# The Ground From Which We Speak

*Joint Speech and the Collective Subject*

Fred Cummins

# The Ground From Which We Speak

## Joint speech and the collective subject

Fred Cummins
*University College Dublin*

To expedite publication, this edition of The Ground From Which We Speak is being published by the author. Commercial publication may follow.

The author may be contacted at:

Fred Cummins, School of Computer Science, University College Dublin,
Dublin 4, Ireland
Email: fred.cummins@ucd.ie
Web: pworldrworld.com/fred

© Fred Cummins 2018
First published, 2018

ISBN 978-1-9999532-0-1 formatted for A4 sheets/large screen
ISBN 978-1-9999532-1-8 formatted for small screens

# Contents

**Prologue** **5**

**I Introducing Joint Speech and the Subject** **9**

**1 Some Initial Observations** **11**

**2 Amen! The structures of joint speech** **28**

**3 Science(s) and Subjects** **42**

**II The Science of Joint Speech** **61**

**4 The Sounds of Joint Speech** **63**

**5 Joint Speech and Movement** **80**

**6 Joint Speech and the Brain** **107**

**7 Language and the Voice** **125**

**III Beyond Science as Normal** **150**

**8 An Enactive Starting Point** **152**

**9 Making Use of Joint Speech Observation** **175**

**References** **192**

# Prologue

The singing of Happy Birthday. The recitation of the Nicene Creed. The countdown to midnight on New Year's Eve. "U.S.A.! U.S.A.!" spontaneously chanted in a high tech control room. The outrage of Syrian citizens during the popular uprising of 2011. The frenzy of "Lock her up!" as chanted during an election campaign. These diverse snapshots open a window onto the use of the voice in making collective purposes and collective identities manifest. They all take place as part of activities apportioned a great deal of significance to those who take part in them. They all make use of the voice in a specific manner: in each case, many people say or sing identical words at the same time. I call this kind of verbal activity Joint Speech.

This book seeks to introduce joint speech as an object of empirical study. In so doing, it uses the empirical study of joint speech to critically examine many assumptions underlying scientific work in those disciplines that deal with the living: biology, psychology, and the social sciences. In the first part the topic is introduced with several concrete examples. The principal characteristics of joint speech are discussed, and readers will be entirely familiar with many of these, as joint speech is an activity all language users partake in. Having established that there is a lot to examine and discuss, a big question then arises: Why has there been virtually no empirical study of this kind of behavior in the human sciences? The absence of a body of scientific work is very revealing, and it points to something of a blind spot. In pursuing this larger question, it is argued that there is an unresolved tension in play about how science should treat subjects, especially collective subjects, when it aspires to some, often unexamined, goal of objectivity. With this, large issues are clearly at stake. Chapter Three considers the way in which subjects and objects become entangled in the sciences of the living, and how joint speech may direct our attention to just those processes in which many of our collective identities are forged.

The middle part of the book then goes on to demonstrate that scientific inquiry of joint speech is both practical and profitable. Worked examples are

provided from the diverse domains of phonetics (the sounds of speech), from movement science (joint speech as synchronized action) and from cognitive neuroscience (where joint speech has some surprises in store). A special consideration is given to how we might think of joint speech within the study of human language more generally. In each case, the scientific work throws up questions about how to handle the divide between subjects of various kinds and their worlds.

Having established that there is a significant absence of scientific work, and then demonstrated that such work is possible and produces novel insights in many different ways, the final section turns back to the difficulties raised at the start. It argues that the study of joint speech might open a window to the empirical study of practices that ground human experience and identity. This may give us a useful and powerful way to approach the study of many kinds of important human activity, and the multiple overlapping collective identities that are thereby brought into being. This points towards a radical reconsideration of what scientific activity is, and how far its truths stretch. To support this ambitious venture, some suggestions are made about how one might appropriately develop a technical language suited to consideration of multiple perspectives, and how one might appropriately handle the relationship between subjects and their worlds.

The empirical phenomenon being discussed, joint speech, should give rich food for thought. As familiar as it is from every day life, whether one indulges in religious rituals, takes part in political protest, or merely chants merrily on the football terraces, it is relatively easy to show that joint speech is a very special kind of language use. It appears to be far older than writing and to have played a role, largely unexamined, in the foundation of all human societies. One might almost question whether it should be regarded as language, in a strict sense, or not, as many things we have come to expect of language, such as the sharp distinction between speakers and listeners, do not seem to apply to joint speech. Indeed, the study of joint speech necessarily leads us to a larger view of just what language is, and how the vocal activity of humans works and how it matters.

The questions raised for science itself may not be as familiar to some readers. In the course of the book, we will have to recognize some commitments within the life and human sciences that are frequently unacknowledged. It will be argued that objectivity in science is a complex issue, especially when subjects of various kinds are in play, as they necessarily are in the study of the living, including humans. One kind of subject in particular, the single autonomous individual or person, seems to carry a very great explanatory load when we are called to account for our behaviors

and activities. This stark individualism has been pointed out by many critics of modernity, especially of a Western, post-Enlightenment and largely Christian modernity. With the introduction of these highly politicized and polarizing adjectives, it is clear that any treatment of subjects and their objects will be potentially contentious, and that is as it should be. Throughout the book, we will encounter arguments in which the scientific debate is inextricably entangled with political concerns and the foundations of cultural identities. In the final two chapters, some recent perspectives from the enactive tradition in philosophy and science are introduced that may be of service in dealing with this kind of complexity. It is my hope that such debates will be enriched by being approached from a scientific point of view, and through the use of worked scientific examples.

This book makes the case that joint speech merits our attention, that we can learn much by considering how it should be accommodated within existing scientific practices, and that those practices might be expanded or augmented through what we learn. For the window opened by such study does not reveal only a strange form of speech. It provides an empirical access to practices by which order in our lives is created and sustained. It has the potential to lay bare the manner in which several sources of order, normally considered distinct, may overlap and become entangled. These include the regularity of the natural world (natural law), the authority of civil institutions (civil law) and the admonishments of tradition and religion. The study of joint speech is thus not only of interest to one or other academic discipline. It bears consequences for how we conceive of truth, what kind of truths may be arrived at within the scientific domain, and how the authority that comes with knowledge is negotiated politically.

## Acknowledgements

My thanks to several friends who contributed through discussion, and engagement with earlier drafts of this text. I have benefitted greatly from many conversations with Marek McGann, Tony Chemero, Stephen Cowley, Nick Campbell, and Jens Edlund. Alexandra Griesser, Paulo De Jesus, Roger Moore, Judit Varga, and Mark James also provided valuable feedback, and opened my eyes to many facets I had missed. The professionalism and vision of colleagues such as Sophie Scott, Kyle Jasmin, Stephen Connor and Daniel Richardson has greatly advanced my own thinking on joint speech. In the background stand the broad influences of Scott Kelso, David Vernon, and in particular, Victor Langheld, who runs the finest Indian Sculpture Garden in Ireland. Thanks also to the many colleagues and students who have

supported discussion on these topics through invited talks, panels, round tables, and the like. Bei Wang, Beatriz Raposo de Medeiros and Carlos Cornejo have played a special role here, as have ongoing discussions with Abeba Birhane, Austin Dwyer, Mark James, and John Francis Leader. Let the conversations continue!

### Image credits

Felice Varini kindly allowed reproduction of his works in Figure 17. Cover images from Wikicommons. From left to right, The Chant, Alex Proimos, CCA 2.0 Generic; Monks chanting at a wedding ceremony in Thailand, User Kborland, CCS-SA 3.0; Protesters chanting, Al Jazeera, CCA-SA 2.0 Generic. Further image credits in figure captions.

# Part I

# Introducing Joint Speech and the Subject

In this opening section, we introduce the topic of joint speech. In addition to definitions and labels, we illustrate joint speech through several concrete examples. Some readily apparent features of joint speech that give pause for thought are enumerated. These include the absence of an obvious distinction between speakers and listeners, the use of a great deal of repetition, the participatory nature of joint speaking, and the absence of any clear divide between speech and music in such activity. The big question is raised of why there is no science of joint speech.

In order to better understand joint speech on its own terms, it is necessary to recognize the importance of uttering. Joint speech is placed within a continuum extending from interior monologue at one extreme, through the to-and-fro of conversational speech, on to dialogical interaction with call and response in ritual and rite, and culminating in the earnest recitation of solemn texts such as a Credo or an oath of allegiance. The centrality of joint speech in highly valued cultural practices provides an incentive to stand well back from our topic, and to consider its contribution to the broad notion of Logos, understood as a generalized sense of order that finds expression in natural law, in civil law, and in the dictates of religion, tradition and habit.

Modern science has progressed from the dispassionate observation of the stars and planets to the more local and familiar territory of the biosphere in which organisms, singular and plural, from single cells to herds of wildebeests, co-exist in a dynamic negotiation of values and concerns grounded in diverse forms of embodiment and lifeworlds. Among the living, science is forced to confront multiplicity of perspectives, a task that must still be regarded as work in progress. In the domains of medicine, psychology, and the social sciences, we encounter different kinds of subjects. Joint speech serves to draw our attention to highly valued human practices that seem to ground collective identity and being for many communities. It challenges us to resist a view of agency and autonomy located only in the individual person, and to see ourselves as essentially and variously collective.

# Chapter 1

# Some Initial Observations

## Example 1: Reciting the rosary

IN THE BEGINNING . . . We are inside a small convent church in County Cork, Ireland. A group of nuns from the Poor Clare Colettine order are leading a recitation of the rosary, an extended form of ritualized prayer, once common in this country, but now slowly dying out. About 20 local lay people are also present. One of the nuns acts as a lead voice. She and her fellow nuns are located on one side of a dividing rail. Most of the nuns are kneeling at small individual benches, but the lead speaker stands in front of a microphone. Her voice is very soft. The lay group is on closely packed chairs facing forward. They are mostly middle-aged or elderly, with some few exceptions. Women outnumber men about 2 to 1. Most people, both nuns and lay, finger rosary beads, to help them keep track of the prayers. The beads, like the prayers, are organized into groups of 10, or decades, each bead corresponding to one recitation of the Hail Mary. Between groups of 10, some isolated prayers are uttered, and the whole suite of decades is bookended by additional prayers spoken together.

The prayer that is so often repeated has two halves. The first half is recited alone by the leader. The second half is a response, uttered by all present. The other prayers are likewise divided into calls and responses. The prosodic, or musical aspects to the voice, are quite pronounced. Each time through, the words are pronounced with the same slightly lilting melody, not quite sing-song, but not like conversational speech either. Everybody present is very familiar with the practice, and when everybody speaks together, there is a gentle acoustic blur, made all the more indistinct by the reverberant character of the room. Individual words or phrases are hard to

hear. Synchronization among participants is loose, allowing some voices to be tracked as individuals. The role of lead speaker is rotated at the beginning of each decade, and the decade is introduced with its title: The First Glorious Mystery: The Resurrection; The Second Glorious Mystery: The Ascension, and so forth. Decades are grouped into sets of 5—the joyful, sorrowful and glorious mysteries—so that the entire recitation has a complex hierarchical formal structure. Adding to the formal intricacy, prayers in successive decades show an alternation such that what was "call" in one decade is now "response" in another. Each part of each prayer is thus recited by everybody. This leads to a little uncertainty at the start of some mysteries, as not everybody seems to be entirely sure where they are in the structure. But the scaffolding of the collective is enough to establish unity and confidence very quickly.

## Example 2: Strife at the Al Aqsa mosque

A MORE VOLATILE SETTING. Leaving the relative calm of the nuns in Cork, we travel now to the plaza outside the Al-Aqsa mosque in Jerusalem. It is February, 2012. There has been friction, and there are wisps of tear gas in the air. Onlookers of many kinds are present, journalists with cameras, tourists, Arabs, Jews. Israeli riot police are also present, all dressed in black uniforms. They group together and an interface forms between civilians and police. There are scuffles. A man is grabbed by the police, and pulled back by his associates. Once he has been recovered, the police and the civilians hesitate. Suddenly a cry goes up from the civilian quarter: "Allahu akbar." This is the *takbir*, ubiquitous in the Arabic world, misunderstood in the West. It quickly becomes an insistent chant, with three beats stressed out of four: "a-LLAH-hu AK-BAR."

**The Takbir**

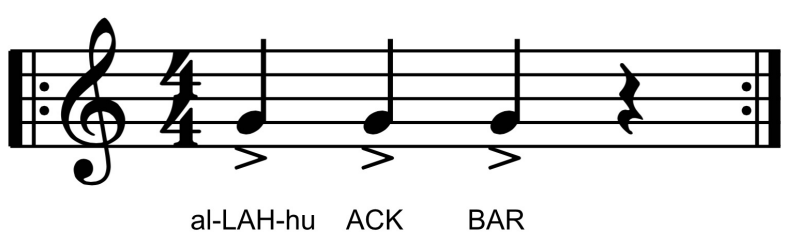


*Figure 1: "Allahu Akbar" as chanted during one violent conflict outside the Al Aqsa Mosque in Jerusalem.*

The phrase "Allahu Akbar" is famous, or better, infamous, in the West. Its meaning is often rendered as "God is great," or "God is greater." It has, unfortunately, become associated with the commission of violent acts, so that to many non-Muslim people, it has become a symbol of violence. Within the Islamic world, it is seen and used entirely differently. It is ubiquitous, so much so that it has it's own name, the *takbir*. If one calls out the name of the phrase, those around will respond with synchronized calls of "Allahu Akbar." The phrase is uttered both individually and collectively, sotto voce and out loud, and under many different kinds of circumstance, both delightful and horrific. It does not at all herald or signal violence, but rather functions as an injunction to the pious Muslim to recognize that no matter what he or she is experiencing or doing, it can be relativized by recognizing that God (Allah) is *greater*.

Now, suddenly, there are two groups, two collective entities, present, where previously there was only one. The riot police are already clearly marked as belonging together. They wear similar uniforms, shields, truncheons; they stand together and move en masse. The civilians were hitherto a colorful mixture, moving like pollen grains on the surface of water in Brownian motion, uncoordinated and various. But when the chant starts, they coalesce, and now there is a second group to stand in opposition to the police. For the brief period that the chant persists, we see protesters versus police. The chant unites, and a collective agent is temporarily brought into being.

## On observing

Prayer and protest make odd bedfellows. The gravitas of many forms of religious worship is far removed from the chaos of violent protest. Yet these two domains of human activity might be argued to share much in common. At a superficial level, we find the quasi-musical unison chanting of texts whose meaning is completely familiar to the participants. There are associated synchronized gestures (making the sign of the cross, fist-pumping), and there is an awful lot of repetition. These overt similarities might be dismissed as no more than the accidental use of specific forms of cultural practice, creating the illusion of commonality, and hiding fundamental differences in the purposes to which these forms are used. In one case we have a highly formalized ritual, designed and perfected by others, and repeated in more or less invariant form in many places and throughout centuries. In the other we have a highly contingent, improvised expression of frustration and anger directed against a very specific and tangible target. To link them by virtue of the relatively trivial characteristic of chant might carry no more weight than observing that clothes are worn at both events, and both happen in the afternoon.

But there are advantages to attending to superficial things. That which is on the surface can be observed without further ado. It can be observed by you and by me. That rather obvious characteristic has some benign consequences. It facilitates the path to consensus. If we both observe something, such that we are happy to use the same words to describe it, then we have a starting point for a discussion about the significance of what we have observed. This doesn't stop the merry work of disagreement thereafter, but it does provide a useful starting point. How many arguments go wrong because of the failure to agree on what it is that is under discussion?

The drive to achieve consensus, even limited and partial consensus, underlies science, politics, diplomacy, much of religion, and, in less formal mode, a lot of everyday conversation. In each case, the chances of achieving some kind of consensus are greatly increased if the discussants can demarcate a field of discourse, within which some basic foundations are agreed upon. In the discussion to follow, the subject matter will range over many disciplines, drawing on the everyday experience of familiar practices, but also linking these to scientific, philosophical, and historical arguments. In order to discipline the discussion, it might be wise to stay close to the surface of things, to lean heavily on observations in which we have some confidence, and to return to simple brass tacks whenever possible. In this manner, I hope both the reader and I may emerge unscathed.

The goals of the scientific enterprise will be of special importance in what follows. Science as I understand it starts with observation. In what follows, I will not treat science as if it were a unified enterprise, with each specialization capable of rational alignment and unification with each other. The kinds of observation and argument found in physics bear little relation to those found in biology, psychology, social science, economics, geography, or countless other fields, some of which purists may wish to exclude from the scientific family altogether. The small descriptions provided above of praying nuns and chanting protesters are observations of a sort. They are uncontrolled, to be sure, mere anecdotes, but we will treat them here as observations worth taking somewhat seriously despite this limitation. In coming to understand human practices, careful ethnographic observation is an essential point of departure.

In a rigorous formal framework, individual observations have a determinate form. If we are plotting star positions in the night sky using an agreed coordinate system, two numbers (and perhaps a time stamp) will suffice for each. But, for better or for worse, we will be very far from a rigorous formal framework here. In the journey before us, most observations will demand some contextual embedding. The anthropologist Clifford Geertz introduced the notion of "thick description," whereby any recorded observation is supported by as much contextual detail and elaboration as possible, thereby facilitating a rich form of interpretation, and holding back from a single, determinate, reading of the data (Geertz, 1973). Within anthropology, such elaborate descriptions become necessary when the objects of study–other people and their practices–are alien to the investigator. Geertz famously used this approach in describing cock fighting in Bali. In what follows, our situation will be somewhat similar, in that we will attempt, at times, to alienate ourselves from our everyday world, to view ourselves from the outside, making the familiar strange, in full knowledge that such an exercise is impossible. The very impossibility of such distance will be a topic we shall have to consider in its own right.

Under these circumstances, the ease with which joint speech can be observed will be a virtue, allowing us to calibrate our observations and to keep two feet firmly on the ground. I will often make use of extended descriptions as starting points and as anchors, tethering the more conceptual arguments to specific instances. In most cases, I will have in mind specific recordings of specific events, so that the details I note are not imaginary ornamentations, but documented features of at least one instance. An archive containing video recordings of specific instances of joint speech is being assembled, and

pointers to the particular examples described in this book will be available there.[1]

## Joint speech defined

The kind of speech to serve as an empirical anchor here is simple to define. **Joint speech is speech produced by two or more people who utter the same thing at the same time**. I intend "joint speech" as an umbrella term to cover many different kinds of speech produced in very different contexts, yet meeting this minimalist definition. The term "utter" is used, rather than "say," as joint speech will extend to include many utterances that lie between speech and song. We can identify a small number of subsidiary varieties:

- **Choral speech.** This is a genre of performance in which a group, such as a class of children, recites a set text for an audience. The audience will not infrequently be largely composed of relatives of the speakers. Choral speaking competitions are found in many countries. They seem to be particularly popular in Malaysia and South East Asia generally, but are also found in Ireland as a specialty in performance competitions (Feiseanna) involving solo recitation, music making, dance, and the like.

- **Chant.** The English word *chant* is ambiguous with respect to whether the vocal activity is considered to be speaking or singing. It can be used with equal applicability for the austere plainsong of a group of Benedictine monks or for the raucous hoots of a bunch of soccer fans. This ambiguity will serve us well in what is to come, as the domain of joint speech does not seem to support a categorical distinction between speech and music.

  **Synchronous speech.** This is a term I coined to specifically refer to joint speech produced under laboratory circumstances, in which speakers are speaking at the behest of a researcher, and not with their own purposes in mind. The texts employed are usually unfamiliar and of no special significance to the speakers.

Joint speech is found in a wide variety of circumstances, and the few specific varieties noted above do little to circumscribe the activity more

[1] The archive is available at jointspeech.ucd.ie

generally. But we can use the simplicity of the definition of joint speech in a singular fashion—to pick out discrete and diverse domains of behavior. We can use the definition of joint speech as a kind of lens with which to frame our observations. To see why this might be helpful, consider the task facing the poor anthropologist or behavioral scientist who wishes to study "ritual." Central examples of ritual are not hard to find—the Roman Catholic mass, or the coronation of a new monarch might provide obvious and plausible examples. But the borders of ritual are not easy to identify. Does your habit of folding your clothes and placing them on a chair before going to bed count as a ritual? What about tooth brushing? Is a football game a ritual? These are not substantive questions precisely because we lack an agreed definition of ritual. But if we use the definition of joint speech as a means of framing our observations, we find that its simplicity allows us to home in on several singularly important domains of human activity, without tripping ourselves up with such definitional niceties. If we ask "Where do people say the same thing at the same time?" we immediately pick out several familiar domains.

