18 Voice and Gaze Considered Together in 'Languaging'

Fred Cummins

From Language to 'Languaging'

Gaze is by no means a purely human phenomenon. Any animal with jointly directed eyes may be described as exhibiting gaze, which makes the independently rotating eyes of the chameleon so perplexing, and thus well suited as a comic vehicle in animated films. Many conversations we might have about gaze will begin by assuming specific affective qualities to any instance of gazing, or, more potent yet, mutual gaze. Here, it seems is the confrontation with the other. Here is the locus at which we fuse, touch, and acknowledge each other, or, conversely, at which we refuse to open up, creating a rift between us.

But we have just leaped somewhat rashly from the impersonal mode of observation, used to pick out patterns of the gaze of chameleons and persons alike, to the qualitative, phenomenological theatre of social intercourse. I wish, in this contribution, to abide a while at the surface and to begin by considering gaze in the impersonal mode, from without, but to do so in order to interrogate the relation between gaze and that most human of characteristics, language. To do so, it will be necessary to first note that the term *language* picks out different objects of interest, depending upon the manner in which it is initially framed.

One way to approach language is to ask about the means by which humans pass messages among themselves. It is this capacity that has led to the characterisation of language as a distinct system, involving several related sub-domains, such as phonology, morphology, and syntax. This view of language gave rise to the structuralist revolution (de Saussure 1916) at the start of the 20th century and led directly to the development of the generative programme in linguistics in the 1960s (Chomsky 1957, 1965) Whether structuralist, generativist, or their several competitors, such fields set out to understand an abstract system, largely independent of the incarnate context in which it is used, expressed indifferently in speech, writing, or sign and facilitating the encoding and decoding of such messages. To be linguistic, within any such account, is to be defined by categorical contrasts, and the language system is understood as a set

of such contrasts, at the level of sound (phonology), structure (morpho-syntax), and meaning (formal semantics). The basic unit to be understood is the sentence, such as 'The cat sat on the mat', and the sentence is a product of a specific instance of a language, in this case English. This broad approach has informed most of the formal academic study of language since its introduction.

This characterisation of language is built upon a Cartesian background in which minds are individual, discrete, personal, and unobservable (Cummins 2018). Many interesting questions and answers can be formulated within such approaches. This essay is not the context to pursue and critically appraise these approaches. But the aspirations of many linguists to develop an account of language that is continuous with advances in psychology and the social sciences, on the one hand, and biology, on the other, remain largely unfulfilled, not least because the relation of psychology and the social sciences to biology is as contested as ever, and probably even less secure than it seemed 50 years ago. The framing that is required to characterise language as encoded message passing is not obligatory, and alternative frameworks that emphasize the importance of the body, of the intersubjective, and of mutual entanglement in concrete situations of embodied co-presence, are also available (Stewart, Gapenne and Di Paolo 2011; Chemero 2011).

One can approach the broad topic of language in a rather different fashion though. If we begin by asking what it was that so radically transformed our species in the 5 or 6 million years since the last common ancestor of the genus *homo* on one side, and our relatives the chimpanzee and bonobo on the other, it is not clear that we are talking about the same thing as before. It is difficult to document where anything that we might consider as human language enters the picture, but the earliest evidence we have of human-like capacities for symbolism and ritual, as evidenced in cave art, bone flutes, bodily ornaments, and the residue of rituals, goes back no more than 100,000 years and probably considerably less. This is a tiny amount of time within which processes of biological evolution might work, and a comparison of human and chimpanzee anatomy does not reveal any obvious differences that might be responsible for the emergence of language.

It is notable that for almost all of the time in which language has existed, it has found expression in an oral mode among participants who are in each other's bodily presence. The very earliest forms of writing are no more than 5,000 years old, while widespread literacy is much more recent, extending back no more than 500 years or so. These are very recent innovations and just as an investigation of writing itself would not start with emojis, so an inquiry into the transformation of the species would do well not to begin by abstracting away from the embodied context of face-to-face exchange.

