

- First there should be a predictable and recognized relation between some parameter, P , of the signals and a feature, F , of the signaler that has some functional import, i.e., $P \propto F$, e.g., the rate of vibration of the laryngeal (or syringeal) folds and overall size of the signaler (where size is a predictor of success in competition).
- Second, that within a specified range, the signaler can vary P .

- Third, that depending on what seems will yield the most favorable outcome from a competitive encounter, the signaler can vary P in order to convey the impression of F' , different from its intrinsic F .

Two important qualifications must be made: in addition to parameters that can be varied, one can also identify anatomically implastic features that serve a similar function in signaling, e.g., piloerection in canids is plastic but the large mane of hair around the head of male lions and some monkeys are implastic. Also, though verbs of volition may sometimes be used to describe the signaler's intent, e.g., "The aggressive bird makes itself look bigger by ruffling its feathers, extending its wings and elevating its tail feathers (in order that it will subtend a larger angle in the view of the receiver)", this is ethological 'shorthand' for "The bird reacts [in such a way] because that behavior provides it (and it has provided its ancestors) a selectional advantage; it is genetically predisposed to act thus". No conscious strategy of willful deceit is attributed to the signaler – even to human signalers.)

The above principles probably apply to other vocal and kinesic displays as outlined the following sections.

The smile and the "o-face"

Ohala (1980, 1984, 1994) speculated that the smile, a non-threatening display, and what he calls the 'o-face' (because it seems not to have any other convenient label comparable to 'smile'), i.e., with lips constricted and somewhat protruded, which is exhibited in threats, may have originated as part of the acoustic component of submissive/aggressive displays: retracting the corners of the mouth (= 'smile') effectively shortens the vocal tract and thus raises the resonant frequencies; constricting and protruding the lips (= 'o-face') lengthens the vocal tract and thus lowers the resonant frequencies. These modifications of the resonant frequencies would contribute to the impression of the overall size of the signaler since normal resonant frequencies vary inversely with other linear bodily dimensions. Through ritualization the smile may have become to a large extent a visual (kinesic) signal but its origin is revealed in apes and monkeys where much the same mouth shape invariably accompanies high-pitched submissive vocalizations (Bauer 1987). This scenario resolves the long-standing paradox of why the smile, an affiliative display, shows so many teeth, normally regarded as potential weapons and thus seemingly more appropriate for an aggressive display.

(It is often thought that in humans the smile, at least, is a reflection (somehow) of inner contentment and well-being. But Kraut and Johnson (1979) have demonstrated convincingly "a robust association of smiling with a social motivation and an erratic association with emotional experience".

Sexual dimorphism of the vocal anatomy

It was mentioned above that signals can be plastic and implastic. Clearly the use of F_0 variations and mouth shape to express aggression or submission fall into the category of plastic signals. But, just like the lion's mane, there are implastic anatomical features in humans which may also convey similar signals. It is well known that there is marked sexual dimorphism in the human² vocal anatomy: adult males' vocal tracts are some 15% to 20% longer than those of adult females. But there is an even greater disparity in the larynx: males' vocal cords are some 50% longer than those of females. The 50% greater length probably implies a 100% greater mass that is free to vibrate. Accordingly, the resonances of the male vocal tract are lower than those of females and the male F_0 is on average a full octave below that of the female. These dimorphic traits begin to appear at puberty at the same time as the growth of facial hair in the male, another implastic signal akin to the male lion's mane (and for that matter countless other cases of sexual dimorphism documented in many species). The implastic signals have the same function as the "aggression"-related plastic signals.

It might be asked why it is the male's lot to have the implastic signals of aggression. The answer is that it is the male who has to compete for acceptance by the female (Puts, Gaulin, & Verdolini 2006) and this is where manes, facial hair and low voices are advantages. Males have more of the hormone testosterone and this inclines them to be, in general, more aggressive. There may be an underlying biological rationale for this: males have the less valued gamete; the ovum of the female is many times larger (than sperm), takes more metabolic energy to produce, there are orders of magnitude fewer of them in the reproductive life-span, and they are metered out at regular intervals. Being aggressive has its risks; males take this risk because, sociobiologically the loss is less than in the case of females.

Another elaboration of the above account of why high acoustic frequency of vocalizations (of F_0 and resonances) is associated with non-threat (and a lead-in to the next topic): it is possible the higher frequencies actively serve to inhibit any threat or aggression from the receivers of the signal because they approximate the kind of signals characteristics of the young. As mentioned earlier, for obvious reasons, every species must have an innate inhibition against aggression toward its young. Consider the case of human babies: given how much trouble and effort they require to be raised, if an innate inhibition against harming them did not exist, then neither would we (exist, that is). It is obvious that we find babies endearing, attractive, and worthy of our attention and sympathy.

I move on to an admittedly speculative account of facial expressions related to aggression and submission and involving the eyes.

Eyes and eyebrows

² But also in certain non-human species, as well, e.g., the howler monkey.