The largest by far is the domain of *prayer*, which for our purposes will generously extend to encompass both *liturgy* and *ritual*, while excluding silent prayer and the supplication of the individual, for these do not fit the empirical definition we are using. When we approach collective prayer, ritual and liturgy with joint speech in our sights, they seem to overlap so much that there is no profit in seeking to carve them into distinct provinces. With that, we are confronted with a widespread human activity that lies at the center of very diverse forms of order. In picking out one or other manner of speaking as prayer, or more specifically as collective prayer, we immediately reveal our own commitments and our own heritage, for that which might plausibly appear to me as prayer will be activity that bears some similarity to the practices I am familiar with, have grown up with, and that are on display around me. Coming, as I do, from an Irish background, the prototype of prayer might well be something like the recitation of the rosary described above. As we move further afield, it will be less clear what counts as prayer and what does not. We might encounter trance-like states induced by twirling, as in the *dhikr* of the Dervishes. This looks little like Roman Catholic prayer. Is it still the same phenomenon? What about the manual rotation of so-called "prayer wheels"' by visitors to Tibetan shrines? Or to take an extreme example that clearly illustrates the difficulty of escaping our own grounding, what are we to make of recent reports of chimpanzees in the Republic of Guinea in West Africa, who have been observed hooting and banging rocks against a specific tree, and piling up stones against it.

The purpose of the activity is not available to us as human observers. We cannot legitimately claim to understand chimpanzee activity that does not wear its intentions on its sleeve. Yet we find reports in the popular press announcing "Mysterious chimpanzee behaviour could be 'sacred rituals' and show that chimps believe in god" (Griffin, 2016). At this point, reason has left the building.

When we start with joint speech as our framing device, we will not delineate the domain of prayer to anybody's satisfaction. But we will be able to group observations together that belong together. We will be able to recognize commonality across many kinds of tradition, despite the fact that the traditions in question employ very different suites of concepts in describing themselves, their activities, and the world. In this sense joint speech studies can inform us about prayer, in a manner analogous to the use of blood pressure measurement to a cardiologist. The circulatory system is complex, and blood pressure measurement provides a very incomplete window into that domain, but it is a useful one, clearly relevant to the functional organization of the system, and it thus provides an empirical anchor to the heart doctor. This may be all the more important when we come to study behaviors and forms of organization that we do not understand, or that are not interpretable with our local suite of concepts, labels, and biases. By focusing on joint speech, we will be looking at activities that might be interpreted as religious ritual, rite, liturgy, or as prayer. The empirical focus provided by the definition of joint speech obviates the need to categorize the activities in too rigid a fashion.

The domain of *protest* also jumps out at us. We find the use of joint speech whenever people gather together to object, demand, or to revolt. There is variation from one situation to another, and such variation will be of interest to us; but it is by virtue of the unison chanting that the domain of protest is approached empirically, helping us to avoid thorny questions about what, exactly, counts as a protest.

A third domain that we must immediately recognize has, on the face of it, very little in common with either the gravitas of prayer or the urgency of protest. This is the use of chant among supporters of sports teams. Not every sport has a chanting tradition. It is rare in tennis, unheard of in snooker, but completely at home in soccer, ice hockey, baseball, American football, and several other sports. (Interestingly, rugby, which has very much in common with soccer, does not have a chanting tradition, though it does have its own remarkable singing tradition instead.) Despite the profound differences in the type of activity here, it will prove possible to identify

characteristics of sports chants that merit consideration along with other forms of joint speech practice.

Prayer, protest and sports chanting are the three biggest domains that joint speech picks out, but there are very many other areas in which joint speech arises, each time with its own domain-specific characteristics, but also with features that reveal commonalities where they might not be expected. In schools, teachers make use of joint speech in various ways, and as educational practices are not unified, so too there are diverse ways in which joint speech is used. Rote learning is common in classes with young children, and so recitation of multiplication tables can be found in every country. Getting children to speak together is also a simple way to marshal their attention, and skilled teachers of young children will use it as a means to gather and unite the children.

In many countries, religious education is an important part of basic learning, and chanting traditions are used here too, as a way to instill sacred texts indelibly. Madrassas throughout Asia and Africa use chanting as a means of learning the Koran. Hindu sacred texts have long been passed down and protected by chanting, and sutra chanting is part of the everyday experience of the young Buddhist monk. There is thus continuity in many cultures and traditions between the use of joint speech in education and later in rite and ritual.

We will encounter many forms of joint speech in what is to come. Often, these will be vignettes taken from everyday life, unremarkable under most circumstances. Everybody has experience of speaking in unison, even if many of us consciously choose not to join in this or that form. If the reader is alienated by people reciting the Nicene Creed, she may nevertheless assent to joining in with a chorus of Happy Birthday, a small ritual that also harnesses the collective, unified, voice. We may opt not to join in a pledge of allegiance to a secular authority, but when we join the circle of onlookers drawn to a street performer, and we hear an energetic appeal "Do you want to see a show?" we too will probably call out "yes" with one voice, and with that, we are no longer innocent passers-by, but are now part of a committed group of spectators with common focus.

Joint speech as a technology

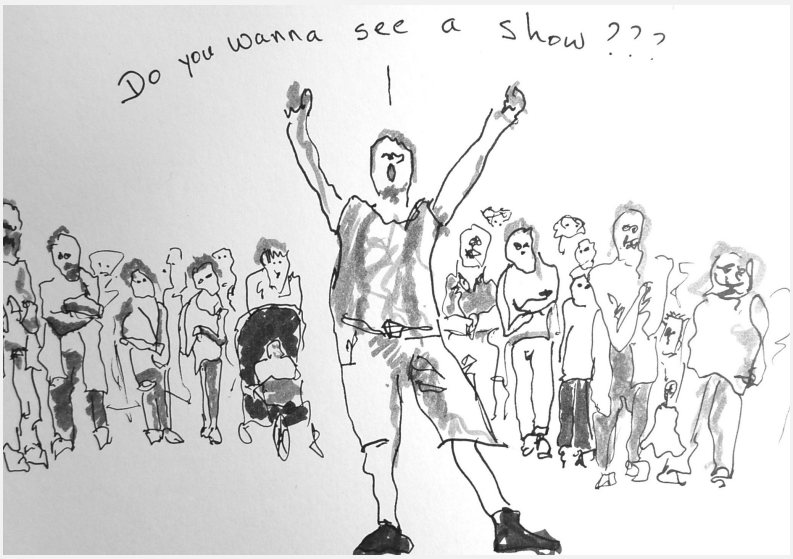


*Figure 2: Do you wanna see a show?*

Joint speech obviously serves many ends. As such, it also constitutes a technology that can be adapted to serve many kinds of goals. When the street performer gets the assembled crowd to shout back "yes," then he knows he has his audience, and they know that they are part of his show. This transient means of gathering attention is common in classrooms, where it assembles the unruly individuals and gives them a common focus. The informal shout of general assent that gathers the crowd on a shopping street has a counterpart in the collective prayer or oath often used to begin a formal meeting, a liturgy, or a ceremony, or in the domestic ritual of reciting Grace Before Meals. Once we speak together, we have common purpose.

## Some perplexing issues

So joint speech is ubiquitous, familiar, almost pedestrian. Yet it harbors many perplexities to be explored in what follows. When examined, these call out for interpretation, but not, I hasten to add, for explanation. An attempt to *explain* any of the following features would be to accommodate them within an agreed interpretive framework, to assimilate them to the

known and secure. I believe that joint speech will resist such a comfortable exercise, and will demand rather more of us. We start by scratching the surface of the phenomenon, expecting perhaps to uncover a novel genre, style, or cultural practice, one that might provide a pleasant distraction. But what we find is something vastly richer, and more challenging.

Here then are just a few features of joint speech that might give us pause for thought.

- In joint speech, there is no distinction between speaker and listener. This is rather obvious. All participants are engaged in something that is both, or neither of these. To say that prayers are addressed to a transcendent deity does not change anything. A Catholic God will, by now, be completely familiar with the text of the rosary. The protesters may be addressing a political establishment, but most of the time, those addressed are not present. There is no news value in what is said, but it must be said anyway.

- Repetition, repetition, and more repetition. Whether on the street, the terraces, or the church pew, repetition is a canonical feature of joint speech; so much so, that when it is absent, as in the collective swearing of an oath of allegiance, it is the absence itself that is noteworthy. The rosary beads of our opening scene are repeated in Christian Orthodox, Hindu, Moslem, Sikh, Jain, Bahá'i, and Buddhist practices of prayer.

- Performativity. The text of the rosary may be known, but that is irrelevant. It must be *uttered*. The necessity of actually uttering some phrases is well known from speech act theory (Austin, 1962). But within that framework, performatives such as "I dub thee a knight" or "I do" (at a wedding) are relatively rare. They accomplish something only under very specific circumstances, and that something is typically singular. Having married a person, it is not really possible to marry them again (at least not right away). Joint speech is performative, but in a rather different sense. We might speak of *enacting* rather than accomplishing. The vocabulary of enaction will be of use to us as we proceed.

- Musicality. Language (including speech) and music (including both chant and song) bear interesting relations to each other, and much effort has been spent in considering both their commonalities and differences. However when we begin to observe joint speech, any firm boundary between the two becomes invisible. In repetition, strong

> syllables are exaggerated; intervals are regularized; strong-weak alternations are enhanced; gestures such as fist pumping or clapping are frequently used. All this leads to an enhanced rhythm. Repetition tends to turn intonation patterns into melodies too. The prosody of joint speech, and the inextricable mingling of the fields of speech and music will be of great interest in what follows. Our definition of joint speech will have to extend to include unison singing, as well as speaking.

A final point to note about joint speech is that the practices we have identified, and others we could pick out by using joint speech to frame our observations, are all accorded a great deal of subjective significance by those who take part in them. The importance of prayer needs no argument. The urgency of protest is evident. The enthusiastic enactment of collective identity on the football terraces makes patent its charms for practitioners. Along with these canonical examples, we might note the solemnity accorded public group recitation of oaths of allegiance and fealty integrated into secular ceremonies. Clearly, this slightly odd business of saying the same thing at the same time is of some considerable importance, and there is ample motivation to seek to understand such practices in a manner that goes beyond the concerns of any single domain.

But now we come to the most perplexing feature of joint speech: it has not been made an object of concerted empirical inquiry at all. There is very little scientific work done on any aspect of joint speaking. As a topic in its own right, it seems to be invisible to those who study speech and language, and to students of human behavior. There are, of course, specialist and scholarly works that approach musical questions such as the history of plainsong and Gregorian chant, or that address liturgical niceties such as the respective roles of priests and congregations in mass. We will even find encyclopedic coverage of the rich and raucous world of football chants. There has been passing acknowledgement within ritual studies of the importance of collective speech and associated gestures in the specific context of religious ritual. There has been a small amount of documentation of protest chanting in specific situations, such as during the tragically misnamed Arab Spring of 2011 (Moghith, 2014). What is missing is the thematization of joint speech itself.

Joint speech is absent from linguistics. Speech is not the same thing as language, and we will have cause to consider features of speech that have no counterpart in language, conventionally defined. The scientific study of language has a history of focusing on the encoding and transmission of

messages, abstracting rapidly away from the messy business of shouting, whispering, cajoling, and imploring in specific communicative contexts, to the more rarified and untethered domains of syntax, semantics and phonology. This has made some aspects of the spoken voice less visible that they might have been. In recent decades, some researchers have begun to probe vocal behaviors more attentively, paying attention to those elements that do not make it onto the page in writing, including such neglected particles as uhms, ah-has, grunts, and silences. Where once linguists approached the sounds of speech to try to recognize the ghosts of letters and words, more recently phoneticians have begun to study the melody and rhythm of speech, the placement and duration of pauses, and the rich variation found in voice quality (Wennerstrom, 2001). The intimate relations between speech and gestures are starting to be traced (Wagner et al., 2014). Joint speech adds a great deal of additional material for study, though I would suggest that it must be studied on its own terms, and not merely as the voice of the individual, replicated many times over.

Joint speech is absent from behavioral and movement sciences. The vast majority of behavioral science looks at the actions of distinct singular persons. The form of any kind of skilled movement bears the signature of the individual. Though we may all reach a similar level of proficiency in writing, in walking, and in speaking, the manner in which we do so marks us out as unique and distinct, and the bodily patterns we exhibit as we perform similar tasks all speak of our individual identity, our accent, our uniqueness. Babies almost all learn to walk, but some do so by crawling, some by bum shuffling, and some by observing quietly. More recently, however, researchers have begun to pay keen attention to important aspects of behavior that are not captured by studying individuals. The shoaling of fish, flocking of birds, even the collective behavior of crowds in various physical environments are all now attracting attention (Vicsek and Zafeiris, 2012). But joint speech has so far been missing. Among the many questions we can ask, we might consider why joint speech has seemed to offer so little to observers of human behavior.

Joint speech is absent everywhere. A quick search on Google Scholar, the search engine of choice for scientific and academic publications broadly considered, reveals little. Part of the problem is terminological. In the absence of an established field of study, I have introduced the term Joint Speech, and I confess it was me also who introduced the term Synchronous Speech. Choral Speech is of greater antiquity. One might also look for Unison Speech. Combine all of these search terms, including both “speech” and “speaking” as variants, and I can find no more than about two or three

thousand works, many of them accidental catches. The greatest number of published works in the field address "choral speaking," and most of those belong to the slightly quaint field of elocution, whereby school children are taught to recite entertaining verses with polished pronunciation to the delight of appreciative parents.

By way of comparison, we might look for scholarly works on the slightly odd phenomenon of glossolalia, or speaking in tongues. This is a vocal behavior found within some evangelical religious traditions, and it is especially prevalent in Pentecostal congregations. It is, by all accounts, a very moving experience to utter syllables without a determinate meaning. Believers generally attribute the source of their utterances to the Holy Spirit. Google Scholar produces over eleven thousand scholarly publications that address this specific behavior. They include many works in the domain of cultural studies, comparative anthropology, ritual studies, and theology. There are neuroscientific studies, psychological studies, consideration of possible relations to psychopathologies, and even phonetic studies. There is work on glossolalia as a learned behavior and as a form of possession. Relations between glossolalia and personality types are explored. Glossolalia is a fine topic for research apparently. Yet when we do some mundane counting, it is clear that instances of speaking in tongues are clearly outnumbered by instances of joint speech, not by a hundred to one, or a thousand to one, but by literally billions to one, for joint speech seems to occur in all societies, in many different domains, and it is difficult to conceive of a vocal individual who has not spoken in unison with another at any point, while few of us, with the respectful exception of Pentecostal congregation members, will have spoken in tongues.

This then is the conundrum I wish to look into in depth. Why has joint speech remained invisible, despite the rather obvious facts that it has quite distinct characteristics, is accorded the greatest significance by practitioners, and is easy to observe. Is this neglect? Is there perhaps nothing to see in a group of people speaking in unison that cannot be found in the voices of individuals? In many respects, the apparent invisibility of joint speech may be its most interesting feature. As we shall see, it is neither difficult, nor unrewarding, to study joint speech as a scientist. There are many aspects worthy of consideration, and with minimal effort, we can generate scientific findings that speak to linguists, behavior and movement scientists, neuroscientists, social psychologists, and many more. But the absence of such work speaks of something even more important.

## Of subjects

The scientific tradition we value has its own history. It did not spring into the world fully fledged. If we squint a little, we might detect the modern scientific viewpoint coming into being first in the domain of astronomy, as the challenge of interpreting the motions of the planets, the moon and the sun from an earth-bound perspective was addressed, and ultimately solved. Objective science at its best has led us to learn how to think of our position on Earth, in a vast universe, most of which is alien, inanimate, and remote. As we move nearer to our terrestrial home, the kinds of studies we now understand as belonging to physics and chemistry were the next to emerge, and with Newton's magnificent construction of a theory of mechanical motion, it was possible to generalize from the movements of bodies close to hand all the way to the impersonal and imperious passage of the planets in the night sky. But the application of the scientific method to the goings on of living beings took a while. A scientific biology did not really appear until the beginning of the 19th Century. Scientific psychology emerged later in the same century; the application of scientific thinking to societies and groups of humans did not begin until the 20th Century. The social sciences might reasonably be considered to be still grappling with the task of finding basic concepts and methods that can garner widespread consensus.

As science has turned towards the living, and ultimately towards our own selves, the cool disinterested gaze of the observer has been challenged. With the study of the living, it becomes necessary to recognize and consider the perspectives of the living themselves. Living beings are subjects, not mere objects. They have *perspectives*. Things *matter* to them. The notion that science might provide a single God's eye view from nowhere, with no reference to value or to the perspective of a subject, now appears somewhat naive (Rorty, 1979; Nagel, 1989). The inestimable profit accrued from the application of the scientific method to inanimate matter makes it inevitable, even obligatory, that we should apply those same methods to the goings on of the living, and to human affairs. But where astronomy can get by just fine in an objective key, any science of the living must grapple with subjectivities of many kinds. As soon as we must appeal to any notion of *function*, then there is a subject lurking behind that appeal. The healthy living body is a subject for whom a beating heart can perform a function. Value-laden battles are fought literally under our skin as we speak of pathogens and anti-bodies, a view predicated upon the body as a domain for whom encounters with microbes are meaningful. We cry out for science to deliver results we can use in medicine, in education, in regulating our own conduct as individuals

and as groups, but this kind of science cannot be done as if the entities involved were mere objects and their interactions were free of value and significance, at least to the entities themselves.

The study of joint speech opens up new opportunities here. The activities we have surveyed above are steeped in values. The values are collective, and the subjects associated with them are collective subjects. Here we begin to see why there is such a remarkable absence of empirical work taking joint speech as its topic. We have not yet developed a language with which such collective values and collective subjects can be adequately addressed. Science in a simplistic objective key recognizes no subjects. The science of the living, or biology, in common with the psychological and the social sciences, has yet to find a way to rise to the challenge of integrating the competing and conflicting perspectives of multiple subjects, each with their own set of values. The old fashioned idea that science does not traffic in values has had its day. Science, when turned to goings on within the biosphere, within society, and by and for humans, has no option but to carefully negotiate the presence of many actors, many kinds of value, and multiple perspectives. This may be old hat to social scientists, but there is work to be done in establishing continuity between, and conversation among, the social sciences, the human sciences, the sciences of life, and the so-called hard sciences. As our gaze is drawn to those practices in which joint speech regularly occurs, we must confront the inextricable mingling of political and cultural concerns along with our scientific practices of observation and measurement.

So when we undertake the study of joint speech, we run the risk, and encounter the opportunity, of pushing empirical science where it currently does not go. We will need to acknowledge various kinds of subjects that arise through the collective activity of many kinds of groups. We will have to do so, in self-conscious awareness of our own limitations, of our own biases, our own grounding. The challenge joint speech presents is not that of an indecipherable object of study. As an object of study, joint speech is fascinating, rich, and ripe for the picking. It is its complement, the *subject* of such speaking, that will throw up the greatest challenges in what follows.

There is no way to address these topics without venturing waist high into contested territory. In so doing, my own shortcomings as an observer and interpreter will become apparent. Joining the dots across radically different disciplines is a challenge, and within the human sciences, it should be recognized as an unavoidable challenge. The strategy to be adopted here is to constantly return to the surface, to observe together specific examples, and to use these to anchor the discussion. Joint speech is an empirical locus that can do service to philosophers as well as to scientists, and can inform the

interested non-specialist too, providing footholds where discussion becomes difficult and fractious, and providing rich material to try out new ideas about how science should be done.

# Chapter 2

# Amen! The structures of joint speech

## Example 3: I need you to help me testify

THE REV. DR. MARCUS D. COSBY IS PREACHING. We are in the Wheeler Avenue Baptist Church in Houston, Texas. It is one of several Sunday services. The congregation fills the ample building. Attendees are almost all African-American, and the preaching lies squarely within a long and rich African-American evangelical tradition. His topic is Psalm 119, Verse 71, "It is good for me that I have been afflicted; that I might learn thy statutes." His words are lent color by an electronic keyboard whose timbre varies from piano to organ; the music is supportive, underscoring the energetic rhetoric of the preacher, and never detracting from the words themselves. At the start, the preaching is measured, and, because the topic is affliction, the words come slowly, with a lot of thought. Rev. Cosby is going to chew over the two lines of the psalm, allowing members of the congregation to recognize a message of hope in their personal affairs, encouraging reflection and the extraction of a personal meaning from the short text. Forty minutes later, the sermon has reached a musical finale; drums combine insistently with a pounding keyboard, the tempo of the preaching has escalated, the faithful are on their feet, there is clapping and stomping, the preacher is somewhat breathlessly repeating the word "Halleluiah" over and over, and we have completed a journey that this meager text cannot begin to convey. Let us simply listen to a little excerpt, taken from around 27 minutes into the sermon. Contributions from the congregations are signaled

by bold font within square brackets, and they are, strictly speaking, the joint speech here.

Now Child of God, this ought to get you happy right here
Because if you don't know anything else about God, you ought
to know
Before you leave this sanctuary today, that your God, my God,
our God,
is faithful [**yeah**]
That that mean he steadfast [**yeah**]
That means he's dependable [**yeah**]
That means we can count on him [**yeah**]
I need two or three folk in the sanctuary
Two or three folk in the worship center
To help me witness to somebody around ya
Who may be on the fence about the faithfulness of our God
I need you to help me testify
Our God is faithful [**yeah**]
If you read Jeremiah's version of Lamentations, Verse Chapter
3 Verse 22
You will hear Jeremiah say it is of the Lord's mercies [**yeah**]
That we are not consumed for his compassion fails not
They are new every morning
Great. Is. Thy. [**Name.**]
Y'all know something about the faithfulness of God ...