In this matter, Tomasello and colleagues have brought to our attention one small biological change within our lineage that, while it can

carry only a limited explanatory burden, serves to bring to the fore the manner in which the voice and the gaze work together (Tomasello et al. 2007). Human eyes have a white sclera, against which the iris and pupil stand in stark contrast, which is very different from the almost uniformly dark appearance of the ape eye. Figure 18.1 illustrates the difference, and adds, for good measure, a Hollywood creation. The ape in the right panel is from the film *Planet of the Apes*, and its white sclera subliminally conveys a human-like intelligence or cunning.

Humans are exquisitely sensitive to the direction of gaze of others. Apes too are interested in where other apes are looking, but they must rely upon the cruder signal of head direction rather than eyeball orientation (ibid). One consequence of this is that human infants are cocooned in a web of joint attention from their earliest days (Baldwin 1995) learning to look where their peers and caregivers look, and, by about the age of 18 months, learning to direct the gaze of caretakers to a spot of their choosing. This sharing of attentive focus offers itself as a suggestive locus for a radical change in the patterning of social intercourse, ensuring a great deal of alignment in the concerns of interactants.

When we look at how gaze is employed in adult conversation, we see a delicate dance that usually goes unnoticed. When one person starts a speaking turn,[1] she frequently looks away, while the other, currently in the role of listener, is more likely to continue looking at the face of the speaker. As the exchange continues, gaze is employed to encourage, or resist, the alternation of the roles of speaker and listener. Even blinks are drawn into the fabric of the interaction. Figure 18.2 shows data from a study in which eight dyads conversed freely for 15 minutes each (Cummins 2012). The plot shows the proportion of time each participant (speakers are indexed by letter, dyad by number) gazed towards the face of the other when he was speaking and when he was listening. In every case, gaze towards the other is more likely when listening than when speaking.

Such considerations open up the way to a more expansive characterisation of vocal interaction and interpersonal coordination, noting other co-occurring features that are frequently observable. Once we admit gaze

Figure 18.1 Left: human eye with white sclera; centre: chimpanzee face; right: ape eyes from *Planet of the Apes*

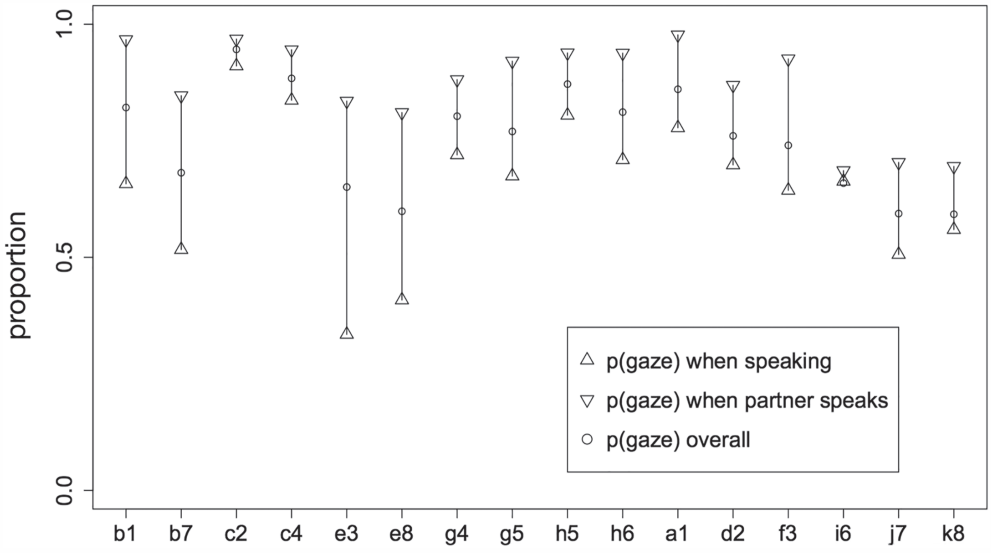

Figure 18.2 Proportion of time gazing towards the head of the partner as a function of speaking turn. Data from Cummins (2012)