It is well documented and it is part of common lore than raised eyebrows signal non-threat or a friendly greeting whereas lowered eyebrows convey threat (Eibl-Eibesfeldt 1975, Ekman 1969). Again, this is true not only with humans but with non-human primates and canids as well (van Hoof 1962). A functional account of this pattern is often given in terms of the needs of vision: raised eyebrows when a wide field of vision is required and lowered brows for narrow, focused, viewing (Andrew 1972). But this would predict that in a situation where the signaler is conveying non-threat, i.e., submission or capitulation, when confronted with a specific recognized threat, lowered brows would be expected in order to closely monitor the source of threat. But the opposite is found. An alternative, ethologically-based, account is possible (Ohala 1994) which is more in line with the explanations given above for voice F_0 , voice resonance (and hence for mouth shape), viz., that eyebrows, like voice F_0 and resonances, can help to convey an impression of the size or maturity of the signaler. As is well known, the dimensions of the eyes do not change as much as overall bodily dimensions during maturation. Babies are well known to have “big eyes”. There are even metaphorical expressions testifying to this, such as the English phrase “wide-eyed innocence”. In fact the eyes do enlarge with progressive maturation but not as much as the head encasing the eyes. Babies at birth have an eye diameter that is about 75% to 85% the diameter that the eye will eventually achieve at maturity. Obviously other bodily dimensions, the head, spine, limbs, etc. grow much more in maturity. Thus the ratio of eye diameter to head diameter gradually gets smaller reaching an asymptote some time after sexual maturation. As a consequence, a visual estimation of this ratio (eye diameter/head diameter) is a rough indicator of age and size of signaler and thus the degree of potential threat posed. Subtle plastic variations in the portion of the eye externally visible accompany displays conveying threat and non-threat: threat involves a narrowing of this opening – what is commonly called in English, a “squint”. (German, “schielen”) whereas non-threat involves a widening of the eye opening. Now the eyebrows are not actually the outer visible limits of the eye but I would speculate that they convey a visual impression of their boundaries.³ This account also may help to explain the popularity (primarily in females) of cosmetics such as mascara, eyeliner, and eye shadow which, as cosmetologists will immediately confirm, make the wearer’s eyes “look bigger”. (<http://www.howcast.com/videos/441-How-To-Make-Your-Eyes->

³ It is an interesting question by itself as to why human have eyebrows at all. Some have suggested that they help to keep rainwater or sweat out of the eyes, others that they augment the function of the bony projection in the skull above the eyes in shading the eye from strong sunlight. As a hiker who likes hills and other sweat-producing exercises, I can personally testify that the eyebrows do not function well to keep rain or sweat out of the eyes. As for shading the eyes, it seems the eyebrows are a meager addition to what the skull shape already provides. It seems to me that all of these explanations ignore the fact that the eyebrows can be elevated or depressed by voluntary contraction of an array of facial muscles (the frontalis, the corrugator supercilii, and the obicularis oculi). Such plasticity seems more congruent with the idea that the eyebrows exist to serve a signaling function.

Look-Bigger). There may be some optical illusion at work here where providing more contrast between the necessarily white of the eye sclera and the surrounding flesh, gives an impression of larger eyes. This needs further investigation. Another possibility is that the contrasting colors draws attention to the eyes and this focusing of attention on them some triggers a cognitive impression of larger eyes.

Another cosmetic modification made in the pursuit of beauty is enhancements to the eyelashes: either making them seem longer or more prominent with mascara or with eyelash curlers or even the use of artificial eyelashes (invariably longer than the wearer's own eyelashes). There is no obvious way that these practices would make the eyes themselves seem larger but, as it happens, this could still be a manifestation of the strategy to achieve appearances that mimic those of babies. In a study quite remarkable (to me, at least), Pucci et al (2005) measured eyelash length in a large group of subjects of varying age from infants to adults. The study was primarily devoted to the effects of the pathological condition known as vernal keratoconjunctivitis but it included a control group without that condition. They reported that "[i]n healthy subjects, a negative correlation was found between eyelash length and age." It needs to be emphasized that the negative correlation was *not* to the *relative* length of the eyelashes, i.e., proportional to other anatomical dimensions, rather it was to the *absolute* length of the eyelashes. Babies, it seems, have longer eyelashes (on average) than do adolescents and adults. That babies do have long eyelashes is confirmed by an informal sample of mothers that I consulted. I refrain from speculating as to why eyelash length varies in this way but it seems credible that the enhancement of eyelashes in adults is another way to mimic this trait in infants and thus to evoke in viewers some of the same innate positive reactions.

Why to babies (of many species) have a relatively large ratio of eye diameter to head diameter? The probable answer is emmetropization (van Alphen 1990). Emmetropization is the feedback relationship between the retina and the processes which govern the growth of eyes. It has been shown that in order to preserve a well-focused image on the retina, the growth of the eye is inhibited.⁴

There is evidence, as a consequence, that, other things being equal, babyish faces (even those of adults – think of the silent movie stars “Fatty” Arbuckle or W. C. Fields – are more closely interpreted by viewers as exhibiting fear than are angular mature faces (think of another silent

⁴ Experiments with animals show that if one eye is covered or the eyelids sutured together, there is abnormal growth of the occluded eye. Nature sometimes implements the same experimental controls, even in humans, when one eye is affected by cataracts or other abnormal conditions which prevent normal focusing of images. In these cases, too, abnormal, excessive growth of the eye results.

movie star, William S. Hart)⁵, which, in turn are interpreted by viewers as exhibiting anger (which is closely associated with threat) (Marsh, Adams, & Kleck 2005).