Even in transcription, something of the rhythm of this performance shines through. The Reverend Cosby is highly skilled, his words are lively and engrossing; he positively invites the congregation to take part in his performance, and they do. They rise up out of their seats, they clap, and they shout back at just the right points, points that Dr. Cosby signals clearly. At one point, he even says "Don't you miss your shout cue, I said ..." Given our focus, it is of course the audience contributions that draw our attention. But it would do violence to the nature of the event we are witnessing if we were to artificially separate the words of the preacher from those of the congregation. The sermon unfolds dialogically, each phrase a response to what went before, and if the preacher is the main source of the message, the congregation most certainly emerge with the sense of having helped to deliver it and having contributed to its unfolding.

## Generalized assent and the co-construction of meaning

As we turn towards the role of the voice in collectives, whether prayer, protest, sports, or elsewhere, we need to develop some way of talking coherently about the sentiments and aspirations of such groups. A recent fashion for speaking of the "wisdom of crowds" stands in opposition to a more longstanding distrust of the voice of the herd. Mob justice is not typically regarded as sophisticated, considered, or rational. If they are coming after you with pitchforks, run, don't argue! Claims of the "wisdom" of crowds have their origin in the dismal science of economics, where they are used to justify a specific view of markets. This is hardly reassuring. When we look at the three canonical domains of joint speech—prayer/ritual, protest and the chanting of sports fans—we might likewise be forgiven for thinking that this is not the place to find the most nuanced and refined use of language. The most common rituals are found in the service of religious observance, and there is a widespread and unfortunate perception among many that religion must be at odds with rationality. Protest chants are often improvised and abbreviated, and it is not in the chants themselves that we expect to find argument, discussion, and consideration of the issues involved. Meanwhile on the football terraces, the massed voices of the fans are often found to be either simply proclaiming the identity of the group, or hurling insults at the opposing crowd. Could it be that joint speech represents a degenerate form of language, divorced from the rarified fields of reason and civilization? Is joint speech the inane and shouty little brother of dialectic and reasoned argument? On this view, the phenomena we are looking at here would be little more than grunts with added syntactic sugar, and would be regarded as marginal, or completely irrelevant, in consideration of the role of language in the origin of humanity.

But let us look a little more closely. We will focus for now on the two domains of prayer and protest, with a keen ear for forms and themes common to both. In both domains, chants very often take the form of call-and-response. The call is sometimes from a sub-group, but more commonly from a single leader. In the Catholic rosary recitation the role of leader was passed around in an egalitarian manner, from one nun to the next, as the ritual progressed through the decades. In Rev. Cosby's preaching, there is no doubt at all about who is playing the role of leader, but the frequent interjection of calls of "yeah" constitutes a group response. The division of responsibility between caller and crowd provides an opportunity

for constructing a far more sophisticated kind of linguistic behavior than the hoots of a mob. Both prayer and protest make frequent use of what we might call generalised assent. In churches of the Abrahamic faiths, this is frequently voiced as "amen," "amin," "aymen" or similar. At a protest rally, it might be calls of "right on!" or in either, it might simply be "yeah." In calling out a token of assent in this fashion, the leader and the crowd are engaged in a single collective act, a dialogical uttering, back and forth, with common purpose. Many variants are possible on this common theme. At some points within a liturgical ritual, this may be further underscored by making the collective assent a two part formula: Priest (after a long and complex prayer of supplication): "Lord, hear us," All: "Lord graciously hear us," or by calling out "Can I get an Amen?" At a rally we might hear Leader (after outlining a plan of action): "Are you with me?" All: "Yeah!" Rev. Cosby makes use of an effective means of integrating the voice of the crowd into his improvised stream of words when he slows down, turning each word into its own phrase: "Great. Is. Thy." and the crowd knows to seize the opportunity to join in with "Name."

If we go to a religious service, we will probably encounter a variety of call-and-response structures that seem to serve several different functions. Some of the most canonical of the prayers will be spoken in a single unified chorus. For this to work, all the participants must be familiar with the words, and so this is typical of the central prayers that are repeated very often in the lives of the adherents. In a Protestant or Catholic service, for example, there will be an "Our Father," which is understood as a model prayer, and there will be some form of *credo*, an explicit expression of shared belief. *Chorusing*, as we might call such extended unison speech, is most often appropriate for specific prayers that are well-rehearsed, memorized, and canonical. The limitations of memory suggest that for most collectives, there will be no more than a fairly small set of such central verses that can be spoken in unison, without either accompanying written texts or the support of a leader.

The texts of the liturgy are crafted with intense attention to detail and propositional subtlety, in which statements are made that express central theological tenets. The *filioque* illustrates the importance of even a small phrase: The thousand-year old split between Western and Eastern Christian churches is indexed by the inclusion or exclusion, respectively, of the phrase "and the Son" when describing the third element of the Trinity, the Holy Spirit, within the Credo. But while the text of the Credo provides a set of core propositions central to the specific church or faith, there is a great deal more content to a full religious rite than any congregation could be expected to memorize. For example, the precise formula of words that is

used around the consecration of the host in a Catholic mass is chock full of careful phrasing that has been developed over centuries of debate, crafted to distinguish between the specific form of words that finds approval, and those very similar forms that pick out heresies, deviant cults, and schisms. This is too fine a filigree to be trusted to the crowd, but the crowd can partake of the collective expression through the careful placement of "amen" throughout the liturgy. In this manner, joint speech is capable of supporting a great deal of propositional sophistication, making fine distinctions, and expressing very precise concepts.

A similar relation between form and content is found at political rallies, where the collective vocalization is asymmetric. A leader or figurehead will enunciate arguments, points of policy, or the precise nature of the grievance at hand. Frequently, the leader is the only person present with a microphone. The crowd partakes in the expression through the collective voicing of assent, frequently punctuated by fists punching the air. It is neither expected, nor necessary, that the crowd speak a nuanced political message; the back and forth between the leader's exposition and the crowd's assent enables all present to partake of the collective uttering.

## Example 4: The human microphone

LET US DROP IN ON THE 2011 OCCUPY WALL STREET PROTESTS. A large crowd has gathered in the expectation of a series of public speakers. The crowd is enthusiastic, but not tense. Most attendees are overtly concerned with matters related to social equality, leading to a convivial and cooperative tone. However the New York Police Department has issued a ruling forbidding the use of amplification during the protest, which necessitates a creative alternative: The human microphone. In order to broadcast a spoken message through the crowd, a source speaker begins with the phrase "Mic check." With that, the phrase is repeated in unison by a broad circle of participants within hearing. This increases the volume, and the phrase is transferred to a larger outer circle, where it is repeated once again. Once this mechanism is in place, the ban on amplification can now be finessed completely. The speaker splits her message up into short phrases, and each phrase is passed from speaker to inner circle and on to the outer circle, growing in volume with each repetition. It is slow but efficient, and in this manner one public figure after another manages to broadcast their speech despite the ban.

Speaking and commitment

*Figure 3: Left: Fingers crossed while swearing an oath. Right: Hand signals used by the Occupy movement.*

To speak is to be responsible for one's words. The technological innovation introduced by the Human Microphone, e.g. during the Occupy Wall St. protests, acknowledges this by providing a parallel channel for expressing one's commitment to the words being spoken, while allowing the person to function as a mechanical cog in an amplification machine. This is a straightforward elaboration of the well-known ruse of crossing one's fingers while making a promise, such as an oath, on the understanding that the manual gesture relieves one of responsibility for the words uttered.

Each participant in the human microphone has consented to take part as a cog in a machine, mindlessly amplifying the message, irrespective of its content. But what happens if the speaker says something to which an individual participant objects? The speeches are polemical, and they range over topics that are of concern to all those present. There is diversity of opinion, but the human microphone only works if the individuals who make it up are willing to speak the words of the speaker, even though they do not necessarily agree with them. In order to deal with this potential conflict between individual perspectives and the collective voice, hand signals have

been agreed upon. If the individual speaker wishes to express disagreement with the sentiments she herself is voicing, she can hold her arms aloft, and wiggle her downward pointing fingers. The gesture goes by the name of "down twinkles," or, as the microphone finds application in other cities, as "squid fingers." "Up twinkles," with upturned hands are used to signal agreement. This provision of a parallel channel to signal affect ensures that nobody appears to be responsible for speaking words they do not stand behind. At some of the protests, a richer set of hand signals is used. Some of these can be traced back to conventions used among Quakers; others to American Sign Language; some appear to be novel. The human microphone illustrates how joint speech is, among other things, a form of technology that can be creatively put to new ends. Nobody needs instruction in speaking together—though the hand signals do need to be taught explicitly.

## Uttering and commitment

Contemporary Ireland is full of so-called "lapsed Catholics." These are people who nominally belong to the Roman Catholic Church by virtue of birth and baptism, but who do not attend regular religious observances, and who do not consider themselves believers. The machinations of society nevertheless conspire to ensure that such folk will find themselves attending the celebration of the mass to participate in weddings and funerals, both of which have import beyond the domain of the church. At such events, most attendees will blend in with the crowd by kneeling, standing, and sitting along with everybody else. Taking part in the joint prayers is rather more fraught though. Some choose to mumble, which avoids the embarrassment of speaking words they do not condone, yet allows them to blend in with the overall diffuse acoustic buzz. Others will voice, or mouth, the words of most prayers. But joining in with the Credo, which asserts "I believe in one God, Father almighty, Maker of heaven and earth. . . " is, for many, a step too far.

Many people are wary too of joining in a protest chant. It is one thing to feel a broad sense of alignment with the concerns and aspirations of an energized crowd. It is quite another thing to join in the chant. One might attend a protest out of curiosity, but if one joins in the chanting, one is *participating* in no uncertain manner. At political protests, it may be the case that a series of chants is used, often selected and orchestrated by one individual with a bullhorn. Changing the chants after a few dozen repetitions is one way to keep the crowd alert and to infuse the affair with

a sense of entertainment. But this can lead to involuntary or unanticipated participation in chanting that is poorly aligned with one's own view, or worse. Political demands can morph into ad hominem attacks. Unless you are the person with the bullhorn, you do not control the words. Participation is a voluntary surrendering of autonomy. It is not a mechanical act.

Uttering matters. To give voice to a proposition is to represent it, to vouch for it, and to be responsible for it. In a literate society, awash with free-floating texts whose connections to their creators are unknown, this may not be obvious. But we are all keenly aware of the difference between a written apology and a spoken one. When we give our word earnestly, that word must be spoken. And who hasn't use the rejoinder "But you said ..." in an argument? In investigating joint speech, one thread to be followed is the manner in which speaking gives rise to commitments. The human microphone includes an explicit acknowledgement of the way in which obligations may arise by uttering, and provides a parallel manual channel to allow the voice and the generative subject behind the voice to be artificially sundered. Ordinarily, the speaker bears responsibility for the act of uttering and can be called to account for what is said. This sounds a little quaint in 2017, surrounded as we are by disembodied voices. Siri chirps inanely from the iPhone; someone who is nobody tells us to mind the gap and that the next train is delayed, and the answering machine reassures us that our call is important to a faceless corporation who is, finally, nobody. But we remain very sensitive to the responsibility of the real flesh-and-blood speaker, and we hold her accountable for her words.

Generalized assent provides a means by which a large number of people can share responsibility for a message. It might be helpful to consider a dialogical continuum, as in the sketch below (Figure 4). On the left is the back-and-forth of a two-party conversation among equals. Each utterance in a conversation builds on the preceding utterance, which itself is a response to its predecessor, and so on ad infinitum. With each utterance, the shared ground of the conversational participants, hinted at by the overlapping colored areas, shifts somewhat. New points of agreement or disagreement are uncovered, and the space of possible continuations is altered. The conversation is dynamic, and the shared perspective of the dyad is constantly shifting. When we can speak of a shared perspective, we are also invoking the notion of a common ground, supporting the conversation. If there is too little common ground, the conversation will fail.

Conversation, common ground, and the Credo

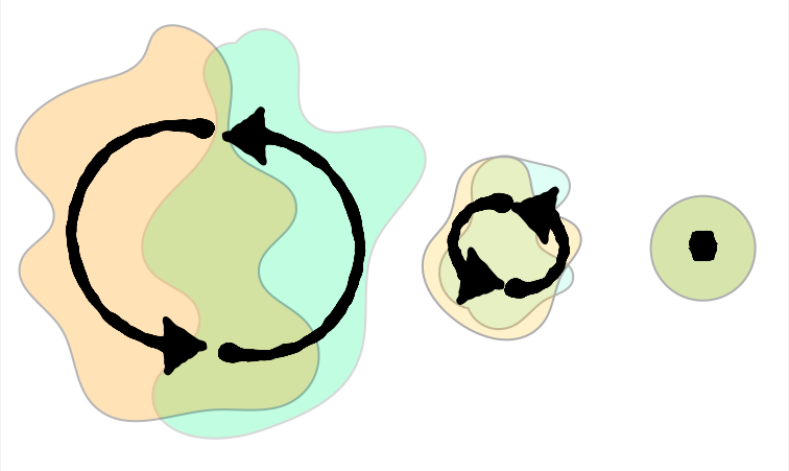

*Figure 4: Left: A conversation considered as a dynamic back-and-forth, in which the common ground (overlap) is constantly shifting. Middle: In liturgy, the back-and-forth is much more highly constrained, and the amount of common ground is greatly increased. Right: In collective recitation of a credo, common ground is maximized and there is no dialogical negotiation.*

In considering joint speech, the poorly articulated, but rich notion of common ground cannot be avoided. If a conversational exchange were just a matter of trading propositions (sentences that are either true or false) then the common ground would be the set of agreed propositions, much like a set of axioms in a mathematical deduction. But human conversational exchange is vastly richer than that, and it is probably better to consider common ground as something like a shared perspective, from which threats, opportunities, and uncertainties appear in a similar light to both parties. As the conversation progresses, the degree to which such a perspective is shared will change continuously. Joint speech seems to be part of a suite of practices that align those who partake of them in a similar manner. The propositions of a credo are not negotiated by those who speak the words. They represent, instead, a shared origin.

In the center we have a much tighter back-and-forth, more representative of the link between preacher and congregation in the opening example of this chapter. The feedback of the cries of "yeah" does not divert the Rev. Cosby

from his message, but it does nudge him in encouragement, invigorating the following phrases. Here the shared ground shifts less. There is little opportunity for divergence, although a less enthusiastic response from the crowd might conceivably elicit a slight modification to the ensuing phrases. Finally, on the right, in the joint chorusing of the credo, there is a unification of the concerns or perspectives of the participants. For this brief while, speakers speak with common purpose and alignment. Participants are well aware of the consequences of joining in, of synchronizing, and fusing, albeit temporarily, with their co-congregants. Of course the fusion arrived at in such moments is, itself, only one point in a continuing flow, from which the individuals will ultimately re-emerge. Utterances are necessarily fleeting.

## Example 5: Swearing an oath of allegiance

THE CONVENTION CENTER IS PACKED. Several thousand people of diverse nationalities are convened in order to swear an oath of allegiance to the country they have chosen to live in, in this case Ireland. The stage is occupied by officials, including the Minister for Justice, along with some musicians who will be silent for the swearing of the oath. There are also many of the formal trappings of the state. A pair of soldiers in uniform holds a large national flag. The crucial swearing of the oath is about to happen, for which everyone is asked to stand, in a manner entirely analogous to behavior in Sunday church at some of the more solemn moments. But first some instructions are required. Not only is it necessary that everyone swear the oath, but the words that are sworn must serve to bring the individual, who has a unique name and address, into the collective fold. And so they are instructed that after the first word "I" there will be a pause in which they are to speak their own name. Then comes the word "of" spoken together, after which each person is to speak their own address. The rest of the words are spoken in common:

> I, <your name>, of <your address>, having applied to the Minister for Justice and Equality for a certificate of naturalization, hereby solemnly declare my fidelity to the Irish nation and my loyalty to the state. I undertake to faithfully observe the laws of the state and to respect its democratic values.

The ripple of voices is quite indistinct. In particular, nobody could hope to pick out the name or address of any specific individual. But intelligibility

is not the goal here. This sequence of words must be spoken exactly as written, and embellished with the trappings of the individual.

## Mirroring

The form of call and response seen in the naturalization ceremony we might call *mirroring*, to distinguish it from the *chorusing* of the credo. In mirroring, the leader utters a short phrase, after which everybody repeats the exact same words. This does not require anybody to memorize more than a few words at a time, and only for the few seconds required to speak them. It is thus admirably suited to the performative uttering of complex individual texts that need be spoken once, and once only. This was also the form employed in the human microphone. The resemblance of the above transcription to a written administrative form in which personal details are entered is not accidental. Mirroring seems to be the hallmark of the kind of performative ritual that needs to be done once, after which the participants are agreed to have changed, from non-nationals to nationals, from laity to priest, from citizen to king. In these rather particular circumstances, there is no place for repetition. The act is strictly instrumental, and there is an overt purpose for the business of coming together. These are performatives of the kind discussed by Austin (1962).

Under these circumstances, the prosody of the speech is strictly speech-like. Musicality does not arise, and the musicians on the stage remain silent. Oaths of allegiance can also be repeated, as is highly conventional in American public schools at the start of the day. This is no longer simply instrumental, in the above sense, and so when that happens, the prosody is altered, and singsong elements become part of the performance. But the basic form found in the Irish convention center is echoed in many countries, and even in the videos issuing from Islamic rebel groups pledging *bay'ah* (allegiance) to the Caliph of the Islamic State in Iraq and Syria. Musical prosody and repetition thus serve to distinguish most forms of joint speech from the amusical occasions in which the purposes of the assembled are purely instrumental, where we find speech-like mirroring.

## Logos: Founding an order

In the previous example we may get a glimpse of the means by which a prevailing order is established. There is much more to a nation state than any set of ceremonies, but part of the means by which the state enacts its

own identity is through the rituals of its various institutions. Law courts, naturalization ceremonies, the daily business of parliament, all have their formal rituals whose conduct is required, typically in public, to bring into being the offices, laws and institutions of state. Formality is reserved for the more important meetings and processes. It would be disastrous if every sub-committee meeting required the attendance of a state mace bearer with white gloves before the participants could get down to work. But convening parliament, passing a law, swearing in a citizen, these rituals are formal, their structures and texts are laid down and are required to be performed in accordance with specific constraints. And with this formality, the particular form of order that is a nation state enacts its being.

There are many kinds of order that influence, shape and dictate our activities. We will concern ourselves with the relation among three: The stars and planets move in the sky in accordance with what we might call *natural law*. Its order is widely considered inviolable, and there is not much point in petitioning any earthly body to modify say the path of Jupiter. The *civil laws* of state and the associated institutions constitute a rather different kind of order. They are less fixed, but they have a rigidity that ensures they are not changed easily, or arbitrarily. They also support judgments that are determinate, and founded on matters of evidentially supported fact. Religious systems establish such orders too, in part through the regular performance of rituals by adherents who, through their very participation, both continue the enactment of the order and subjugate themselves to it. Thus along with natural and civil law, we will have to discuss the *injunctions of religion*.

In contemporary Ireland, in 2017 as this is being written, most of my fellow citizens would see these three kinds of order as strictly distinct. Natural law is inviolable, civil law is inevitable, and religious stricture is entirely optional. Natural law pre-exists any particular social order, and is considered entirely distinct from it. Civil law is constructed laboriously, through means agreed in advance, and always capable of modification. The dictates of religion and tradition are viewed as belonging to a different kind of domain altogether, and one that may be rejected in principle. But this partitioning of the several kinds of order that constrain, guide, and impel us is a rather local phenomenon. It is certainly not the case that a division between church and state is accepted universally–quite the opposite. The division of those two kinds of authority is, and has always been, highly contested. The history of any religious tradition is inextricably intertwined with the dynamics of power and authority. The very notion that one could separate church and

state is fairly recent, dating back only to about 1800, and it is by no means a universally shared aspiration.

The division between such human orders and natural law is also not quite as clear as some may think. It is undoubtedly true that we have no say whatsoever in setting the magnitude of the gravitational force exerted by the planet or the timing of an eclipse. But the domain of science extends to the domain of the living, and once we treat of the living, we are faced, whether we like it or not, with the perspectives and values of living beings who are subjects and not mere objects. In what follows, I will argue that contemporary science seems to exhibit a commitment to a specific kind of subject, as described by scientific psychology, but that this subject is not a natural object, and treating it as a natural object has the unfortunate result that we fail to recognize, let alone respect, many other kinds of equally real subjectivities. This is contentious stuff, so let us proceed cautiously. We will be reviewing several kinds of science as we go along, and we will need to be sensitive to the difference between science conducted in a strictly objective key, e.g. in monitoring the stars, and science that must, of necessity, be more cautious in its pronouncements, as it addresses the conflicting concerns of many kinds of subjects.