into the salient aspects of linguistic communication, we are drawn inexorably to also consider co-produced manual and head gestures, which elude a classically linguistic description, but which are inevitably present in conversation, even among blind discussants or when speaking on the phone. There are many other aspects of the voice itself that do not lend themselves to categorical description and symbolisation, such as the tempo of speech, the intonation contour, the rhythm of production, and the tone of voice. All of these loosely grouped under the heading of 'prosody' have attracted a great deal of attention in the strongly empirical world of phonetics, despite the difficulty of assimilating them under a symbolic linguistic description. None, of course, are represented in writing. Finally, the purposes to which the voice is employed, and its relation to the context in which it is used, all require scrutiny. Collectively these considerations encourage a broader view of vocal communication that we might term *languaging*, rather than *language*, in recognition of the obvious possibility that we are not picking out a single 'system', but rather identifying many ways in which different forms of mutual coordination arise and in which participants (sometimes speakers, sometimes listeners, sometimes neither or both) become dynamically entangled as they interact.

But dyadic conversation does not exhaust languaging, and the study of vocal coordination and communication has focussed almost exclusively on such conversation. Most of my work over the last 15 years has explored a rather different use of the voice, attending to those situations in which *multiple people utter the same words at the same time*. I call such speech *joint speech* (Cummins 2018).

Joint Speech

It has been my experience that whenever joint speech is found, there is something interesting going on. Joint speech is very common, but it is not found indifferently everywhere. It is found primarily in several highly specific domains of social organisation, and these domains play an important role in founding the many social identities we have adopted. The practices in which joint speech arises serve to create and sustain a common lifeworld.

The first, and largest, domain is that of ritual and prayer. Praying or chanting aloud together is found within most religious traditions. Secular rituals too employ joint speech, as when swearing a collective oath of allegiance together, whether it be to the United States of America or to the Islamic State. Prayer and ritual admit of a dizzying variety of forms and purposes, but it is evident that joint speech is reliably present in such activities.

Another domain that is picked out by this simple definition is that of protest. Protesters the world over chant in unison as they make manifest their grievances. Despite the radically different overt purpose of protest, many of the surface features of joint speech in a religious or official context are also found in protest demonstrations. Thus, we may find alternation between leader and follower in call and response. The cherished intricacies of a liturgical text or a political argument may best be enunciated by a leader, while participation in the collective act of uttering is made possible by the jointly spoken indication of generalised assent, e.g. as 'Amen' in the church or 'Right on!' in the street.

A third domain that becomes apparent is the use of chant in the eager expression of group identity among sports fans. Not every sport has a chanting tradition. Soccer famously does, rugby has a very different singing tradition, while tennis fans do neither. But where fans indulge in chanting, it is an important vehicle for making manifest group membership. (Tennis, of course, is not based on teams that persist over generations.)

The final broad domain in which joint speech is predictably found is in the education of young children. Here, as with ritual and prayer, there is great variability in the specifics as we move from one culture and country to another, but in each case, joint speech will be used to marshal the collective attention of young children, to inculcate culturally valued texts, and to aid in rote memorisation.

Joint speech occurs in many other situations as well, both formal and informal, and it serves many purposes. However, these four domains speak to the centrality of the associated activities in establishing, enacting, and proclaiming identity, concern, and belonging, and it is clear that when joint speech is at work, it is collective concerns, not individual ones, that are in play. Cummins (2018) provides an overview of the topic, summarising the scant scientific work done that does address joint

speech, and further resources pertaining thereto can be found at https://jointspeech.ucd.ie.

Evidence of liturgical structure employing joint speech can be found as far back as 2600 BCE in the Temple Hymn of Kesh, a Mesopotamian text regarded as perhaps the oldest literary text in existence. This text, which survived more or less unchanged for one thousand years, even as the ambient language switched from Sumerian to Akkadian, has a clear verse-chorus structure, with identical recurring lines at the end of each verse, suggesting that chorusing was well known long before writing emerged. This is independently supported by the observation that joint speech is found in every human society, including those that are purely oral in character.