Conclusion

In the preceding discussion I have focused on cases where a given vocal and some non-vocal signals of what is called “emotion” or “affect” are designed to evoke from the viewer a response favorable to the signaler. These include signals involving plastic modulations of voice F_0 and vocal tract resonances (the latter primarily due to the shape of the mouth), as well as plastic modulations of the eyes and nearby structures. I have also brought into the same explanatory scheme implastic signals such as the growth of hair around the perimeter of the face as well as quasi-plastic cosmetic modifications of the eyes. This is not to deny that some aspects of emotional or affectual signals may in some way reflect the inner psychological or physiological state of the signaler, e.g., tremor or rapid eye-blinking of nervous or fearful individuals. And many other signals may be completely voluntary and possibly culturally-derived signals emerging from the conscious state of the signaler, e.g., tipping the hat, waving with hands, nodding the head, etc. There is, as we all know, a very rich “vocabulary” of signals of emotion and affect and only a fraction of them have been adequately explained by theory or explored empirically. The message I would like end with is: we can gain much in being able to differentiate these various types and origins of such signals from comparison of signals humans give with those given by animal in this same domain.

References

- Van Alphen, G. W. H. M. 1990. Emmetropization in the primate eye. Ciba Foundation Symposium 155: Myopia and the control of eye growth. 115-125.
- Andrew, R. J., 1972. Information potentially available in mammal displays,. in R. A. Hinde (ed.), *Non-verbal communication*. Cambridge University Press, Cambridge, 179-206, 1972.
- Bauer, H. R. 1987. The frequency code: oral-facial correlates of fundamental frequency. *Phonetica* 44.173-191.
- Bolinger, D. 1978. Intonation across languages. In Greenberg, J. (ed.). *Universals of human language*. Stanford University Press. 471-524.
- Darwin, Ch. 1872. *The expression of the emotions in man and animals*. John Murray, London. [http://www.darwin-literature.com/The_Expression_Of_The_Emotions_In_Man_And_Animals/index.html]

⁵ If these names or their images are not familiar to the reader, see <http://www.goldensilents.com/>.

- Darwin, E. 1803. *The temple of nature; or the origin of society: a poem with philosophical notes*. J. Johnson: London,
[<http://www.english.upenn.edu/Projects/knarf/Darwin/templetp.html>]
- Davis, N. B. & Halliday, T. R., 1978. Deep croaks and fighting assessment in toads *Bufo bufo*. *Nature* 274.683-685,
- Eibl-Eibesfeldt, I. 1975. *Ethology. The biology of behavior*. 2nd ed. Holt, Rinehart, and Winston, New York.
- Ekman, P. 1969. Pan-cultural elements in facial displays of emotion. *Science* 164.86-88,
- Helmholtz, H. 1868. *Theorie physiologique de la musique*. Paris.
- van Hoof, J. A. R. A. M., . 1962. *Facial expressions in higher primates*,. *Symposium of the Zoological Soc. of London, No. 8: Evolutionary aspects of animal communication*. 97- 125.
- Kraut, R. E. & Johnston, R. E. 1979. Social and emotional messages of smiling: An ethological approach. *J. Personality & Social Psych.* 37.1539-1553.
- Marsh, A. A., Adams, Jr., R. B. & Kleck, R. E. 2005. Why do fear and anger look the way they do? Form and social function in facial expressions. *Personality Social Psychology Bulletin*. 31.73-86.
- Morton, E. W. 1977. On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *American Naturalist* 111.855-869.
- Ohala, J. J. 1980..The acoustic origin of the smile [Abstract]. *J. Acoust. Soc. Am.* 68.S33,
- Ohala, J. J. 1984. An ethological perspective on common cross - language utilization of F₀ of voice. *Phonetica* 41.1 - 16,
- Ohala, J. J. 1994.. The frequency codes underlies the sound symbolic use of voice pitch,. In L. Hinton, J. Nichols, & J. J. Ohala (eds.), *Sound symbolism*. Cambridge: Cambridge University Press. 325-347.
- Ohala, John J. 1996. Ethological theory and the expression of emotion in the voice. In *Proc., ICSLP-1996*, 1812-1815.
- Pucci, N., Novembre, E., Lombardi, E., Massai, C., Bernardini, R., Caputo, R., Campa, L., de Libero, C., & Vierucci, A. 2005. Long eyelashes in a case series of 93 children with vernal keratoconjunctivities. *Pediatrics* 115.e86-e91.
- Putz, D. A., Gaulin, S. J. C., & Verdolini, K. 2006. Dominance and the evolution of sexual dimorphism in human voice pitch. *Evolution and Human Behavior*. 27:283-296.