One might object that I have skirted around, or ignored the roles of tradition and culture in organizing human affairs. Religions may pronounce on matters of fact, and they may even develop their own bespoke legal systems, whereas cultural trends and traditional practices are rarely formalized. However for the purposes we envisage here, tradition, culture, and religion may usefully be grouped together to point to a very broad range of influences, distinct from "nature" and "civil state," that clearly influences and organizes our activities. I will interpret the term "religious" very broadly, to allow it to extend to non-theistic traditions such as Buddhism and Daoism, which juggle with both philosophical and soteriological issues, i.e. which seek both to pronounce with authority on matters of "fact," but also to provide direct guidance in the living of one's life. The thrust of the argument herein will not be to crisply define any such influences on our affairs, but rather to use the study of joint speech to show how fuzzy such distinctions are in the first place. In the study of joint speech, we will encounter various ways in which order is introduced and maintained, in which the unshakable foundations of diverse forms of being are rooted, and we will be challenged to acknowledge our own grounding too.

There is an old Greek term, *logos*, beloved of the pre-Socratic philosopher Heraclitus, which may be of use to us in what follows. Logos is sometimes translated as "word," but that is an anemic rendering. It is the word of

the beginning of the Gospel of St. John ( *"In the beginning was the Word"*), and it features as a central foundational term in many philosophical and religious traditions. We will be attracted to it in this work, as it captures the notion of an order, whether natural, civil, or religious. Logos may also be used as a generative term, used to signify the bringing into being by speaking. For now this may be merely evocative, but as we keep practices of joint speech squarely in our sights, it is with the establishment of order, and the bringing into being of order through the act of speaking that we are concerned. As we explore the absence of any science of joint speech, and the related question of how science treats of the *subject* (as opposed to the *object*), we will constantly run into boundary issues between the three kinds of order.

# Chapter 3

# Science(s) and Subjects

## Example 6: U. S. A.!!!!

THERE IS GREAT EXCITEMENT IN A ROOM IN AMERICA, but for now our TV screens just show a view of a floating barge somewhere in the Pacific ocean. We, both viewers and technicians, are waiting with baited breath for the latest in a long series of attempts to achieve the engineering feat of landing a part of a space rocket, the first stage of the Falcon 9 vehicle, on this barge. This is a task of unimaginable complexity. A large cylinder has propelled itself to the upper reaches of the atmosphere, and now it must return in a controlled fashion, landing vertically and precisely on a bobbing barge. It seems impossible. Several previous attempts failed, and there is no strong optimism that this will go any better. This stage of the mission is regarded as experimental, as work in progress. Nevertheless, it succeeds. The rocket enters from the top right of the picture, flames coming from its rear. Magically, improbably, it hovers in mid air and then touches down on the barge, remaining upright the whole time. The crowd erupts, and the TV screen now splits to show both the barge and rocket on one side, and the large crowd of scientists, technicians and engineers who have worked on the project, on the other. Suddenly a chant breaks out in the control room: "U.S.A., U.S.A., U.S.A." There is celebration going on here, but there is something more.

The project has been completed at a critical time in the American space program. The space shuttle has been decommissioned, leaving no access to the International Space Station except through the Russians, and things have not been going very well there. Today's success bodes well for the re-establishment of a regular supply line through American firms, the govern-

ment having decided it was no longer in the business of building spaceships. Of course excitable Americans are prone to chant "U.S.A." at just about any plot twist, even if it is only a successful pitch in a baseball game. But it is certainly not accidental that the national chant surfaces in this context.

To many, this feat we have just witnessed will clearly demonstrate the absolute indubitable solidity of the scientific worldview, trumping all other explanatory frameworks by virtue of its clear demonstrable results, not least in the exciting domain of rocketeering. But is that an appropriate way to describe what we just witnessed? Most of the science underpinning this feat was not done by these people, or by anybody alive. A lot of it was done by Sir Isaac Newton himself, in the development of a theory of mechanics. What we have witnessed is a fantastic engineering achievement, but there is precious little in it that can inform and improve our understanding of our world or our selves. Meanwhile there is a remarkable group behavior taking place that the science we currently have is mute about. We lack any understanding of what this chanting is, why or how it happens, and how it is related to both the individuals, and the many collectives they live in and among.

## Individuals and collectives

Joint speech presents us with a conundrum. We have grown accustomed to viewing language as a game of message passing. I have a secret thought, and I pass it to you in encoded form for your personal use. Our roles in this exchange are clearly defined and distinct. On the one hand we have an activity carried out by individuals; on the other we observe cumulative results that affect the collective, up to and including the whole species, allowing us to cooperate and solve shared challenges. In our everyday thinking about language and its effects we routinely maintain a clear separation between the collective and the individuals that pass the messages.

But now we turn our attention to joint speech, which certainly appears to be language of some form. It is found in every human community; it seems to be as old as humanity itself (of which more later); it arises during just those activities that seem to found collectivities of all sorts, in the rituals, assemblies and ceremonies that underlie and enact the social order, and yet it defies the conventional picture of the message passing game, and challenges us to examine how we distinguish between individual and group concerns. As we do this, we are led to consider the difference between two

types of descriptive and explanatory accounts we might seek to provide of ourselves.

On one side we have the individual person, to whom we attribute a singular mind, acting as an autonomous agent, whose behavior we interpret through the prism of such psychological notions as (individual) memory, perception, attention, and emotion. It is hard to do without such notions in describing our actions and their causes. On the other, there is the group or collective, who may also be granted agentive qualities, but for whom such terms are available only in a weaker, metaphorical form, if at all. When we speak of collective attention, or shared memory, it is conventionally assumed that we are borrowing terms that have a more literal home in the description of an individual and their personal inner constitution. In this way, the social sciences defer to the claims of psychology.

But the social sciences are even younger than the psychological sciences, and may be reasonably said to be still looking for an appropriate set of foundational concepts that can do duty. The founders of sociology, such as Émile Durkheim, insisted without ambiguity that there were causal forces at work at the collective level that could not be simply reduced to aggregations of individual contributions. For example, Durkheim was greatly impressed by the unity of experience that happens during religious rituals. At such moments, there is a profound coming together, evident not least in the synchronization of speech and gesture, that unites those taking part. This kind of group experience is difficult to put into words, and Durkheim's notion of "collective effervescence" has not become a term of art, but it demands acknowledgement, and Durkheim's concern that it be recognized as a genuine causal force in its own terms, and not merely as a shorthand for very many individual experiences, resonates with the challenge we meet as we try to see joint speech as language and as essentially collective at the same time.

As we bring joint speech into our sights as an object of study, we are obliged to consider several kinds of scientific explanation that are normally kept at a careful distance from one another. The uneasy distribution of explanatory competence between the disciplines of psychology and sociology arises as an urgent consideration. These two branches of the so-called "softer" sciences need to be considered together, and the very idea that there is a neat separation between their respective domains is called into question as we square up to joint speech. To this heady mix, we must then consider the language sciences, which have a long tradition of existing at some remove from most other scientific fields, with few shared concepts, methods or explanatory principles. Heinz Von Foerster phrased his "Theorem Number Two" thus: "The hard sciences are successful because they deal with the

soft problems; the soft sciences are struggling because they deal with the hard problems" (Von Foerster, 2003, p. 191). The soft sciences he refers to include all those that provide accounts of human affairs, and thus includes psychology, sociology, linguistics and more. Boundary disputes among these activities are very far from settled, and their respective explanatory remits remain to be sorted out. It is to such vexed concerns we are drawn as we approach joint speech, in the hope that by careful alignment of the various parties involved, science itself may be nudged to better address human concerns. To do this, we will have to understand how these "soft" sciences stand in relation to the more established landmarks of both physics and biology.

## The Mechanical Universe: Science in a simply objective key

Modern science may reasonably be said to have come into being between about 1500 and 1700 CE. This was the period in which the very idea of natural law emerged, supported in large part by the magnificent mechanical theory of Isaac Newton. Copernicus and Galileo and many others had paved the way by transforming the best available account of space. The pre-scientific cosmologies of Europe employed many elaborate schemes to situate human affairs at the center of the universe, consigning the stars to a uniform outer sphere, with the planets and other bodies of the solar system artfully arranged on intermediate structures. With the displacement of the Earth to a path now understood to be an orbit around the sun, on par with that of other planets, space became homogenized, transformed into a neutral container, within which things happened irrespective of whether that was on the surface of the moon, or in the familiar environment of the kitchen. Newton's laws of motion appeared to invest this view of space with a reassuring regularity. Three small equations sufficed to scaffold an account capable of linking the familiar observation that things fall to the ground when dropped with the imperious passage of the planets in the sky. A single theory of gravity, and three laws of mechanical motion, allowed a unified story to be told in which the motion of massive objects under idealized conditions was intelligible, because it was entirely predictable.

This is the birth of modern physics. Physics came first among the sciences. It has its roots in the observation of the stars and planets, and it applies the craft of measurement to its objects, interpreting those measurements within a mathematical framework. To this day, physics has a reputation for grounding our most reliable stories about the nature of reality.

When people use the term "physical" in every day discourse, they typically do not mean the domain of the physicist, which in the most modern physics is entirely remote from the familiar grounded world of everyday experience. They mean instead the quality of being "real" or "indubitable." The kind of proof a Doubting Thomas seeks as he insists on bringing the risen Christ before him, and placing his finger in the wound, bearing witness to its warmth and wetness, has nothing to do with either Newton or Einstein. It is a proof available only to a living, experiencing, active agent who uses his body to explore what is before him. Yet we might describe it as physical proof, in this sense of "very very real."

The magic of precise measurement is nicely illustrated by a recent success story, the LIGO project. This involved the detection of a remote and enormous collision between two black holes, generating a gravitational wave (not something Newton would have been able to make sense of) that propagated throughout the cosmos. At the terrestrial level, where we live, the wave had become attenuated to such an extent that it could only be detected using a device capable of registering motion smaller than the width of an individual proton. Neither the original collision, nor this deflection could be directly observed, of course, but the tiny deflection had in turn to be reamplified until a trace could be drawn by a needle on a dial that the physicists could see. It is worth remembering that our sense of reality is secured in this manner—through seeing, touching, feeling the indubitable presence of the world as it makes itself available to an embodied being. The physical account of the empirical world must be built on measurement, and the process of measurement must arrive at some point at the witnessing of the physicist herself.

The difference between contemporary physics and that of Newton is vast. The mechanics of Newton made a kind of immediate intuitive sense as it dealt with the motion, inertia, momentum, of objects about the size of a breadbox, and idealized so as to ignore such complexities as friction and turbulence. Since the early 20th Century, observations that are relevant to the latest physical theory have been made at spatial and time scales that are vastly removed from the familiar embodied lifeworld of any individual person. But modern physics grew from the taming of space and time within a framework established, among many others, by Galileo, Copernicus, Descartes, and Newton. Here, for the first time, the space of the heavens and the earth are brought together to form a single isometric container, such that a meter measured at any point is comparable to a meter measured anywhere else. Once this framework was in place, the challenging business of mapping the globe with increasing accuracy could really get underway, a

practice that continues to this day. Time too became homogenized, with the development of calendars, clocks, and conventions that allow an interval measured here to be compared to an interval measured under different circumstances there. The early modern cosmos became a 4-dimensional container, made intelligible by the use of standardized measurement units and practices.

**Isometry of space and time**

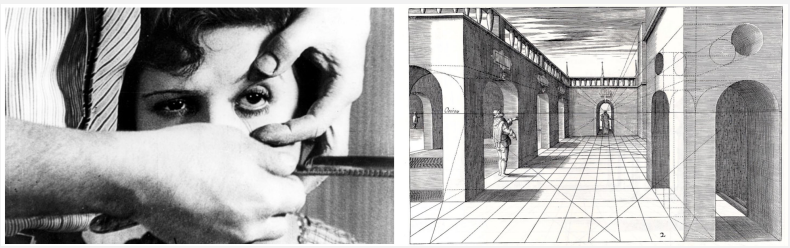

*Figure 5: Left: A scene from Luis Buñuel and Salvador Dalí's film "Un Chien Andalou" (1929). Right: Strongly perspectival scene by Hans Vredeman de Vries (1605).*

Space and time *as experienced by a subject* are decidedly non-isometric. The centimeter that extends from your eyeball in the direction of your gaze is charged with significance in a way that a centimeter in the empty desert is not. A wasp in one space or other will be of radically different import to you, as a subject. The last ten seconds you experienced are strictly incomparable to ten seconds from an arbitrary period in the past or future. It seems remarkable that we think of time and space as isometric, although no person has ever experienced them as such. This modern view denigrates the subjective, attributing reality to our models, rather than to our experience. The development of linear perspective, around the same time as the development of the Galilean model of space, has contributed greatly to our intuitions about the geometry of space. Clocks, calendars, and the working week do something similar for time.

For all its undeniable utility, this cosmological model served to enhance an artificial distinction that had emerged over many centuries, perhaps as far back as ancient Greece. Alfred North Whitehead called it the "bifurcation of nature" (Whitehead, 1920) and by that he meant the imposition of

a seemingly unbridgeable gulf between the qualitative reality of lived experience (let us call this, tentatively and unhappily, "subjectivity"), and the impersonal model of "that which is" as described by an "objective" science. On the one side, we find the very substance of experienced reality, in which things are heavy, warm, sharp, or tasty. On the other, we find a reality that can only be indexed by number, and described as the motion of impersonal matter. Bruno Latour describes it thus:

> Bifurcation is what happens whenever we think the world is divided into two sets of things: one which is composed of the fundamental constituents of the universe—invisible to the eyes, known to science, yet real and valueless—and the other which is constituted of what the mind has to add to the basic building blocks of the world in order to make sense of them (Latour, 2005, pp. 225–226)

As science developed in the following centuries, this artificial gap was spanned by the postulation and elaboration of the individual mind, imagined as a hidden domain which transformed input from an external world into the qualitative stuff of individual experience. The general picture has become so very familiar that it is difficult to see it for the elaborate artifice that it is. Specialists in the philosophy of science are well aware that a simplistic and stark division between the nominally subjective and the objective is anything but a simple matter, but in everyday talk about ourselves and our lives, the particular form of realism that insists that the cosmos is simply existent, and one's experience of it is a derived function of an individual mind (implemented in brain tissue) is almost impossible to get away from (Daston and Galison, 1992). It might be considered to be the unexamined dogma of those of us who look to science rather than religion to ground their being. It is a familiar picture, but a restricted one, and there are alternative ways of considering the relation between (many kinds of) subjects and their respective objects.

Before we further explore the notion of a subject in the sciences, let us look at a joint speech example that, while recent, hearkens back to, and is continuous with practices much older than modern science that served to orient people with respect to time.

## Example 7: Countdown

A LARGE CROWD IS GATHERED in Dublin city center, outside Trinity College. The mood is festive. The crowd is energized and somewhat drunken in places. Colorful light patterns are projected onto the front of the picturesque university facade. At some point, the patterns and images turn into numbers, and bit by bit, the thousands of disparate conversations everyone was engaging in stop, and the countdown begins. Everybody present chants in unison. No instructions are necessary. 10 . . . 9 . . . 8 . . . to zero, at which point a large cheer goes up, fireworks explode, and the myriad conversations restart, so that the overall acoustic landscape disintegrates once more, from coherence to variousness. It is, of course, midnight on New Year's Eve.

This familiar celebration happened at the same time in countless places within one time zone, and was repeated at hourly intervals until the entire globe had passed from 2014 to 2015. From a suitably disinterested point of view, midnight means nothing at all. After the construction of coordinated time zones and clothing the globe with a meshwork of latitude and longitude, it is possible to agree, by convention, on a time measurement system that allows more than one person to recognize something called midnight. And of course, the idea of a new year makes no sense without the convention of the calendar, through which patterns observed in one year can be brought into alignment with patterns observed in another. These days, it is easy to forget how much artifice and convention underlies such a banal celebration.

But atomic clocks, accurate GPS, and live television transmission are newcomers to the human world, even though nothing material has changed in the constant rotation and orbit of the earth. The oldest ways we know by which humans have marked time lie in ritual. Rituals have provided the conjunctions necessary to mark both singular transitions, e.g. from one stage of life to another, and recurrences, e.g. the beginning of the annual harvest. Roy Rappaport's masterly work (Rappaport, 1999) on the role of ritual in the making of humanity expounds in depth upon the way in which collective rituals provided the foundation for collective coordination and the establishment of shared collective perspectives that ground a sense of order. (It is certainly no coincidence that Rappaport is also the only author I have come across who seems to have articulated the importance of joint speech as one component of ritual.)

## Medicine and the body as subject

The strict division between an "objective" material world that exists, but is devoid of such meaningful attributes as color, taste, or feelings, and a "subjective" domain in which all qualities reside, though they may not be measured directly, is very hard to get away from. To some, psychology is the science that enforces this separation and that takes on responsibility for all questions of meaning and quality. To others, psychology is not a science at all, but the pragmatic business of using whatever accounts of the person may be employed to make people feel and function better. This is sorely vexed territory, and most people choose to avoid it. Scientific psychologists are committed to their role in allowing this division of responsibility between disciplines. Physicists, by and large, want no part of it. The last mainstream physicist to insist that the reality understood by physicists be relevant to the reality experienced by living subjects was probably David Bohm (Nichol, 2005). Bohm regarded science as a practice that directly extends human perceptual abilities, by contributing to the intelligibility of one's experience of the world. The role of the telescope, microscope, high-speed film, time-lapse video, these all serve to make events and things directly perceivable, and they thus extend the human capacity to encounter, experience, understand, and cope with, the variety of the universe.[1] To quote Bohm:

> [T]here is always finally a stage where an essentially perceptual process is needed in scientific research—a process taking place within the scientist himself. (Nichol, 2005, p. 73)

It may help, then, to briefly pursue some examples that make it clear that a simplistic division between an objective material world and a subjective mental domain cannot be sustained. Rather than worrying about minds, it will be simpler to consider first the sciences of the body as they have grown within the domain of medicine.

Medical practice is as old as humanity. The accompanying science is less so. If we take our understanding of science as it began in the observation of the stars, then introducing the body demands a reconsideration of what we mean by science, how the fact/value distinction is handled, and what we mean by a subject. When we discuss the orbits of the planets around

[1] In this spirit, we might view joint speech, not as a phenomenon in need of elucidation, but as a tool for directing our awareness as we study our own constitution, more akin to a microscope than a puzzle; something to be seen through, rather than stared at.

the sun, or indeed the chemical reaction of various elements, we have before us objects in a simple sense. Their activities are independent of us, and our observations do not interact with them. Under these circumstances, we never resort to the notion of *purpose* in our scientific accounts. Aristotle might have understood natural law as an expression of built in purposes, or teleology, but such an explanation would no longer satisfy contemporary astronomers or chemists.

Medicine has its own particular trajectory over the centuries in which the various aspects of logos have bumped off one another, as the body was contested as a natural object, as something in need of social control, and as the locus of received tradition and custom. If we understand medicine as the field that interprets the body, with a very applied intent of preserving its various functions, and restoring those functions in case of breakdown, then we have already taken a very significant step from the simplistic subject/object divide we were faced with when contrasting physics and psychology. Functions do not fit in any vocabulary that asserts a simple clean split between subject and object. When we assert that some process or thing fulfills or enacts a function, we are describing that process or thing in a way that presupposes the notion of a goal, for a function may be successfully implemented or it may fail to be achieved. The role of the goal here makes it clear that functions are necessarily *for* someone, or some thing. Functions, in other words, demand a subject for whom things are of sigificance. In medicine, we encounter the body as a subject in its own right, neither identical to, nor separable from, the person. Indeed, somatic medicine can be viewed as the domain in which the body underwrites all attributions of function. We might recognize that consensus is reached in the domain of medical science precisely because those who take part in the conversation are embodied in a particular way, and they thus share a common understanding of function grounded in the body.

As we now consider medical science, or more precisely, physiology, we find we must ask questions such as "What is the function of the heart in the body?" There is consensus that the heart acts as a pump for the blood, and in so doing it provides the cells of the body with nutrients, oxygen, and it transports waste. Nobody would question this simple, if inadequate, account of the role of the heart in the body. But this is not a simply objective account, in the sense of a freestanding mind-independent fact in need of no qualification, at all. It introduces the notion of purpose, or function. With such language, we have introduced normative concerns, as the heart may be said to serve the body, or to fail in serving it. A comet may, or may not, emerge intact after its most recent orbit around the sun, but there is

no question of it failing, as there is no goal. If we speak of it failing, it is perfectly clear that we are projecting something onto the comet that has nothing to do with the comet's own being. Two chemicals may or may not react, but they do not fail to react.[2] But in the domain of the living, things are different. We can recognize, describe and understand the heart if, and only if, we frame our observations on the assumption that there is a body for whom the activity of the heart makes a difference. This makes the body into a subject, and the activity of the heart is intrinsically meaningful from the point of view of the body. The normative character of the activity of the heart brings with it a requirement that we acknowledge a limited form of perspectivalism: We consider the significance of the activity of pumping from the perspective of the body. In the absence of such framing, there is no sense in which the heart can fail. A non-beating heart is entirely on par with a healthy beating heart from a strictly objective, sufficiently disinterested, point of view, but with recognition of the integrity of the body, we may recognize success or failure.