How remarkable then that the science of language, linguistics, has not made joint speech an object of concerted study at all. In fact, before its thematization (Cummins 2018), no single term existed to pick out the *collective simultaneous uttering of identical words*. But when we examine the obvious simple structural features that reliably characterise joint speech across these several domains, it becomes clear that joint speech will not fit into the mould crafted for the exchange of encoded messages. Happily, the more expansive term *languaging* allows us to avoid such narrow definitional anxieties.

In joint speaking, there is no longer a distinction between speakers and listeners. With this simple conflation of roles, most of the concepts used to characterise language are rendered useless. What is more, the texts that are spoken together are typically known by all who participate, though they been authored elsewhere. This contrasts starkly with the emphasis on the creativity of natural language, which has played such a significant role in establishing the generative tradition in linguistics. However, it aligns with the keen observation of Rappaport (1999) that one of the defining features of ritual, understood as a foundation for human society, is the 'performance of more or less invariant sequences of formal acts and utterances *not entirely encoded by the performers*' (emphasis mine).

Another feature of joint speech, which is not obligatory, but so prevalent that its absence is more remarkable than its presence, is repetition. It is not enough that something be said, it must be said again and again. The act of uttering reveals a performative dimension to joint speaking, drawing our attention to the fact that something is happening in the uttering, something no written trace could replace, which is of urgent importance to those who take part. The role of rosary beads among Catholics has its counterpart in the prayer beads or mala of Hindu, Buddhist, Islamic, and Sikh traditions too, once more showing how the observable characteristics attending joint speech transcend the extravagant cultural variation exhibited by the attendant activities.

As we survey the many and varied activities that the definition of joint speech picks out, another important characteristic comes to the fore.

Where it is entirely conventional to distinguish between speech and song, apportioned to the distinct domains of language and music, respectively, no such partition can be maintained when looking at joint speech and associated practices. Some joint speech is clearly spoken and has the prosody of read text, as when an oath of allegiance is read one time only by initiates at a naturalisation ceremony. This particular purpose requires that the words be spoken aloud and collectively, but its function is entirely instrumental. The speaking of the words a single time is a performative act in the strict sense of Austin (1962) in that the status of the speaker is changed by the act of uttering. However, this particular situation is somewhat anomalous, and lacks the much more common feature of repetition that usually accompanies activities of prayer, of protest, and even of support for one's home team.

When a short phrase is repeated over and over, the prosody (i.e. the more musical aspects of the voice) become exaggerated. Stresses acquire a greater degree of temporal regularity with heightened contrast between strong and weak syllables. Strongly enunciated stresses are frequently combined with gestures such as bowing, fist pumping, or clapping. A repeated intonation contour may undergo a perceptual shift from speech to song simply through repetition (Deutsch, Lapidis and Henthorn 2011). Furthermore, although some instances of joint speech are clearly musical in a strongly aesthetic sense, e.g. the plainchant of monastic communities, many other instances occupy an odd and poorly documented space that is neither clearly musical, nor non-musical, and which has many of the characteristics of formal ritual without the attendant formality and explicit ideology. The familiar act of singing Happy Birthday comes to mind. The words are sung, but the resulting sound is never heard as music in an aesthetic sense. As familiar as this little ritual is, it is worth drawing attention to its coordinating function in drawing all participants into the collective performance, such that refusing to join in would be considered anomalous. Although not part of any institutional agenda, the ritual nonetheless is part of the means by which the individual is identified, demarcated, and with that, celebrated, within a community.[2]

Languaging may thus be continuous with a broad notion of *musicing*, a term that now invites consideration of situations of joint coordination beyond the practices of performance and participation we usually associate with the term *music*. That the boundaries of music are not given, but are constructed in a culturally specific manner is obvious and is well illustrated by the prohibition of many forms of music within the austere Wahhabi form of Islam, which lives alongside the use and propagation of sacred monophonic chants, as found, e.g. in the propaganda of the Islamic State or in the *adhan* call to prayer. We can now note the role of drums as used sometimes in the coordination of teams of rowers, or the use of working songs among slaves and chain gangs, further

suggesting that a sharp demarcation between languaging and musicing would obscure rather than reveal the import of such activities.