Is the activity of the heart as a pump then an objective fact, a subjective interpretation, or something else entirely? As I see it, it is clearly an objective fact, but one that we have constructed together. You and I, and presumably most human beings, recognize the role of the heart in the body. We agree on how to frame its activity (view it as a pump, tasked with acting as an engine of transport in the circulation of nutrients and waste). We share this framing because our own being is grounded in our own bodies; because we share this framing, we can agree on a whole bunch of facts. But they do require framing, and the facts are not free floating. They are tethered to the consensus we create by framing things in a similar manner. We are implicated in them, and they are secure to the extent that there is common ground among us. Because of the common ground, we can observe the heart together and arrive at a consensus about its role, its function, its purpose, or whatever teleological term we might use.

We might contrast this consensus-based account with the slightly odd consideration of the function of litter, as found, for example in the car park of a national park. Obviously, from our perspective, litter is an undesirable annoyance, which hopefully does not arise intentionally, and so we would be unwilling to interpret it as having any function or purpose at all. But consider its role in the foraging activity of the local crows. A car park that

[2] Of course, you and I might have a bet about the outcome of mixing the chemicals, and we might then have a sense of success or failure as they do, or do not, react. But it is clear that the chemicals have no stake in our gambling.

persists over decades will reliably generate a certain amount of litter, and this litter may be important in sustaining a local population of corvids. From their perspective, the trash most certainly has a function. So who is right? The question is meaningless, as the coherence of the notion of a functional account will always depend upon our framing. Are we considering the perspective of the visitor to the car park? Or are we considering the rather different perspective of a group of crows? Other framings are also possible. Ants may have a stake in this, as does the local recycling company, who generate profit from the trash. The lesson to be learned is that every time we consider the function of something, we have framed our observation in a particular way, and we are making use of a specific kind of subjectivity– the perspective from which the notion of a function makes sense. We do not normally make this framing explicit, because we share a lot of common ground and will tend to adopt similar attitudes towards the objects of our discussion.

## Biology and the individual

Biological subjects come in all sizes and shapes. The quintessential biological entity, the single cell, represents a minimal form of organization that we might choose to identify as an individual, for whom its interactions with the world are normative. A cell may thrive on its own terms, or it may fail and die. This makes interactions with its local environment meaningful in a very precise, and cell-centered way. We explore this limiting and minimalist case in more detail in the final section. Cells make up organs, which we can also see as subjects, for they can succeed or fail in serving their superordinate bodies. Individual multicellular organisms can be seen as subjects, for they live or die. Groups of individuals form coherent wholes that thrive or perish. In each case, the organizational characteristics that lead us to regard the entity as a subject will be contingent and specific. The organizational characteristics of a cell are different from the organizational characteristics of a shoal of fish, of a nervous system, or of a nuclear gorilla family. As these characteristics change, so to do the features by which we judge the "health" or integrity of the subject. The term "health" is somewhat misleading here, as we habitually apply that term to the individual multicellular organism, or body, and little else. The older Greek term *εὐδαιμονία* (eudaimonia, literally "good spirit") might be of use here. It refers to a general sense of flourishing or thriving, with the implication that the terms in which that thriving is to be understood are predicated upon the intrinsic teleology of the organism

itself. What is good for a cell is not the same thing as what is good for a lion or for a colony of seabirds, but each may be said to thrive (or fail to thrive) on its own terms.

The recognition of the interconnectedness of all life within the biosphere should not come as news, at least not to theoretical biologists. It was living beings that transformed the surface of the earth, bringing forth the biosphere as rocks turned to soil, as the atmosphere was oxygenated, and as countless other processes reshaped and crafted the common resources that feather all our nests. In the approximately 4 billion years of life on the planet, all living beings have depended on other living beings to survive. One important consequence of this state of affairs is that no biological entity can really be considered as a free-standing individual, distinct from those interactions with its world that serve to keep it thriving. In a strict sense, there are no individuals in biology. For example, the flourishing of the human body requires the presence of vast numbers of bacterial symbionts whose genetic lineage is independent of the host. In a useful review article, Gilbert et al (2012) go through the various ways in which it is often convenient to consider a given organism or entity as an individual, and they demonstrate that in each case, a closer analysis reveals complex reciprocal relations of dependence that make the entity strictly inseparable from everything else. This works on anatomical and physiological terms; it works for developmental and for evolutionary considerations, but also if we consider genetic individuality or individuality based on the immune system. Within theoretical biology, symbiosis, or co-existence within webs of mutual co-determination, is rapidly replacing an older model of the autonomous individual.

As we recognize such reciprocal dependencies, it becomes clear that biology, as a science, must acknowledge and learn to discuss, the perspectives of many kinds of subjects. Subjects need to be understood on their own terms, and thus biology has no choice but to learn how to invoke a constrained notion of purpose or teleology. In the absence of such a way of interpreting entities and processes, the notion of function is simply not available. In particular, we have no license to speak of function at all if we rely on a strict separation between the subjective and the objective, as we frequently do in everyday discourse.

For those who might be ill at ease with the use of such terms as function, purpose, and teleology within science, it might help to note that the framing that licenses attribution of function is only possible if we agree on that framing. The kind of objective account arrived at in biology is more complex than in astronomy, and any such account that leans on notions of purpose implicates the observers themselves. Such accounts are unproblem-

atic precisely where different observers share a framing understanding of the kind of systematic organization that warrants attribution of function. The role of the heart within the mammalian body has been our guiding example here. Where the framing assumptions are not explicit or self-evident, misunderstandings are likely to arise, and it may be impossible to arrive at a consensus-based account that makes use of the notion of function. This will be the case as the discourse moves from biology to the social sciences. We will leave the biological account behind for now, to pick up this discussion in the final section, where a brief introduction is given to the enactive approach within cognitive science, which leans on the processes of the living, rather than on minds, to address human experience and behavior.

## Subjects in the psychological and social sciences

It has proven extremely difficult to develop concepts within the social sciences that can be understood as continuous with the insights of other branches of the so-called harder natural sciences. The agencies that are operative within society are varied and powerful. We speak glibly of class warfare, of the corrupting influence of the media, of nations, races, and tribes. We attribute grit and resolve to football teams, and we complain of how the rich, the poor, the right, or the left are responsible for the mess we find ourselves in. This talk presupposes that it makes sense to attribute agency to these diverse and poorly defined entities. Such entities seem to flourish within political discourse rather better than they do within scientific accounts. This will come as no surprise to anybody. But if our sciences are to be of utility to the kind of complex multi-faceted being that we are, we cannot do without some means of referring to such collective agents within scientific discourse. When we do so, the division of labor within logos, between natural law, civil law, and tradition, is again not clear cut or simply given. It seems unreasonable to demand of science that it speak the same language to discuss the activity of an organic human body, of a citizen, or of a pious devotee. We are all these, but they are not a single thing, nor do they act within a single domain.

An essential part of our scientific worldview that developed after the emergence of the physical sciences is the view of mind as a thing, singular, personal, and unobservable. This concept has become so deep-seated in our understanding of ourselves that we can hardly see it as anything other than given, despite the difficulties that arise when such a picture is subjected to scrutiny. The conventional place to point when discussing the modern view

of mind is to the French philosopher and mathematician, René Descartes. In truth Descartes is frequently and somewhat inappropriately treated as if he were single-handedly responsible for this metaphysically challenging notion, which is sometimes called the *cogito*, from the well-known phrase *cogito, ergo sum*, or "I think, therefore I am." The concept is older, and it has a long and complex history as the Christian notion of the soul gradually decoupled itself from the domain of theology, and became the problematic notion of mind we grapple with today (Reed, 1997). When we speak of a Cartesian view of mind, we mean the idea that minds are separate from the physical world, that minds are separate from one another, and that minds are unobservable, personal things that persist in an individual from birth to grave. The cogito is this notional domain, and its delineation as the cogito at the birth of modern physical science was a way of punting many questions about reality that science was not yet in any position to address. Material properties of the world are relatively easy to measure, and the physical sciences exploited this with obvious success in the following centuries. Qualitative properties of experience demand a different approach, and the postulation of individual minds allowed most of science to avoid having to answer hard questions about meaning, value, and perspective. We cannot pursue this in detail here, but we might note in passing that the singular, unobservable, and autonomous mind has not yet become something on which a scientific consensus has developed, and the place of psychology among the sciences, which frequently relies on such a foundational notion, is very much a matter of contention.

In a society in which the individual person is granted a unique and essential form of agency trumping all others, a yawning divide opens up between two domains of explanation: the psychological and the social. By treating psychology as if it were a science capable of generating objective facts, on par with those found in the hard sciences, without recognizing the need to interrogate any simple subject/object division, we summon into being an entity, the mind, which helps us locate ourselves as individuals, but thereby blinds us to the fact that we are also constituted by our participation in many kinds of collective entities. An overly restrictive attribution of agency to one being only, the individual person, must make any kind of social agency or social cause appear parasitic upon the acts of individuals. We may describe social phenomena as emergent, as when dancers in a vigorous mosh-pit spontaneously begin to move in concentric circles, or when banks are toppled by a frenzy of withdrawals, but such emergent effects do not dislodge the presumed locus of agency within the person. In this situation, the social sciences are dependent upon the claims of the psychological sciences in an

asymmetrical fashion. The claims of the psychological sciences that identify agency or actors are treated as if they were prior to claims of collective agency, not least because of the widespread respect for the integrity and agency of the individual person upon which the post-enlightenment view of society is built. Nor will this respect for the individual person be readily unseated, for it is supported by important, if ill-defined, consensus on such democratic humanist ideals as the concept of human rights, and the role of the individual voter within a democracy.

Here we can see that there is an intertwining and co-mingling among the three kinds of order, subsumed here under the single term logos, that we have cause to consider. One might appeal to the principle of natural law in staking out a claim for any particular account of individual psychology. Scientific psychologists understand their work in this manner: The mind or the person is considered an object, to be understood as other objects are understood. But any attempt at developing such a science must be committed to specific ways of delineating the subject, or agent, presumed to be acting on his or her own behalf, exercising free will, and independent, to some extent, of circumstances. And when we do so, we rapidly find ourselves embroiled in disagreements that lie elsewhere, in the negotiation of authority and of the right to act within a political system, or in the overlapping claims of diverse cultural and religious traditions.

When we think we are arguing about objective facts, and we interpret "objectivity" as if it were a simple matter of establishing that which exists independently of any belief system, but we then find that our arguments have immediate and urgent consequences in the political sphere, in law, or in codes of conduct, an important corrective needs to be applied. It seems as if there are many kinds of "person" here that are becoming confused. The person as a biological entity (skin clad, somatic) is readily confused with the person as subject of political and social moralizing which in turn is in danger of morphing into the person as a character in modernist fiction, whose inner life is a babbling stream of consciousness unseen and unseeable by others. In living our lives, we move fluidly among collectives, and enact different kinds of collectives that thereby also define us: the tribe, the congregation, the mob, the playgroup, the quartette, the audience, or the platoon.

It is perhaps unsurprising that when psychology is required to address the experience and behavior of the person, it becomes tasked with providing normative explanations, with distinguishing the good from the bad, and with providing guidance in the conduct of the affairs of the person. Soteriology, or telling people what is good for them, has been a central part of psychology since William James, in contributing to the foundation of the dis-

cipline, adopted a strongly pragmatic, rather than metaphysical, approach. It is here that we can locate the beginnings of the self-help tradition within psychology. But this intrusion of what amounts to lifestyle advice, or claims about what is good or bad for a person, should sensitize us to the unresolved issues about the delegation of authority among the various branches of the logos, or common order. As I have sketched it above, any normative statements in psychology, even any functional statements, are meaningful in as much as they are framed by the assumption of a psychological subject. But we have not yet managed to delineate any entity corresponding to such a psychological subject, we have not managed to transplant it into any kind of objective framework, and we find ourselves in need of some argument if we are to understand what kind of subjects we can recognize.

For if agency is the exclusive prerogative of the psychological subject, then the collective subjects we encounter in joint speech are mere fictions, as are the distributed and collective agencies we refer to in everyday discourse. The intentions, purposes, and ardent desires of an inflamed crowd chanting for the fall of the regime would be no more and no less than the sum of the individual passions of the participants, just as the aggression shown by one nation to another would have to be understood to be a metaphorical ascription, denied any form of literal truth. And the insistence of social scientists that there are perfectly well formed social facts, accounted for in terms of social agents, would continue to be in vain. This requires some unpicking if we are to do justice to the collectives we constitute, but that also constitute us.

Let us take the collective subject of Example 2 (Strife at the Al Aqsa Mosque) as a worked example. During the period of chanting, we can readily see a collective agent enacted through the coordinated and synchronized behavior of chanting. This collective entity comes together under very particular circumstances. It is a response to the unrest, and to the presence and activity of the other collective entity, the riot police. But it is not just a response, for it is active, it seethes, pushes back, engages with the police. No single person is in charge, and the collective entity exists for just as long as the activity of its constituents, the people, brings it into being. The collective subject here is a dynamically individuated entity, recognizable against the general background confusion by the relatively coherent activity of its members. The activity of the constituents is what keeps the collective entity in existence. The environment in which it arose and within which it persists is itself active, agentive, animated, and a source of challenges, even existential threats, to the collective.

This entity cannot be transported to a laboratory. It is distinguished from its surround, but it is not divorceable from it. It exists in context, and only in context. It is not well described as an input-output system. If it is perturbed from the outside, say by a baton blow, its response is not dictated by that blow. To the extent that a given system acts as an autonomous agent, there is no deterministic link between things done to the entity and the subsequent response of the entity (providing external conditions do not become so overwhelming as to destroy the entity completely). But it is a subject. We can speak coherently of its perspective, and contrast that with the perspective of the riot police, or contrast it with a more disinterested point of view that can recognize both collective groups as well as the motley collection of individuals who belong to neither. We can bring the classic vocabulary of intentionality, or aboutness, to bear on it: This collective is purposive; it is one pole in a field of relations that arises in this context. Its actions can be interpreted from without as arising from purposes, goals, and concerns.

Are all collectives to be granted subjecthood, then? This might appear unreasonable, or facile. There has been a deal of discussion within philosophical circles of the manner in which we might, should, or could regard many kinds of collectivities as subjects or as agents (Carrier, 1986; Pettit, 2004; Huebner et al., 2010). It is clear, for example, that we effortlessly co-opt the everyday psychological vocabulary to describe the goings on of nation states, who are not only described as agents (Germany intervenes, China declines, etc.), but as if they were perceiving and acting in manners strictly analogous to individual people (France takes offence, Brazil worries, Japan sees...). Most people are not tasked with deciding the degree to which such language use should be taken literally, considered as metaphor, or neither, e.g. regarding such usage as one more language game.

The contribution to be made here is not to take sides in such issues, but to demonstrate how joint speech may be used as an index of specific kinds of collectivities, allowing the identification of groups and their associated environments that need to be taken seriously. Joint speech can function as a guide for our collective attention, bringing to the fore and making explicit the practices that together give rise directly to collective subjects, and, as I shall argue, by extension give rise to the three facets of logos. The collective subject observed outside the Al Aqsa mosque must be addressed, but it also cannot be understood in isolation from the context in which it arises and persists. The kind of observations we need to make here then are quite unlike the two spatial coordinates and a time stamp that are necessary and sufficient to specify the position of a star in the sky. Rather, they are in need

of rich and thick description that seeks to make the context and the subject intelligible at the same time. In science conducted in a strictly objective key, it may be enough to locate an object in space and time. In the sciences of the living, the context of any observation must form part of the observation and such contextual elaboration must inform the recognition of competing and distinct perspectives.

Joint speech has a role to play here in directing our attention to the many ways in which each of us exists, not as an autonomous monolith, but as a participant in, and hence a constituent of, many kinds of collectives. Some of these are quite obvious, such as the groups to which we pledge and display allegiance through symbols, clothes, and rituals. Joint speech is an overt and reliable index of such allegiances. Other collectives may be transient, as shared purpose causes many people to act in unison, chanting their anger, or expressing their elation. Our complex web of interdependencies will not be exhaustively illuminated in this fashion. We are multitudes, and we are various, and we pass fluidly from one mode of collective being to another without even noticing. But attention to a clear empirical index such as joint speech may help us to move beyond the restricted view that we are subjects of one kind only, bounded by the individual body, and divorced from the world in which we are, in fact, inextricably enmeshed.

# Part II

# The Science of Joint Speech

Although Chapter 1 began by bemoaning the absence of the science of joint speech, it is possible to identify some modest attempts to study joint speech with the methods and assumptions of contemporary science. Beginning in the domain of sound, we can look at joint speech both from a phonetic point of view (the sounds of speech), and as it relates to the sounds of music. A strict speech/music division does not appear tenable when looking at joint speech.

Moving to the study of movement and behavior, joint speech can be studied as a specific example of synchronized movement. Movement science is introduced as a domain that is not beholden to a single kind of subject, whereas behavioral science typically assumes a single kind of agent or subject. We examine the hostility shown by experimental psychology to movement, and consider its consequences for the notion of the subject.

Turning to neuroscience, it becomes increasingly difficult to get away from the individualistic commitments of scientific psychology. In line with recent embodied approaches to the mind, it is argued that experience does not arise in the brain. An argument is developed that the treatment of the brain within contemporary cognitive neuroscience is riddled with assumptions that are beholden to culturally specific aspects of modern Western culture rather than any objective picture based on science. Nonetheless, it is both possible and profitable to study joint speech using the tools and methods of contemporary cognitive neuroscience.

Joint speech needs to be considered as a form of language, but it does not find easy accommodation within the discipline of linguistics. An everyday, common sense understanding of what language is, and the scientific treatment of language, are both unwittingly founded upon the notion of message passing from one Cartesian domain to another. Joint speech may be more profitably regarded as a form of orality, long predating writing. In this context, it is worth pausing to consider how the voice and the eyes work together in different kinds of communicative contexts. Consideration of the link between the eyes and the voice suggests how a story of the evolution of language might be constructed.

# Chapter 4

# The Sounds of Joint Speech

The absence of scientific work into the behavior of joint speech was singled out as an important and telling absence, that might call us to fix our sights resolutely upon the way in which subjects, individual and collective, feature in our consensus-based accounts of the world. In so doing, we will have to raise questions that call into question any single simplistic stance with respect to the practice of science. It is possible that this might be taken as reason to dismiss the entire point of view being assembled here. It therefore seems appropriate to lay down two statements that any reading, critical or sympathetic, might do well to bear in mind.

When we come to study joint speech empirically, we are not witnessing some mechanical means of creating, altering, or even destroying worlds that has hitherto been ignored. Joint speech is speech spoken by many people at the same time, but it is still speech. It does not have efficient causal powers in any straightforward way. But as an object of study, it has the potential to provide access to the many, various means by which logos arises, by which humans create order, and by which the relations between the various kinds of order—natural, civil, religious—are understood, brought into being, and maintained. The topic of joint speech can serve as a lens, helping to bring specific phenomena and practices to our attention, and helping to frame our discussion, particularly around the vigorous border disputes that arise as the different kinds of order are negotiated.

The second point to insist upon is that joint speech is entirely amenable to perfectly unremarkable scientific investigation, and there is much to be found there. The empirical science of joint speech is almost non-existent, but not quite. In this chapter, we will look at some things to be found as we examine the sounds of joint speech. This provides points of contact to

language science through the discipline of phonetics, and to musicology, as we look at the speech–music divide. Later, we will extend our coverage of the little scientific work that exists to movement science and to cognitive neuroscience. These happen to be the areas of science with which I and my several collaborators have some passing familiarity. I will not assume that the reader has any expertise in the individual disciplines, and I will attempt to frame the findings in a manner that speaks to the much larger discussion in which I wish to embed this work. Each distinct discipline brings different aspects of the grand theme of the subject to light. We begin in a relatively tranquil area of science, phonetics.

## The phonetics of joint speech

Phoneticians are concerned with the production, transmission, and perception of speech. Because they take speech as their object of study, they have a rather particular relationship with the overall field of linguistics, and it would be well to make that explicit. In a later chapter, we shall have to consider the relation between the voice, on the one hand, and something called "language" on the other, so it might help to clear some ground at this point. To the confusion of undergraduates, there are two separate fields that study the sounds of speech–phonetics and phonology–and in order to understand the difference between them it is necessary to sketch very hastily the history of the scientific study of language, beginning around the end of the 19th Century. We might divide the following period into two halves, one with roots firmly in the 19th Century, and the second, in the information age after the Second World War.

In the first period, we saw the emergence of what is known as "structural linguistics," and the guiding figure was the Swiss academic Ferdinand de Saussure (1857–1913). Science in this period had a habit of systematizing its observations in order to better grasp the underlying relations among things that exhibited different kinds of surface properties. In Figure 6 you can see a comparison of the famous periodic table of the elements originating with Mendeleev (1869) and the consonantal chart of the International Phonetic Alphabet (IPA), first published in 1897. In each you see that a set of discrete elements has been ordered systematically in such a way as to make clear family relations in more than one dimension. The phonetic symbols are intended to cover the set of possible consonantal sounds that occur in all the world's languages.

## Systematization and structure

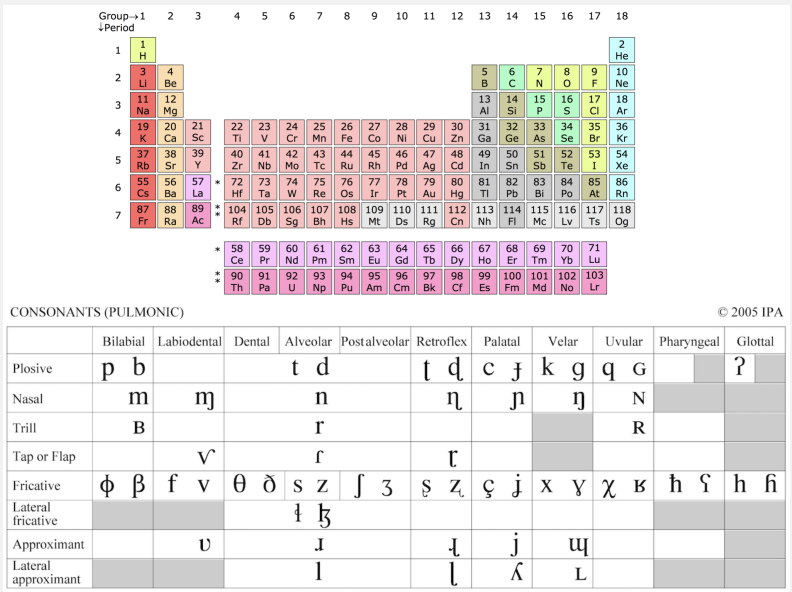


*Figure 6: Top: The Periodic Table of the Elements. Bottom: The International Phonetic Alphabet (Consonants only). (Periodic table: Attribution User Sandbh, CCA-SA 4.0; IPA: courtesy of the International Phonetic Association, CCA-SA 3.0.)*

Both tables seek to bring order to a large collection of distinguishable entities by sorting them along multiple dimensions at once, thus capturing a suite of contrasts that apportions each element its distinctive place. One strength of such systems is their ability to predict the feature constellations of novel entities, not yet encountered. The assumption of any such structuralist approach is that the domain in question may be exhaustively described using only such features and contrasts.

Somewhat fondly, perhaps, there was a view that study of the better-known languages, English first and foremost among them, would provide the kind of information required to capture categorical sound distinctions in all languages. The individual consonantal and vowel sounds captured by the IPA were assumed to be the basic building blocks of speech, from which syllables, and hence words were constructed by sequencing. Each language was assumed to have its own discrete set of abstract contrastive sound units,

or phonemes, that could be combined in accordance with language specific rules. Thus a sequence we might write as pfiff (or, phonemically, /pfIf/) would be fine in German, but would violate the sequencing rules of English. Moreover, there were assumed to be underlying regularities, so well defined that they might be called rules, that governed the sequences of sounds within a given language. Studying these regularities was the task of the phonologist.

We could go much further here, but for our present purposes it is perhaps most important to note that the phonemes that formed the basic atomistic units of the phonologist might be best understood as the ghosts of letters. These hypothetical units of speech are defined by their position within an organizing system, and beyond that, the posited elements bear no necessary relationship to actual sound. Here, English is really not a good place to cultivate one's intuitions, as the English language carries within it the a long history of many peoples, including the Anglo-Saxons, the Vikings, the French Normans, and the more recent history of the British Empire, in which the colonialists encountered very many other languages and cultures, and the language picked up a little here and a little there. As a result, the relation between sound and spelling in English is probably the least systematic in the entire world. In most languages that use alphabets and that have more modest colonial pasts, spelling is a simpler affair and bears a more systematic relation to sound.

Where phonologists deal with the symbolic abstractions that are phonemes, phoneticians get down and dirty with meaty tongues, spit, air pressure, and ears. They study the continuous movements that generate speech, and the resulting acoustic consequences. Mapping from the continuous stream of movement or sound to phonemes is not straightforward. In order to highlight those aspects of the sounds that are relevant to the categorical distinctions of phonology, much of the structure of the speech signal is necessarily ignored. The musical elements of the voice, for example, grouped together under the label "prosody" find no representation in spelling beyond the few characters of punctuation, and so they were largely ignored for a long time. Some have seen the relation of phonetics (meat and sound, observable) to phonology (phonemes and features, inferred) as a form of distillation, whereby the noise introduced by the necessary complexity of a biological system is cleaned up, revealing the underlying sequenced elements that linguists ought to care about. A more recent view would have it that these hypothetical sound atoms are, in fact, a construction of literate scientists imposing their literate understanding onto a complex continuous signal (Port, 2007).

The study of speech has moved a lot in the intervening period. Much of current phonetics addresses prosody directly. Many researchers are paying attention to the gestures that accompany speech, the interaction between the voice and the eyes, and the embedding of speech within specific contexts that greatly constrain and shape the speech itself. In this landscape, it becomes possible to pay attention to those aspects of joint speech that single it out as unique, for if we were to attend to the manner in which it could be transcribed, as a phonemic analysis might suggest, we would probably find nothing at all to report.

A great deal happened in linguistics during the second half of the Twentieth Century to make this structuralist picture look rather old-fashioned. We will defer consideration of those developments for now, noting only that the development of a highly literate understanding of language as something that can be captured indifferently in speech or in writing may be relevant to uncovering why joint speech has not been a topic of scientific inquiry. A reconsideration of the relation between speech and writing will bring with it a reassessment of what we mean by language (Chapter 7).

But what of the phonetics of joint speech? Over the last ten years or so I have conducted many experiments in which experimental subjects (usually students) are brought into the laboratory, equipped with microphones, given a short text to familiarize themselves with, and then asked to read it in synchrony. Typically, I will count down 3-2-1 and they start reading at 0. I began calling this kind of speech "synchronous speech," as I think it important to have a name that differentiates this from the murmured prayers, the shouted protests, the multiplication table recitations, and the many other varieties of joint speech encountered in the wild. It differs in important ways from all these. The subjects are reading a text I have given them. Their words are thus unmotivated, at least compared with those who demand the fall of the regime or who assert their faithfulness to their God. There is no real commitment involved. Subjects tend to try to do what an experimenter asks them to do, and many subjects are surprised to find that the task of speaking in synchrony with another person, often a stranger, is surprisingly easy.

Synchronous speech

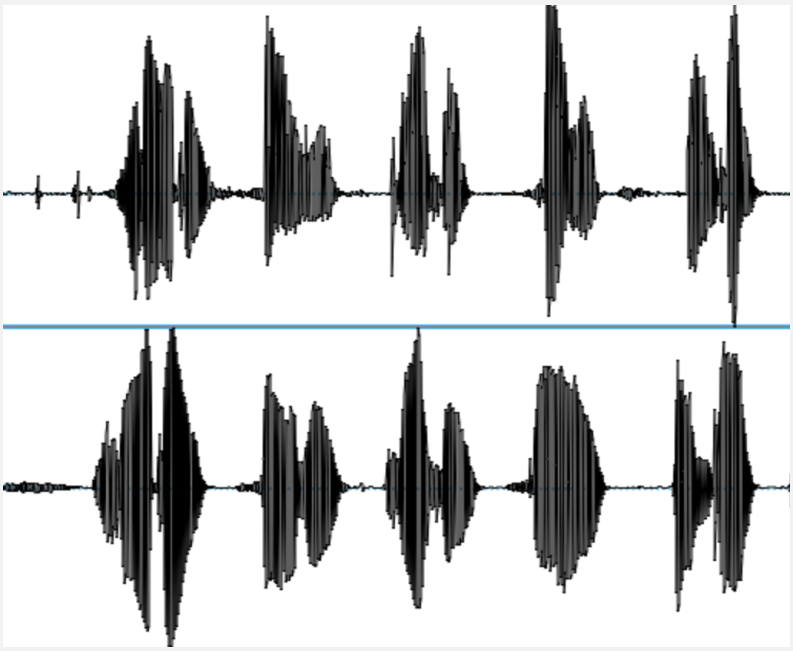

*Figure 7: Speech waveforms of two speakers reading a word list in unison.*

The first finding is so obvious it took me a while to recognize it (Cummins, 2003a). Subjects can do this. They can do it without any practice, and they are extremely good at it. When we measure the asynchrony, or lag, between speakers, we find an average value of 40 ms., or 1/25th of a second. This rises to about 60 ms. at the start of a phrase after a silent pause. This is entirely in keeping with what we know of pause duration in speech; pauses are highly variable and are not precisely timed. After speaking the first syllable or two though, subjects are once more as tightly synchronized as ever.

Now consider the plasticity of the voice. In the course of a single day, you might shout upstairs to your kids to get them out of bed, whisper in your lover's ear, speak calmly and gravely to a friend in trouble, unleash a rapid torrent of syllables at the idiot who parks in front of your garage, and slowly, patiently, enumerate menu choices for your elderly mother. You will do this without attending to the changes in speech. The voice is highly plastic. We adapt it to suit many different kinds of communicative contexts, to suit listeners to whom we stand in a motley variety of relations, under

environmental circumstances ranging from industrial noise pollution on the street to somber enforced silence in church. Despite this flexibility with respect to tempo, volume, musicality, and clarity, subjects who have never met, and have no more instruction than "read in synchrony" manage to shear their speech of all unnecessary expressive variation, and fall into lockstep.

The basic finding can be explored in several ways. For example, we can ask whether it matters that the speakers see each other. Turning the chairs around provides a simple experimental manipulation, and we find that this makes a difference of about 10 ms., but only for the first syllable after a pause. After that, the speech is just as highly synchronized as before. This suggests that even though the speakers are reading from a sheet, there is an effect of peripheral vision that alerts them to the moment of the initiation of speech after a pause. Perhaps it is the simple expedient of breathing in before speaking that they are sensitive to.

We can give subjects some additional practice at speaking with one another (Cummins, 2003a). This turns out not to be much use. Only if we give extensive practice on a specific sentence with a constant co-speaker can we register improvement. In essence, speakers are as good at this task when they walk into the lab as they are going to get. We do find that speakers who are very familiar with one another tend to synchronize more tightly and with greater ease than strangers. Like many of the findings we will review, this is not too surprising. We might expect familiar partners to cooperate better in filling a dishwasher or digging a hole as well.

One question often asked is whether we see clear leaders and followers when people speak in pairs like this. There are two aspects to this question that need to be unpacked. Firstly, if the question is interpreted to mean "Is one speaker ahead of the other in time?" the answer is a very clear "no." It is virtually impossible to remain behind another speaker, a fact that is well known to speech psychophysicists. Indeed, if we artificially introduce a short delay between what one speaks and what one hears as feedback, a condition known as Delayed Auditory Feedback (DAF), we create conditions that make speaking virtually impossible. Speaking under DAF is unpleasant, and stuttering is almost impossible to avoid. Remaining constantly behind a co-speaker would create conditions very similar to DAF. But we can read the question in another way: Are their contributions to the joint speech equal? Not necessarily. There are many ways in which an asymmetry can occur. A more confident or assertive speaker may dictate the tempo of the joint speaking—synchronous speech is usually produced at a slow to moderate tempo. One speaker may be much louder than the other. Different abilities in reading (not everyone can speak fluently while reading aloud) may

influence the performance. So there is no enforced egalitarianism. Speakers remain individual, but they do become locked in time.

If speakers can speak in unison with another person, can they also speak in unison with a recording? This is an easy proposition to test, and the answer is yes, they can, but not as well as with a live speaker (Cummins, 2009). If the recording they are speaking along with was, itself, generated by someone speaking in synchrony with another person, that helps, but it doesn't remove the deleterious effect of the inflexible recording. The reciprocity that inheres in the live situation makes a measurable difference. On the one hand this is still not terribly surprising. When two speakers are live, each can accommodate their voice to match the other, much as two people carrying a table will sensitively and reciprocally modulate their actions to support the collective goal. On the other, it brings to our attention the significant difference between a live face-to-face encounter, and the use of a recording. This becomes very important when we approach joint speech within cognitive neuroscience. For now, we will simply note the introduction of the very important theme of liveness and co-presence that arises when we seek to understand joint speech.

We noted above that speech is highly variable, adapting itself effortlessly to context, partner, noise, and the like. Variability has been the bane of the phonetician for as long as we have been recording subjects. The early goals of phonetics included trying, and failing, to find invariant characteristics of the presumed underlying atomistic elements of speech, the phonemes. When we write words, we use discrete symbols, and a "p" at the end of a word is the same symbol as a "p" at the start of a word. When we speak the words "pat" and "tap" however, the resulting /p/ sounds are very different. The former typically has a strong burst of air after the release of the lip closure, while this is greatly attenuated, or even absent in the latter. The expedient adopted by most researchers for decades was to carefully control the speech that they recorded, that is, to ensure that contextual variability is ruled out as far as possible. So, to examine the difference we just noted, subjects would read sentences containing the target word, with identical surrounding material. A typical carrier sentence framework would thus see subjects reading aloud "I say the word tap again," "I say the word pat again," "I say the word pit again," "I say the word tip again," and so on. No wonder subjects rarely come back for seconds. This kind of rigorous experimental control is reasonably successful at obtaining coherent data that allows us to make generalizations across speakers, to index important categorical differences, and to shore up the structuralist view of speech as assembled from atomic units. Unfortunately the results obtained do not generalize straightforwardly

to speech obtained outside the laboratory. Once we remove the constraints, subjects speed up, become sloppy, or hyper-precise, they shout, slur and generally behave with no respect at all for the phonetician's desire to deliver invariants that will justify the phonologist's work.

Because of the great deal of variability found in speech occurring naturally, phoneticians have long had a fondness for tasks and conditions that can rein in the variability somewhat. For example, if we play loud noise at speakers through headphones, they reliably respond by increasing their volume and articulating their speech in an exaggerated fashion, a phenomenon known as the Lombard effect. We might restrict our observations to very specific kinds of exchange. Speech produced by mothers interacting with their infants, for example, has some very unique characteristics. Unfortunately, a lack of variability is not one of them. We might pretend that we didn't hear an utterance and have speakers repeat it more clearly (Harnsberger et al., 2008). Or we might record speech while subjects are engaged in performing a mentally demanding task at the same time. All these serve in some small way to reduce variability that seems inherent in vocal production.

But the demands of synchronous speech seem to achieve a great deal here in reducing variability. Indeed, the surprising thing is that speakers manage to decide for themselves what can be eliminated and what must remain in order to stay in lockstep with a co-speaker. It thus appears that the simple expedient of having speakers speak in unison might provide a tool for phoneticians, allowing them to obtain speech that has been constrained using the intuitions of the speakers themselves, rather than the artifice of the experimenter (Cummins, 2003b). Figure 8 illustrates this. In a small experiment, subjects read lists of eight words. Each word was a trochee, which is a two-syllable unit in which the first syllable is stressed. A sample word list would be *tango lighter daddy wiper pony cutter pinky mango*. In the experiment, we were particularly interested in the time between the onsets of successive words. Eight word onsets produces 7 intervals. In Figure 8 you can see the sequence of 7 interval measurements for two such lists, each read alone ("solo") or in synchrony with another. The interval durations have been normalized (i.e. divided by the average interval duration) so that a value of 1.0 is the average interval length. This allows us to focus on variability rather than absolute timing. The two sets of measurements on the left exhibit wild variability. Some intervals are much longer or shorter than others, but it is the variation within a given list position that most impresses.

Variability reduction in synchronous speech

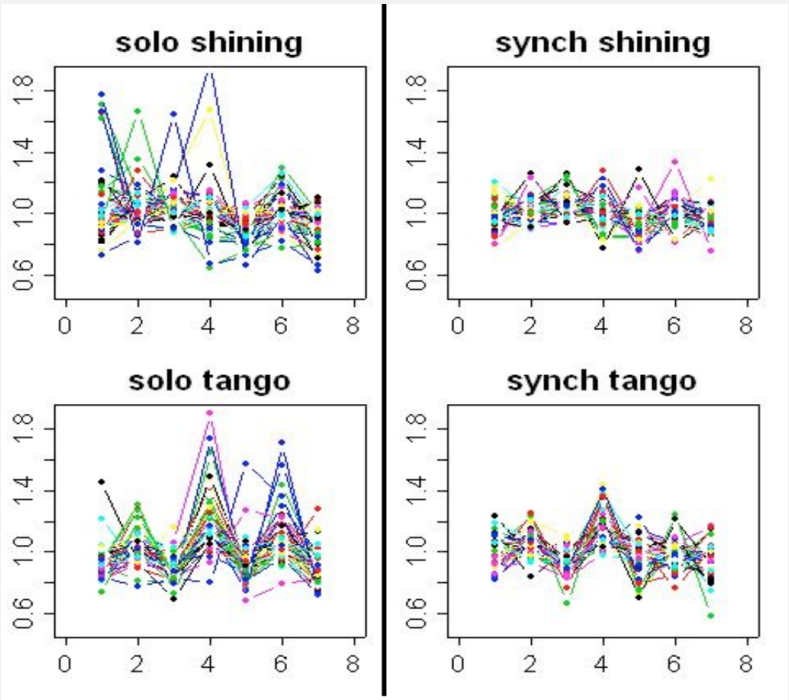


*Figure 8: Intervals between the onset of lists of 8 trochaic words, read alone (left), or in synchrony with another (right). The x-axis records the interval number (8 onsets = 7 intervals). The y-axis records the relative length of each interval (1.0 = mean interval duration).*

By contrast the data on the right appear positively domesticated. Variation from one position to the other is still evident (and this was the focus of the experiment), but variation within a given list position is greatly reduced. The motivation for this experiment, and the conclusions drawn need not concern us here. Phoneticians have their hobbyhorses. What we need to note is that speech produced when people speak in unison is far less variable from person to person than speech produced when we speak alone, and that this fact alone is of special interest to those who study the speech signals we can record.

## The music of joint speech

One of the more obvious ways in which joint speech displays structural or formal characteristics that transcend the domains of prayer, sports, and protest is in the use of repetition. The use of beads within the collective praying traditions of all the Abrahamic faiths, as well as Hindu, Buddhist, Jain and Bahá'i traditions illustrates the pervasiveness of repetition in such religious practices. But chanting on the football terraces or in the protest march is likewise typically highly repetitive, frequently making use of several well known rhythmic templates that allow situation-specific words to fill slots in an otherwise well-practiced structure.

A familiar example of this reusable structure is the rhythmic pattern illustrated below (Figure 9). This is the basic structure of the chant used in the election campaign of Salvador Allende in Chile in 1970. The words used at that time were "¡El **pue**blo u**ni**do, jam**ás** se**rá** ven**ci**do!" or "The **peo**ple, u**ni**ted, will **ne**ver **be** de**fea**ted!" Strongly emphasized syllables are rendered in bold font. After the CIA-backed coup that led to the installation of the fascist Pinochet government in 1973, the chant became a folk song composed by Sergio Ortega. Since then, the basic rhythm of this chant has been re-appropriated in very many contexts. The same rhythmic pattern can now be heard from Irish student protestors–"No ifs, no buts! No education cuts!"–and it gained a new lease of life during the events around popular protests across the Arab world in 2011, where the words became "Ash-**sha'b** yur**īd** is**qāṭ an**-ni**ẓām**", or "The people demands the fall of the regime."

¡El **pue**blo u**ni**do, jam**ás** se**rá** ven**ci**do!

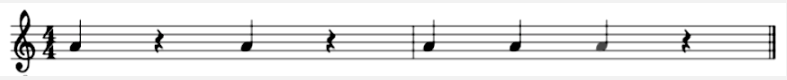

*Figure 9: Rhythmic pattern used in many protest chants. This rhythmic template is traceable back at least as far as Chile of the 1970's.*

With repetition, the irregularities of speech become more orderly. Short phrases become rhythmically accentuated. The beats associated with strong syllable onsets become more regular in time, and are frequently accompanied and exaggerated by co-produced gestures. In the more spontaneous settings of protest and football, such gestures may be fist pumps or hand claps, while in a ritual setting, they have quite likely become formalized into sequences of head bows, hand gestures, and changes of position (kneeling, standing,

etc.). Repetition also has the seemingly inevitable effect that the pitch contour of speech becomes more melodic. There is some evidence that mere repetition alone can transform the perception of a spoken pitch contour into what seems to be a sung melody (Deutsch et al., 2011). It is certainly the case that very well practiced verses acquire a stylized form of prosody that is quite unlike the sounds of conversational speech. We noted this with the introductory example of praying the rosary, and we might also recall the stylized lilt with which the Pledge of Allegiance is recited in American schools.

Can we draw a line, then, between speech and music, when we study joint speech? Certainly, joint speech seems to be continuous with some forms of music, and a strict speech/song division does not seem tenable. On the other hand, there are frequently good reasons to treat speech and music as rather distinct domains, serving different functions, and with markedly different requirements for participation. In Figure 10 below, I have mapped out a hypothetical continuum as a way of thinking about the manner in which speech becomes gradually more musical as we move from speech, conventionally considered, through various forms of joint speech, into territory that is more clearly musical.