Joint Speech and Gaze

We noted earlier that the gaze in a two-party conversation is a constitutive part of the mutual dance and negotiation that unfolds. Disturbance to this patterning is familiar from the gaze behaviour of people on the autistic spectrum (Hutt and Ounsted 1966), which has informed diverse approaches to the characterisation of the attendant social difficulties such people face (Baron-Cohen, Baldwin and Crowson 1997; Dickerson et al. 2005). We might place this dynamic negotiation alongside the far less contentious sharing of common ground among participants in a ritual such as a liturgy. As indicated in Figure 18.3, this alignment among participants reaches an extreme in the recitation of the Credo. Viewed in this light, we might expect the patterning of gaze to depend strongly on the manner in which understanding is being arrived at through dialogue, negotiation, and argument or emerging from common alignment with common purpose. Indeed, what we find is that the gaze in joint speech is markedly different from the patterning of gaze in dialogue.

We can distinguish two principal ways in which gaze is employed when people utter in unison. In one case, there is a common central focus, such as a leader, priest, a blackboard, or just a person with a bullhorn. In this case, all eyes are turned towards this central element, but there is no negotiation anymore, no back-and-forth or jockeying for the floor. This is an obvious, but potentially overlooked, basic structural property of many such situations. As with so many of the reliable features of joint speech,

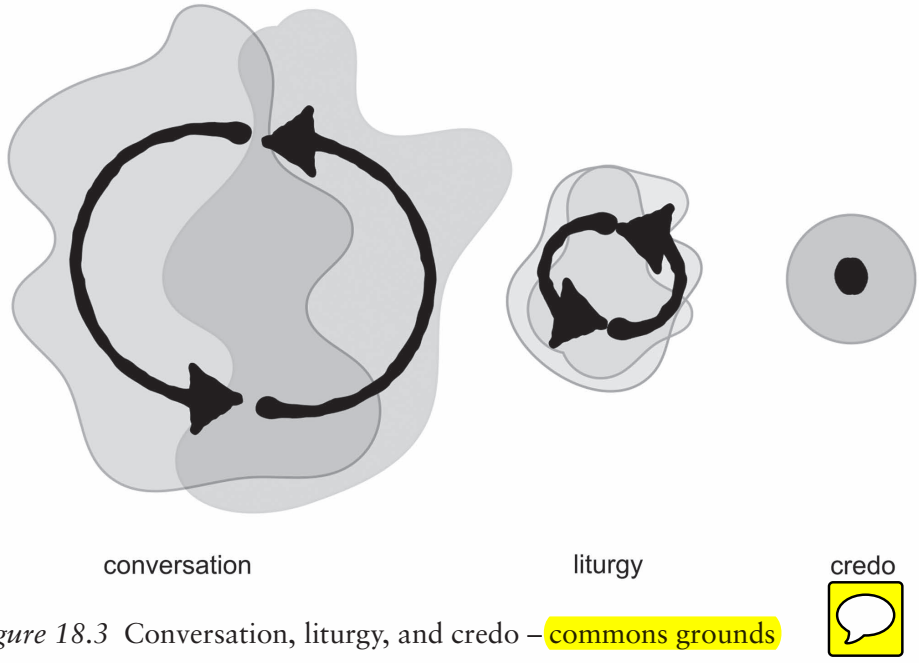

Figure 18.3 Conversation, liturgy, and credo – commons grounds

it is a common characteristic that transcends domain, from the temple to the street, the stadium to the classroom. Coming back to our speculations about the role of the signal available through the white sclera of the eye, this is a situation of exaggerated joint attention, ensuring that those who participate share, at least transiently, a common outlook on the world, its opportunities, and its threats.