On the right hand end, we have the inner voice, unobservable and somewhat mysterious (for who is talking to whom?). It is unclear to me, and I believe everyone else, how to adequately address the subject or subjects underlying such linguistic thought. The next point on the continuum is a situation rather like the preaching of Dr. Cosby (Example 3) or a classroom or lecture theatre in which one person does most of the speaking, but the presence of the others contributes to the event. This is asymmetrical, but it is nevertheless a collective activity. Nobody lectures to an empty classroom, or preaches to deserted pews, and it is very difficult to practice a public address in private. Depending on the conventions of the meeting, the interactions between the principal speaker in a monologue and the listeners may be more or less formalized, but there is always an obvious and important back-and-forth between one and the other. In that respect, the shared experience of the event is co-constructed, albeit with rather different roles for speaker and listeners.

Speech and music

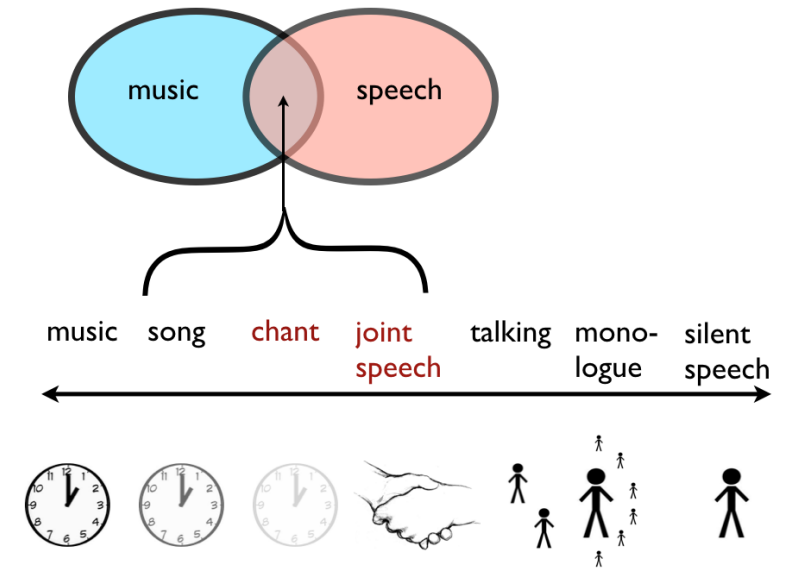


*Figure 10: A hypothetical continuum from speech to music, with associated consideration of the subjects involved.*

The figure shows a hypothetical continuum extending from silent (internal) speech on the right, through conversational and joint speech, and extending into the domain of music. The lower illustrations illustrate the changing nature of the subjectivities involved, with a reliance on external timing, or a beat, emerging as the sounds become more music-like. Participation becomes easier, but less solemn, as joining in becomes easier due to the beat.

The manner in which such an asymmetrical form of speech may nevertheless be co-constructed by speaker and audience together becomes starkly evident in bad play writing. On stage, if one character stops to tell a story, the listeners frequently become entirely passive, conveying that they are listening. This kind of scripting makes the storytelling on stage appear artificial and stilted, the collective equivalent of not knowing where to put your hands while standing, unexpectedly, in a spotlight.

Moving one place further left on this notional continuum, we encounter everyday informal conversation. This situation is illustrated on the left in Figure 4 (p. 36) where we saw that conversation is characterized by a con-

stantly shifting common ground that alters with each contribution to the joint event. A conversation is rather more obviously a shared event as the participants are typically taken to be equal contributors to the joint project. This apparent symmetry is, of course, frequently broken in real life. Furthermore, there are many cultures in which this apparently level playing field is structured differently, with differing roles based on age or social status, for example. Nevertheless, conversations are genuine co-constructions, but highly dynamic ones. They are intrinsically dialogical. Each utterance builds on, and responds to, that which went before.

Before we get into musical territory, it would be as well to point out something important about speech rhythm, loosely conceived. While we may use the notion of the "rhythm of speech" in an everyday sense, conversations do not typically have a beat-based structure, and evenly spaced intervals, so common in music, are not a reliable feature of speech (Cummins, 2015). There have been many claims that the onset of one person's contribution is timed to occur based on a sequence of beats (e.g. syllable onsets) in the speech of the other (Couper-Kuhlen, 1993). Although such studies keep cropping up, there is no convincing evidence that this is the case. Indeed, Stephen Cowley has rather convincingly argued that the dialogical nature of conversation is poorly described in terms of "turns" and "turn-taking" (Cowley, 1998). In a game of chess, one turn follows another, and it is clear in advance when one turn runs out and when the next must begin. Contrast that with the sequence of alternating punches in a boxing match. Here, each punch responds sensitively to everything that goes before, there is a sequence of blows, but it is obvious that the boxers are not "taking turns" swinging their fists. In a conversation, everything is negotiated, constantly. Every word uttered is in response to the preceding utterances, gestures, glances, postures, and events. This unpredictability makes conversation a means for joint sense-making, and prevents it being merely a formal dance with alternating moves. It also prevents it from being music.

The central point on the continuum is illustrated using a handshake, rather than a speaker. This is joint speech, before repetition and stylization conspire to move it into musical territory. This is the commonality of a credo or of an oath of allegiance. Here, common ground is maximized. The message being spoken is meant by everyone, everyone vouches for it, and it gives rise to a common perspective, or a shared way of facing the future and the world. The handshake is intended to suggest the collective nature of the speaking. A handshake cannot be reduced to the sum of two hands. They are in constant contact, or to use a phrase we met before, there is real-time reciprocal interaction among the hands. The owners of the hands

are necessarily co-present to each other, and the handshake cannot be done by mail or text message. The handshake will provide a convenient iconic symbol of non-dissociable commonality in what is to come.

Now, having entered the world of joint speech, we move further to the left, towards the musical end of the continuum. I have made use of the word "chant" to signal that this speech has musical elements, whether it be the enhanced rhythmicization or melodic exaggeration occasioned by repetition, or indeed, the use of melody pure and simple. Chanted music is very akin to prayer. It is a common form of worship and practice found in monastic communities in many faiths. The monks of the Greek Orthodox monastery on Mount Athos speak of God being among them when they chant. The chanting practices of Benedictine monks have much in common with the kirtan of the Hindu monks of ISKON (the so-called "Hare Krishna" devotees). Chant in any of these settings has some surface-level characteristics that differentiate it from musical practices more generally. Indeed, in some traditions, this kind of singing is not covered by the term used to describe "music" played for the purposes of entertainment. The austere world-denying outlook of the Wahhabis, whether they are clerics in Saudi Arabia or jihadis in the Caliphate, professes to hate music. They burn instruments, and persecute those who play and enjoy music. But their propaganda is filled with unison chant, which they see as an entirely distinct form of activity.

Chant of this sort is sung in unison. There is usually little in the way of accompaniment, and nothing that would draw attention to the instrument over the voice. The phrasing of the words that are sung comes from speech, rather than from poetry, so we do not find strong signs of metrical organization. Successive musical phrases may be of very different lengths. This contrasts with most kinds of music in which an underlying beat is organized into hierarchically nested groups of invariant size, giving rise to waltzes, foxtrots, or the ubiquitous 4/4 beat of rock and pop music. In Gregorian chant, in plainsong, and in related traditions, there is an underlying beat or pulse, but no meter. Polyphony, or the presence of different melodic voices at the same time, is not found. Indeed, the absence of polyphony might be taken as the principal identifying characteristic of chant.

If we now move further towards song and music, these restrictions vanish. Multiple voices are found in parallel. Sung lyrics are organized based on the strictures of hierarchical units such as bars, verses and choruses. Rhythm is fully elaborated, with alternating strong and weak beats in regular succession. From chant to music, I have replaced the underlying images of individuals, collectives, and handshakes, with a single image, coming more and more into focus–a clock. The weak beat of chant is replaced by the

obvious and reliable structures of metrical rhythm. Something important happens with the introduction, first of the beat, and then of meter. It becomes easier for anyone to join in the synchronized activity, but at the same time, participation becomes more trivial, lacking the urgent commitment of speaking together. Joining in the recitation of the Credo is costly and commits the utterer, in the sense that it establishes, and licenses, a normative framework through which the subsequent actions of the speaker may be judged. The speaker may thereafter be a good Catholic or a bad Catholic, for example. Mouthing the lyrics of a Bob Marley song on the dance floor entails no commitment to Rastafarianism, or anything else.

If we think of speech as categorically distinct from music, we miss this gradient that runs from the immediacy of the experience of inner sdpeech alone, through a variety of ways in which we co-create meaning, culminating in the civilization-founding pronouncement in unison of deeply held beliefs, and then into forms of mutual coordination that become both more open, but also more lightweight. Not all activity can, or should, be as charged as swearing an oath or reciting a credo. The continuum is sketched tentatively above. But it seems to demand some further justification. We can clearly recognize different communicative contexts. It is also obvious that some will be more musical than others. But on what basis is this a continuous dimension of variation? What exactly changes smoothly as we move from right to left.

The short answer is: I don't know. The continuum as sketched above poses a challenge to make sense of the way in which speech and music become entangled, and of how music enters into practices of joint speech and changes them. But there is a better answer we can give in the context of this book. That which varies from right to left is nothing other than the subjectivity involved. We start with the solipsistic introspection of the Cartesian skeptic, which seems to underlie the only scientifically blessed form of subjectivity, the separate and distinct mind that is detached from the world. As we move across the continuum to the left, then, the subject changes. It immediately becomes something collective, even in the case of a monologue. But the subject is not fixed here. It can be thought of as enacted, or brought into being, in the dialogical back and forth, between listeners and speakers. It reaches its most stable form in the middle of the continuum, where musical elements are still foreign, but where all that is uttered is uttered together. Then, as music enters, the subject bleeds out, and the objective scaffolding of the repetitive beat within a fixed meter, becomes that which holds it all together. The continuum is a challenge to understand how there is never just one kind of subject, but that subjectivities are enacted, brought into

being, in may ways, with different forms of commitment, different forms of commonality.

# Chapter 5

# Joint Speech and Movement

Speaking in unison is one among many kinds of synchronized behaviors humans are capable of. In order to consider it as a special case of synchronized movement, we had better clarify these terms, because both synchronization and behavior are words that are prone to misunderstanding. But of the two, synchronization is by far the easier concept to discipline. To many people, scientists, engineers, and non-experts alike, two processes are synchronized if they seem to be linked or strongly related as they change over time. On this view, all the planets in the solar system are highly synchronized, or to be a bit more literal about it, they share time, and so constitute a kind of clock. This broad approach to synchronization works fine for many purposes, especially in the study of inanimate objects, but I will need to make use of a rather more strict definition of synchronization that may best be illustrated by comparing a group of line dancers with a couple dancing the tango.

Line dancing, in its common cowboy form at least, requires that each dancer do the same thing at the same time. When one dancer makes a right turn, everyone is making a right turn. If one leg is raised, everyone raises one leg. I will be happy to say that these dancers are synchronized because they are doing the same *thing* at the same *time*. This may sound unexceptional, but it is not a common definition. It demands that we be ready to discuss just what we mean by both *thing* and *time* for any given instance. It draws distinctions that are necessary in the context of this book, and my own work generally, but it is not quite what most people mean by synchronization.

**Synchronization and coordination**

*Figure 11: Top: Line dancers. Bottom: Couples dancing the tango.*

In the present work, we will use the term *synchronization* to refer to people doing the same *thing* at the same *time* (as above, top), while *coordination* will apply more broadly to collective movement that belies non-independence of the participants (bottom).

The distinction I wish to draw is clear if we consider as a second case, the couple dancing the tango (Figure 11). At any given time, the two dancers are exquisitely coordinated. Their movements are entirely non-independent. Each dancer accommodates and responds to the other sensitively, in real time. Indeed, they exhibit a lot of the *real time reciprocal interaction* that we have begun to draw attention to. However, for our present purposes, I will choose not to consider the tango dancers synchronized, in contrast to the line dancers, even though, in my experience at least, the line dancers are typically far less coordinated.

To be quite clear, I am not suggesting that the man and woman dancing the tango are in any way uncoordinated. But each dancer within a pair

has a distinct role. The man leans over, the woman leans back; the man extends an arm, the woman twirls, and so on. By way of contrast, if we were observing a dance floor full of tangoing couples, this strict definition could apply to the couples within a room, but not to the individuals within each couple. In that case, we would say that each couple is doing the same thing at the same time. The synchronizing unit would be the couple (often called the dyad in movement studies). With that, I hope we have a solid definition of synchronization with which we can proceed.

## Behavior versus movement

Behavior is much trickier to define. There are so many definitions and distinct opinions on the matter that I will be forced, once again, to use a somewhat strict and slightly unconventional definition, and this may not align well with many other productive uses of the term. I ask for indulgence, as the definition adopted here plays an important role in the picture of joint speech. It does not rule out other ways of approaching the large topic of behavior. As with the previous term, an illustrative example will help to get us started.

Suppose you observe my right arm, unmoved by external forces, raise to the right side of my face, so that my right index finger brushes against the cheek. In an everyday situation, this would be a rather unremarkable event, but it would be an event. It would be me scratching my cheek, for whatever reason. Perhaps I had an itch. We might reasonably call this specific movement of my hand a behavior. It might happen while I am engaged in other activities, such as writing or talking. Interpreted as a behavior, this description would not be an exhaustive account of what I am about at any given moment. To describe my hand movement as a behavior ("scratching an itch") is to parse the continuous flow of movement of my body and the world in a specific manner that makes this particular movement intelligible.

Now suppose I were to suffer a grand mal epileptic fit. This is a pattern of abnormal electrical activity within the brain, leading to violent and uncoordinated movements of the body. Among those movements, it is entirely thinkable that my right arm might rise to my right cheek and go through exactly the same motions as before. In this case, I want to suggest that we should not characterize that movement as behavior. It is certainly the case that calling that "scratching an itch" would be misleading; even "scratching" seems a step too far; such a characterization would not contribute to the in-

telligibility of the scene more broadly considered. We might agree that here the same movement requires a different framing to be understood. That framing includes the attribution of purpose, intention or unseen agentive cause by us, the observers.

And so when we come to study, describe, or account for behavior, some framing has gone on in advance that serves to parse the continuous flux of the world in specific ways. Those movements that are framed in this manner, that we might label as behaviors, are very real, but their characterization as behavior is not independent of the observer. Specifically, the use of a behavioral label ("he is *walking*," "she is *scratching*," etc.) becomes possible when the observer (the person using the label) has identified a purpose, function, or goal that serves to make the observed movement intelligible. I hope this appears strongly reminiscent of the discussion of the function of the heart, which likewise required a specific kind of framing. In that case, understanding the movement of the heart as *pumping* made it more intelligible, but that required framing it in the context of a body for which this pumping can be of significance.

So we will consider behavior as goal-directed movement, that is, as movement that becomes intelligible when we recognize some system or entity *for whom* the movement is serving a function, advancing a goal, or is of significance. In this manner, we remain quite uncommitted as to the system involved, and we will exploit this flexibility as we recognize different behaviors for different kinds of subjects. But there are many kinds of activity we will exclude, most importantly those activities that do not present as movement that can be made intelligible through the ascription of goals or function. So some things psychologists might consider behaviors, specifically "thinking," "remembering," "planning," "problem solving" and the like will not be considered here, as they are not to be found by parsing movement in any specific way. This is in keeping with our strategy at the outset, of staying close to the surface of things, and starting with observations upon which we can agree. There is no such uncontroversial observation we could make for the ill-defined business of "thinking," for example.

It might be obvious that the definitions we just adopted are not independent of each other. Because we will be interested in synchronized behaviors, we need to pick out those behaviors (goal-directed movement) and in so doing, we have delineated the *thing* that is potentially done at the same time by several individuals.

## Synergies

Throughout the 20th Century, movement science and scientific psychology carried on their respective inquiries with relatively little reference to or awareness of each other (with a few notable exceptions such as Ecological Psychology (Warren, 2006)). We will address the antipathy of scientific psychology to movement in a little bit. First, though, it is time to introduce one of the most important discoveries from movement science, a discovery that has informed all subsequent work in the field. The work in question dates back to the 1920's in newly Soviet Russia, where the physiologist, Nikolai Bernstein, developed a method of accurately recording the movements of people carrying out specific tasks. In his best-known study, he examined the movements of blacksmiths hammering a chisel on an anvil (Bernstein, 1930). To do this, he had to attach small light bulbs to various places on the body, and then use cutting edge photography to capture up to 500 images per second as they hammered. The movement tracks had to be then reconstructed through manual measurement from the photographic images.

The principal finding of Bernstein is one that continues to inform all of the study of coordinated action. It is a discovery that greatly changes what we think about skilled action, and any account we might like to give of control of that action by a subject. He found that the blacksmith, together with the hammer, exhibited movement that was entirely intelligible if the blacksmith+hammer were considered as a purpose-built hammering machine, that is, as a mechanism built for one task—hammering. He did not find evidence for one part of the body playing the role of issuing instructions while other parts follow those instructions. This distinction is important, and may be familiar to engineers as the difference between open-loop control (where the brain is interpreted as a controller, the body as the controlled entity, and the anvil and the hammer as outside the system) and closed-loop control (where the brain, body, hammer, anvil together conspire to achieve the hammering goal, without any partition into controlled and controller). This latter constellation is known as a synergy (or, elsewhere, as a coordinative structure).

That the blacksmith and his hammer have become temporarily organized into a purpose-built hammering system is evident from two empirical observations. It can be seen, first, through an analysis of the variability, including noise, or error, as measured at various points from shoulder to anvil. It is also to be seen in the manner in which such a system responds to an externally induced perturbation. We will consider these in order.

If the brain controlled the arm (open loop control)

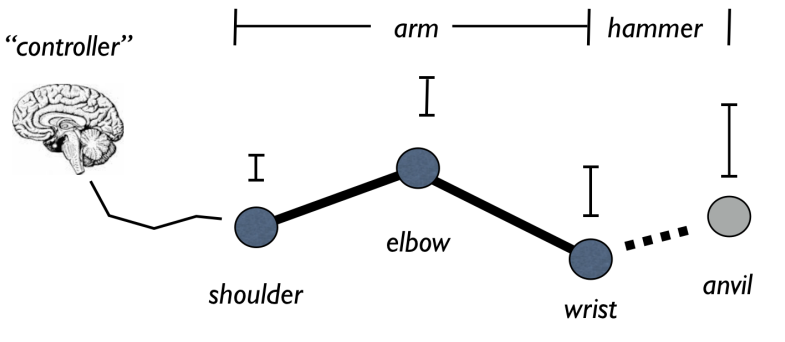


Figure 12: Here we lay out from left to right the elements in a series from brain to anvil. If we were to interpret the brain as the controller, and the body as the controlled entity, then there are several joints between controller and the final point of contact with the anvil. At each joint, there must necessarily be some degree of imprecision, however small, as this is a biological meat machine we are considering and not an abstract mathematical structure. Because this particular system consists of a series of linked segments, any imprecision or deviation from a desired value that arises at the shoulder will be propagated down the system from left to right; an error in shoulder position must introduce a corresponding error in elbow, wrist, and point of contact. But some degree of error will necessarily be attributable to both elbow and wrist too, as these represent distinct control challenges. As a result, the sequence of linked segments, notionally driven from one end by a brain, must exhibit variability that increases from left to right, being maximized at the point of contact. Indeed, the brain, on this account, does not control anything beyond the hand directly, and there is certainly no direct link between nervous system and anvil. This is emphatically not what was found.

Imagine, if you will, trying to recover your keys that have fallen down a drain covered by a grill. You find that you can only get your fingers through the grill, but they are too short. So you elongate them by tying a pencil to your longest finger. You still can't reach them, so you tie a second pencil onto the first, with a movable joint in between. By the time you have tied three pencils together, you have an appendage that is long enough to reach, but well-nigh impossible to control, as every joint, every potential degree of freedom, complicates things, and small errors in finger position become larger errors at the far end of your assemblage. This way of directing the movement of the fingers+pencil system is also known as

"open-loop control," as the influence of a controller is distinct from, and not in immediate contact with, the distal point at which the target of the control resides (see Figure 12). This is not what Bernstein found in his analysis. He found that variability was minimized at the point at which the hammer met the anvil. That makes excellent sense, as the purpose of the activity is best expressed at that point. It is well nigh irrelevant how the hammer gets to the chisel, but the task at hand demands that it be aimed precisely once it hits. This finding rules out an interpretation of the relation of brain to body and tools as one of controller on the one hand and controlled on the other, suggesting instead that the body (including brain) and the tools and the surrounding surfaces are linked into a single non-decomposable system in which all parts of the system work together.