Such situations can be differentiated from joint speaking occasions in which there is no obvious central figure. Here, uniquely in situations of languaging, the eyes are uncoupled from the voice and are free to roam independently. This is probably the case in many situations in which Happy Birthday is sung, where a cake can provide an optional central element, but typically participants are not obliged to stare at the person being celebrated and may allow their gaze to roam more or less freely. It is the case in a moving street protest, as participants march with common direction, and it seems to frequently be the case in collective prayer, such as recitation of the rosary or celebration of kirtan. I have argued that we might usefully regard the joint silence observed after tragedy (and frequently held in a sports stadium) as a limiting case of joint speech in which the verbal content is reduced to nil, but the participatory element remains as strong as ever. During such silences, there is no particular constraint on the gaze.

The integration of joint speech into rituals of many sorts means that gaze may be caught up in the group choreography, and group gaze may be strongly constrained. Thus, we find collective gaze shown by those practicing the Maori Haka, and this is aimed with precision at those they stand against, while synchronized head movements and collective gaze are a signature feature of the Balinese Monkey Chant.

In all of these cases, we can see that the dialogical dynamic that inheres in conversational exchange is absent in joint speech, thereby radically altering the relation between the voice and the gaze. Of course, for all of these situations of dialogue or chorusing, the voice and the gaze are merely two prominent identifiable contributors to what is a whole-body-and-context involving process of mutual coordination, but such identifiable characteristics offer us a secure means of empirically distinguishing among classes of activity, with markedly different relations for the purposes of participants. As we move from the narrow preoccupation of traditional linguistics with an abstract, bounded system of oppositions to the broader and largely uncharted territory of languaging, such landmarks may allow us to retain the analytic frame of mind required to make such interactions intelligible, without prematurely collapsing them into a single domain of theoretical construction.

Joint Gaze?

Speaking is usually thought of as something done by one person at a time, and a reframing of our consideration of speech to focus instead on joint,

synchronised speaking has made it clear that the practices in which joint speaking arises are precisely those that serve to enact social identities, as fans, as passionate believers in a cause, as a people, or within a religion. There is no claim here that joint speech achieves this in any mechanistic fashion. Rather, by attending to where, when, how, and why joint speaking occurs, we provide a valuable lens into the means by which social identities are enacted and collective subjectivities arise. Can we pursue a similar strategy with gaze?

Gaze is obviously normally considered as something associated with a single viewer, though we can readily recognise the highly charged nature of mutual gaze between two individuals. In the pursuit of this investigation of language and languaging, then, I wish to conclude by indulging in a little speculation about the possible role of synchronised gazing by very many people. We already noted that gaze is aligned and unified among many individuals during certain rituals and situations of common attention. However, we might extend this line of thought to consider a situation in which the voice is mute, but the eyes of many people are strongly coordinated. This is the case when viewing cinematic images on a large screen. If one could observe not the movie but the eyeballs of the viewers in a cinema, it seems certain that they would be very strongly aligned in what looks like an act of collective sense-making. Alfred Hitchcock appositely said that he enjoyed playing his audience like a piano.

Some evidence that this is, indeed, the case can be found in a study by Hasson et al. (2004), which revealed a high degree of temporally correlated activity in several cortical areas of subjects watching a single clip from the movie *The Good, The Bad, and The Ugly* (Leone 1966). Hasson et al. (2004) claimed that they were revealing a 'surprising tendency of individual brains to "tick together" during natural vision'. While recognising and seeking to understand the emergent synchrony in the firing of neurons across subjects seems like a compelling task for the empirical scientist, and we might applaud the technical challenges overcome in recording such synchrony, one might reasonably object that the very concept of 'natural vision' is deeply problematic, especially in a situation in which immobile viewers consume highly constructed, technically intricate forms of media. But reading brains is something of a black art that demands a kind of Cartesian orientation foreign to the spirit of the present discussion.

We could choose rather to stay once more at the surface, and look instead at the eyes themselves, as different viewers watch the same film clip. Figure 18.4 shows the horizontal movement of the eyes of ten subjects as they watched a 3-minute clip from the same movie, as captured by an eye tracker (unpublished data of my own).[3] Below the movement tracks themselves is a computation of the synchrony among two independent groups of five subjects: points at which both the blue and red lines lie above the dotted line are ones where we can confidently say

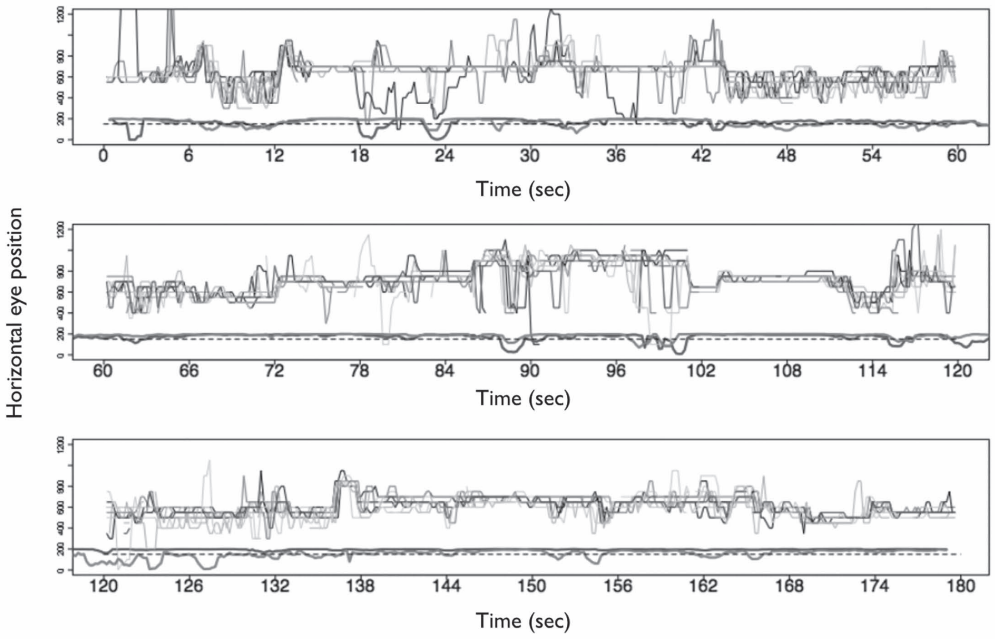

Figure 18.4 Horizontal eye position viewing a 3-minute clip from *The Good, The Bad, and The Ugly*; estimates of inter-viewer synchrony

that viewers' direction of gaze is strongly coordinated.[4] Sergio Leone, the director of the film, is renowned for his stark cinematography which makes it very clear just what the viewer is supposed to be attending to. This provisional foray into gaze and film cannot be presumed to generalise to arbitrary cinematographic material. However, it does demonstrate that, under specific viewing conditions, we can observe the yoking of the eyes as the voice is silenced and gesture stilled.

Synchronised voices and synchronised gaze both suggest that there is a story yet to be told about the body as a medium through which collective sense-making occurs. In the sparring and dialogue of interpersonal interaction, we may be striving for common understanding, or we may be asserting an individual stance and distinguishing ourselves from others. But in joint speaking, there is a collectivisation of sense-making, and in the associated practices we can see the laying of the foundation of a common social life and the enactment of a collective subjectivity. This seems to have been a constitutive part of human society as far back as we can peer in history and may provide one clue as to why speech became such a powerful medium in the construction of a common human world.

Treating gaze in a similar fashion provides food for thought (and nothing more is being claimed here). The forms of media change over time at an ever-accelerating rate. Cinema has without question been a potent part of this evolution of media, and its persistence and refinement

are testament to its efficacy in yoking together the sensibilities of people engaged in otherwise quite disparate lives. In a very short period, we have become immersed in a world of moving images that we read with a vocabulary developed, trained even, in cinematographic terms. The qualitative leap in interpersonal coordination we have tentatively associated with the facilitation of joint attention at the origin of our species may not be the only such leap we are capable of. While this essay is a very preliminary sketch of this rich research domain, it is clear that the syntax and semantics of the image, and the moving image, provide fertile ground for further discussion than we can attempt here. Nonetheless, our consideration of joint speech and joint gaze may provide an impetus to consider how our shared world arises, in part, through such synchronisation in sense-making.

Notes

1. To speak of turns in conversation is somewhat misleading. Players in chess and speakers in formal debate take turns. Conversation partners alternate in a manner more reminiscent of boxing opponents, opportunistically taking and defending the floor and negotiating its relinquishing.
2. It may be worthwhile to ask which collective gatherings support the singing of Happy Birthday and which do not. It would be inappropriate in a transactional social context like a supermarket, but it may be a form of solidarity among workers in a commercial office or factory.
3. In an unpublished pilot experiment conducted at University College Dublin (2017), subjects sat at arms' length from a computer screen with built-in eye tracking cameras as they watched several media clips. Eye movements were recorded in the horizontal and vertical directions separately. Only horizontal movements are shown here. Recordings were done at 60 Hz.
4. Synchronisation among viewers must here be understood to be with respect to film time, rather than time indexed by clocks and calendars.

Reference List

Austin, J.L. 1962. *How to Do Things with Words*. Oxford: Clarendon Press.

Baldwin, D.A. 1995. Understanding the link between joint attention and language. In C. Moore and P.J. Dunham (eds.) *Joint Attention: Its Origins and Role in Development*. Erlbaum, 131–158.

Baron-Cohen, S., Baldwin, D.A., and Crowson, M. 1997. Do children with autism use the speaker's direction of gaze strategy to crack the code of language? *Child Development*, 68 (1): 48–57.

Chemero, A. 2011. *Radical Embodied Cognitive Science*. Boston, MA: MIT Press.

Chomsky, N. [1957] 2002. *Syntactic Structures*. Berlin: Walter de Gruyter.

Chomsky, N. [1965] 2014. *Aspects of the Theory of Syntax*. Boston, MA: MIT Press.

Cummins, F. 2012. Gaze and blinking in dyadic conversation: A study in coordinated behaviour among individuals. *Language and Cognitive Processes*, 27 (10): 1525–1549.

Cummins, F. 2018. *The Ground from Which We Speak: Joint Speech and the Collective Subject*. Newcastle, UK: Cambridge Scholars.

de Saussure, F. [1916] 2011. *Course in General Linguistics*. New York: Columbia University Press.

Deutsch, D., Henthorn, T., and Lapidis, R. 2011. Illusory transformation from speech to song. *Journal of the Acoustical Society of America*, 129: 2245–2252.

Dickerson, P., Rae, J., Stribling, P., Dautenhahn, K., and Werry, I. 2005. Autistic children's co-ordination of gaze and talk: Re-examining the 'asocial' autist. In K. Richards and P. Seedhouse (eds.) *Applying Conversation Analysis*. Springer, 19–37.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., and Malach, R. 2004. Intersubject synchronization of cortical activity during natural vision. *Science*, 303 (5664): 1634–1640.

Hutt, C. and Ounsted, C. 1966. The biological significance of gaze aversion with particular reference to the syndrome of infantile autism. *Behavioral Science*, 11 (5): 346–356.

Leone, S. 1966. *The Good, the Bad and the Ugly*. Los Angeles, CA: United Artists.

Rappaport, R.A. 1999. *Ritual and Religion in the Making of Humanity*. Cambridge, UK: Cambridge University Press.

Stewart, J., Gapenne, O., and Di Paolo, E. 2011. *Enaction: Toward a New Paradigm for Cognitive Science*. Boston, MA: MIT Press.

Tomasello, M., Hare, B., Lehmann, H., and Call, J. 2007. Reliance on head versus eyes in the gaze following of great apes and human infants: The cooperative eye hypothesis. *Journal of Human Evolution*, 52 (3): 314–320.