The response of the system to perturbation is also enlightening. A perturbation is any unpredictable externally applied force that alters the configuration of the system. In an open-loop system, any perturbation must be compensated for by the controller, who is tasked with implementing a response to the deviation from the desired state or trajectory. Perturbations to a synergy, on the other hand, lead to distributed compensatory responses throughout the system, which collectively act to stabilize the overall purposive end of the system. That is, a perturbation to a synergy evokes a functionally specific response. This is perhaps well illustrated by the response of a skilled footballer whose goal-directedness (literally) is perturbed (severely) by the opposing players. As each leg or arm is interposed to stop the attacker, the footballer's whole body reacts, so as to try to ensure that the overall goal (a goal) is reached. If one observes the twists and turns of a successful run towards a goal, every flick, feint, and lunge makes sense in light of the organizing influence of the (literal) goal.

Another dramatic illustration of the immediate distributed compensation we find in a synergy is provided by a centipede. If the legs of a centipede are removed in pairs, the resulting gait of the centipede remains fluent, effortless, and smooth. At no point does it look as if the centipede is learning to control a novel number of appendages. This fluid self-organization of many parts into a smooth, goal-directed whole, is the hallmark of volitional movement in biological organisms. Deliberation, consideration, thinking, are all rather more at home in the static, unmoving subject.

A centipede was happy—quite!
Until a toad in fun
Said, "Pray, which leg moves after which?"
This raised her doubts to such a pitch,
She fell exhausted in the ditch
Not knowing how to run. (Attribution uncertain)

**Synergy**

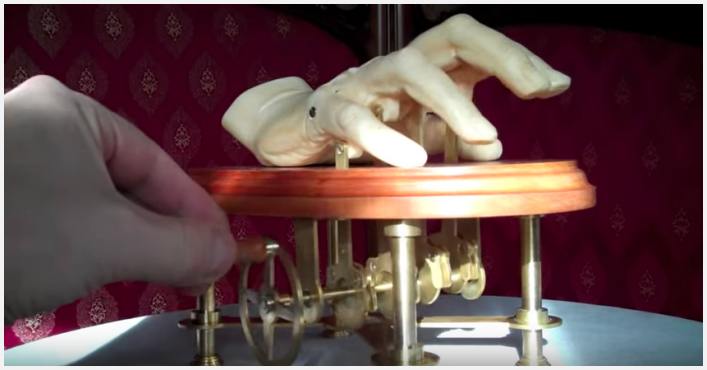

*Figure 13: An automaton illustrating the notion of a synergy.*

The automaton here has been designed by an engineer, and it admits of only one configuration. A real hand may fluidly adopt a similar configuration, and temporarily become constrained to move in the same fashion. When the hand and fingers are so constrained, the owner has essentially a single degree of freedom: You may drum your fingers more slowly or more quickly, but you can't readily alter the sequence of fingers. You may, of course, stop doing the task at any time. Skilled movement is always better described as a form of *coordination* among the various parts of the body and world than as the effect of a controller (brain) on a controlled body.

Synergies arise fluidly and they dissipate again. They are not co-extensive with a single body, and certainly not with the more loaded notion of the person. When I drum my fingers impatiently on the tabletop, it is the hand and the fingers and the tabletop that together make up the elements that exhibit such coordination. Figure 13 shows a beautiful automaton that makes explicit the temporary linkages among the various parts that arise when the fingers are drummed on the table top. The automaton, like the real hand,

is constrained during this behavior. For me, there is essentially only a single degree of freedom left. I can drum faster or slower. I can't really alter the sequencing of the fingers. Similarly, the automaton is controlled using a single controller, the rotating wheel with the handle on it, that regulates the overall speed, and nothing else.

And so in movement science we have about a hundred years in which we have observed synergies arising in a task-specific manner (Kelso, 1995; Latash, 2008). Some synergies are made up of parts of the body; some of the whole body; some of the body plus tools. Usually, the synergy includes some reliable elements in the environment, such as support surfaces. Sometimes we can recognize a synergy in the collective activities of several people—indeed, we will shortly encounter just that when we examine synchronous speech from this angle. But look at what we do not find: We do not find a singular subject who directs and instructs the body. Each synergy constitutes a goal-directed organization of many parts, but such a system is enacted, not simply existent, and it is not co-extensive with the person. It comes into being through the coordinated activity of the parts, and it exists as long as the parts work together in the service of the goal. Then it goes away again. The hand, no longer constrained to drum on the tabletop, is now free to write a letter, scratch an itch, or grab a mug of tea.

## Wu wei, or whodunnit?

Within the Western intellectual sphere, the most influential accounts of skilled movement lean upon the notion of the psychological subject, who is assumed to be the sole agent, to exist independently of context, and to be causally responsible for its own actions. This creates a considerable explanatory burden upon such accounts when it comes to skilled action. The hallmark of skilled action is the apparent absence of ego. The process of skill acquisition is a gradual progression from clumsy, individual, uncoordinated bits whose serial execution requires great attention, to smooth, coordinated movement in which no conscious control is exerted over individual movements. A challenging passage at the piano appears first as a sequence of individual notes that must be laboriously hammered out, one after the other, with errors, and without grace. The same phrase, when polished and mastered, almost plays itself. Its component parts are no longer separate from one another. The only input by the performer is to modulate the tempo or the intensity as a whole, but the phrase has made itself largely independent of the player (Sudnow, 1978).

Wu wei

*Figure 14: The Chinese characters for wu wei, or "non-volitional action" displayed above the Emperor's Throne in the Forbidden City in Beijing. (Image credit: User star5112, Wikicommons, CCA-SA 2.0)*

The idea of non-volitional action refuses the notion of an autonomous agent who is separable from context and is in executive control of action. From a Western perspective, this may appear as a challenge to the notion of free will. It is nothing of the sort, but not because it claims that the person does, or does not, have free will, but because it instead negates the notion of the entirely autonomous person of whom such a claim could be made. If no such entity exists, then there is nobody to have, or not have, free will. The graceful yielding and blending found in the martial arts illustrates this graphically, as does the movement of a fish in turbulent water. The fish does not fight the water, for that would be worse than useless. The fish+water form a kind of unity that is not brought about by any individuated locus of agency. The consequences for imperial rule are left as an exercise to the reader.

If we move beyond the Western/Christian tradition, things look very different. In the Daoist tradition, for example, it is generally understood that the idea of the control of movement by an autonomous agent is an illusion. To understand the form of movement is to see movement as arising spontaneously, without any notional intervention by a supposed controller. Smooth spontaneous action is called *wu wei*, which is often poorly translated as "doing nothing." A much better translation, from the Anglo-Irish scholar Terrence Gray, a.k.a. Wei Wu Wei, is "non-volitional action." For it is not that nothing gets done. It is rather, that there is no *do-er* behind the action, there is nobody who sets the goal. He describes the Daoist view rather well in the following passage:

> **He who gets slapped**
>
> When I was a child I was taken to the circus. There I saw a long series of entrancing performances that caused men and animals to execute every kind of astonishing and unexpected maneuver. And throughout, but particularly when the scenario and its appurtenances were being changed, there appeared a grotesque personage, vaguely resembling a human being, who interfered with everything but effected nothing. He fell over the carpets, bumped himself against every object, was slapped and kicked, and then took all the applause as though he were responsible for everything. We thought him very funny and laughed at him like anything.
>
> Now that I am no longer a child he seems to me to be a perfect image of the I-concept, [...] whose performance corresponds in all respects with that of the clown, in the circus which is our life. In all respects but one: we laughed at the clown in the circus, but we take seriously the clown in the circus of life, although the one is as ineffectual as the other. We even believe that he is responsible for the performance, whereas as children we could see that he was responsible for nothing that happened, that his 'will' was totally ignored by the circumstances to which he was subjected, and that in every event he was an unnecessary nuisance.
>
> In one respect, however, our attitude is unchanged: in both the circuses we love the clown dearly and consider him more important than anything else in the show. (Wei Wu Wei, 2002)

The "I-concept" referred to here is, of course, the psychological subject that underlies all Western thinking about the person. This concept does not travel well. The Buddhist notion of *anatman*, or "no-self" is the assertion that no amount of looking will allow you to pin down a soul, a mind, or a psychological self. The degree to which an individual self (*Atman*) may be considered distinct from the ground of all being (*Brahman*) is a constant concern of all schools of Hindu thought. The Daoist notion of *wu wei* suggests that movement is not to be understood with respect to an underlying controller who is in charge. It will not be helpful if we adopt a Western view of Buddhism, Hinduism and Daoism as different "religions." The philosophical underpinnings of such traditions are venerable, cross-cutting, and highly sophisticated. Within all of those traditions we can find sizable literatures, and traditions of debate, that revolve precisely around the nature of action and its relation to various ways of conceiving of an individual self. Within such discussions, the psychological subject and its relation to action appear as just another variant on some well-worn themes that do not resolve into a single positivist account.

When we observe a goal-directed action, such as a hand drumming on a tabletop, or a blacksmith wielding a hammer, we are liable to get into trouble if we insist on asking: To whom should we refer the goal? This is a very serious challenge for any science of behavior. Teleology, or the postulation of goals, is necessary if the goings on of the animate are to be intelligible. We noted this with respect to the heart, where I think the ascription of function (and hence purpose) is uncontroversial, given the shared framing assumption of the relevance of the continued integrity of the body. But the ascription of goals is far more controversial when it comes to the classical territory of psychology, including all discussion of volition, intention, and purpose. Psychological science has relied on a notional subject who acts as controller, who exists independently of context, and who persists from birth to grave. Movement science finds no evidence for this. The insights of movement science speak rather of the enactment of temporary domains of autonomy, dedicated to this goal or that. Usually, we speak of behavior as if there were a person, and hence a mind, behind them ("Johnny is playing football"), but this entirely conventional summary is best thought of as an informal account of activity generally, suited to everyday conversation. When we become more careful observers and examine the form of movement, it frequently resists this kind of personal description. When two dancers dance the tango, we see an obvious synergy at the level of the dyad, not the individual. When I drum my hand on the tabletop, the synergy lies in the

hand (and table). When Johnny plays football, we might observe synergies arising and disappearing in many combinations of legs, bodies and the ball.

Now we must acknowledge once again the role of the observer in a behavioral description. When we recognize that these parts (hand, fingers & table top; two dancers) are mutually coordinated in a fashion that is intelligible only through the postulation of a goal, we necessarily implicate the observer in the analysis of the situation. It means that this kind of description cannot pretend that it is continuous with science done in a purely objective mode. Observing behavior is not like observing the stars, and an empirical science of behavior must, it seems, demand some degree of tentativeness and indefiniteness, some recognition that assertions that are made are enabled by unstated framing assumptions, and will be valid only to the extent that we share those framing assumptions. Such a tentative science, it seems to me, must be dialogical, and cautiously negotiated, rather than pronounced and fixed, a never-ending process of affirmation and correction, rather than a finalized product.

The study of joint speech has much to contribute here. Firstly, as a form of movement that allows synchronization, it bears some unique features that distinguish it from all other forms of synchronized movement. Secondly, in a laboratory context, two speakers are found to become temporarily organized into a dyadic (two person) synergy while they speak in unison. This is demonstrated by the occurrence of a specific kind of speech error unique to the experimental context. We will examine both of these in a moment. But perhaps the greatest contribution that the study of joint speech can bring to this discussion is to point out the limitations of the psychological subject, and the manner in which any assumption of such an entity blinds us to the fluidity of our being, in which we partake in many kinds of collectivities which we are not distinct from. Joint speech offers a way into a consensus-based examination of the manner in which many kinds of subjects arise, and many kinds of foundations are laid. In order to see this, it will be necessary to come back to the assumed subject of psychology and to examine its tortuous relationship to movement.

## Movement and the psychological subject

The science of movement has existed on the periphery of the psychological, cognitive and social sciences, without ever finding integration into those fields. Psychology students are typically spared any direct contact with movement science, which seems odd, as psychology was founded with the

dual (and perhaps irreconcilable) goals of providing a science of behavior as well as a science of the foundations of experience. But behavior and movement are not the same thing, of course, and recognizing that already threatens any simplistic notion of a purely objective science of behavior. The psychological subject that finds no support in the movement sciences turns out to be constructed in a manner that is positively hostile to movement! Let us have a look at how movement features in the scientific activities that serve to build up a picture of this problematic subject.

We will first have a look at the kind of topics taught to psychology students. Here are the chapter headings from a classic introductory textbook (Eysenck and Keane, 2000), which serves to illustrate the structure of the discipline of cognitive psychology:

- Visual Perception: Basic Processes
- Perception, Movement and Action
- Object Recognition
- Attention and Performance Limitations
- Memory: Structure and Processes
- Theories of Long Term Memory
- Everyday Memory
- Knowledge: Propositions and Images
- Objects, Concepts and Categories
- Speech Perception and Reading
- Language Comprehension
- Language Production
- Problem Solving, Puzzles and Expertise
- Creativity and Discovery
- Reasoning and Deduction
- Judgment and Decision Making
- Cognition and Emotion
- Present and Future

Movement appears in one chapter among 18. Almost all of the chapters deal with goings on that are not directly observable by any means whatsoever. Upon examination, the treatment of movement is slight, and the

text fails to engage with any research into coordinated movement, or with movement science at all. The syllabus also mentions speech and language, but it does not treat of speech as a phonetician might, as a form of coordinated movement. Its concern is rather with the supposed symbols or atomistic units hypothetically underlying the observed movement and air vibration. Cognitive psychology has consistently regarded movement as an outcome, while it concerns itself with supposed processes going on in the background, usually taken as meaning in the brain. Rather than examining movement and working backwards to a best account, it starts with goals, intentions, purposes, and hypothetical underlying machinery required to implement them in a body that is considered to be some kind of machine in need of control. That control is provided by the hypothetical psychological subject.

But there is something very odd about the manner in which evidence is gathered for this view of the psychological subject. On the view taken by cognitive psychology, the subject has an enduring existence; it is not temporarily enacted, as a synergy is. The psychological subject is considered to be the same thing as the person, housed in a singular body, existing independently of context, and extending from birth (or before?) to grave. The presumed architecture of the psychological subject is located in an abstract interiority between perceptual input and action-oriented output: Perception provides input, feeding processes of cognition in the middle, which supplies commands generating action as output. This basic structure has underpinned almost all inquiry in scientific psychology, and has provided the basic terms with which we address our individual being in every day conversation. It seems innocuous, to the extent that it can be hard to see that it is a story constructed on questionable foundation.

The view of the minded subject as an unobservable interiority receiving perceptual input and controlling a body in the manner of an executive has had its strong critics all along. There were a few at the time of the birth of psychology (Dewey, 1896) but there are very many now, loosely grouped under the heading of "embodied" approaches to mind and behavior (Varela et al., 1991; Shapiro, 2010; Chemero, 2011). We can contrast the two broad approaches, and the stances taken by them with respect to movement, if we consider how each treats of the process of seeing.

Vision is the richest of domains. We understand ourselves as visual creatures first and foremost, and discussion of experience, immanence, or consciousness is saturated with visual metaphors in every culture. To see something is to be in its indisputable presence. Only the most confident connoisseur will trust her nose, but we all trust our eyes and take great

delight in the world of optical illusions and visual trickery. The domain is too rich to explore properly here, but let us look in particular at the way that movement features in our understanding of vision.

**Vision, immobilized**

*Figure 15: Left: An anesthetized cat, as used in the neurophysiological experiments of Hubel and Wiesel. Center: Involuntary visual exposure in the film A Clockwork Orange. Right: The ubiquitous fixation cross used in most psychological investigation of vision.*

Perception, and vision in particular, as understood within cognitive psychology, involves constructing a representation of the "external" world based on input from the senses. Any such account is greatly facilitated if the subject in question is constrained not to move. The consequences of this unacknowledged limitation include an apparent separation of the experience of the subject from the world it inhabits.

In 1981, David Hubel and Thorsten Wiesel got the Nobel Prize in Medicine and Physiology for their influential and foundational work on the activity of nerve cells in the brain related to vision. They were indeed pioneers. They were some of the first scientists to use the very fine electrodes that were necessary to record from individual nerve cells. The very first such electrodes were at that time created by crafts people capable of drawing glass out to a fine tip far thinner than a hair. The extreme thinness was necessary because the tip of the electrode had to be introduced through a cell membrane to record from the inside of the cell without damaging either cell or electrode. It goes without saying that the electrodes were fragile in the extreme. By the late 1950's, electrodes might also be made out of very fine tungsten wire. In their most famous experiments, they used anesthetized cats as their subjects. The experimental procedure being used demanded repeated precise projection of visual stimuli onto the retina of the cat, and so it could not support any movement on the part of the animal. These unconscious and inert cats had their eyes propped open (echoes of

the treatment of Alex in A Clockwork Orange are not far from the mark. See Figure 15) and visual patterns were presented in front of the cat while recordings were done from individual cells in the retina, the lateral geniculate nucleus and the primary visual cortex. In this manner, the responses of the cells were interpreted as if these were the building blocks from which an image-like representation of the visual scene in front of the animal was to be constructed. The cells closest to the retina were found to be sensitive to very local and simple properties of the visual "stimulus," and as cells were explored deeper within the brain, they seemed to respond to more complex, higher order, and more meaningful properties of the supposed visual world. It was almost as if peering into the brain from the sensory periphery, one penetrated a vast Cartesian interiority, populated with the perception of an outside world.

These pioneering experiments were the start of a substantial industry, in which the role of the brain was cast as extracting information from the play of light on the retina, leading to the construction of a representation of the world. The cat was assumed to see through this form of mediation. The representation stood in for the world. In a summary article from 1979, the two scientists review a great deal of work in mapping this presumed mapping from the visual field, through projection onto the retina, and onward into the interior of the brain, assuming at each point that the visual field is presented as if it were a static image (Hubel and Wiesel, 1979). The cat is not only immobile; it is insensate. It is not engaging with the world at all. It is not behaving in the world and negotiating its own relation to the surrounds. This is the basis on which a baroque inner Cartesian theatre is constructed.

But real cats move. In the same year that Hubel and Wiesel won the Nobel Prize, Hubel published a paper with Margaret Livingstone which revealed that all the response properties of the cells in visual cortex were greatly altered when the cat regained some level of consciousness (Livingstone and Hubel, 1981). They still weren't allowed to move though. Their heads were taped in place, and they wore contact lenses to focus an image on the retina. Nevertheless, allowing even this little breath of life back into the cats showed that the basis for the representational story for which the Nobel Prize had been awarded depended sensitively on the stillness and inactivity of the cats.

In the 1960's another set of experiments on cats showed something else—that self-initiated movement is essential for the development of vision (Held and Hein, 1963). Held and Hein allowed kittens only very restricted movement opportunities. The kittens were examined in pairs. In each pair, one

kitten (A = "active") got to move in a harness tethered to a central pillar (Figure 16). The other kitten (P = "passive") was passively moved as the first one walked. All walking took place inside a cylindrical chamber so that the visual stimulation received by both kittens was as nearly matched as possible. The kittens spent 3 hours a day in this chamber for some weeks. The A kittens developed normally by the measure of the tests employed (paw placement, response to a visual cliff, etc.). The P kittens, on the other hand, did not behave as normally sighted kittens. They had not learned the relation between activity and seeing.

Held and Hein's experimental setup

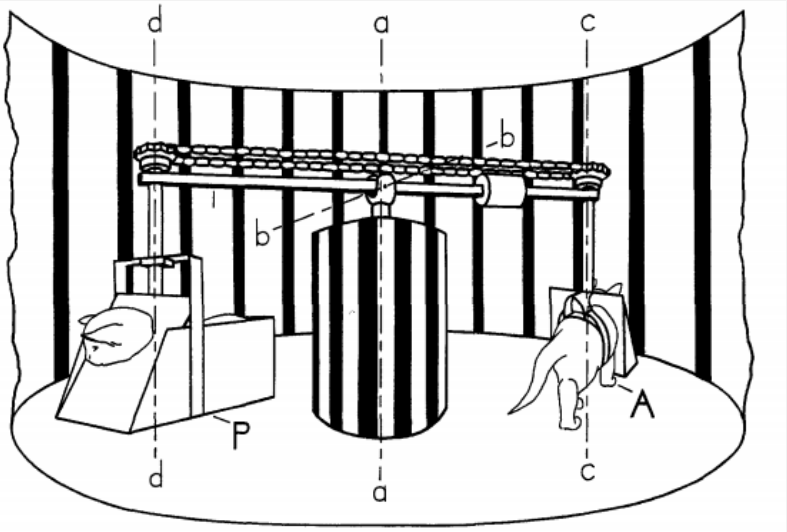


*Figure 16: Apparatus used by Held and Hein in which one kitten (A) is active, and the other (P) passive, though both receive comparable visual stimulation (Held and Hein, 1963).*

There are very many schools of thought when it comes to vision. Some of them insist that seeing is mediated by images. Others dismiss such talk as incoherent. We do not have time to sort such matters out here. But we do have the opportunity to make a crude distinction between two fundamentally different approaches: those who understand vision to be the process of constructing a representation of the world that informs a subject, and those for whom vision is an activity. The former camp subscribes to the psychological subject, whose inner world is populated by "information" extracted

from an ambient world through the eyes (Marr et al., 1991; Vanrullen and Thorpe, 2001). The latter study instead the relationship between a moving entity (person, organism) and its environment, and seek to uncover the role of the patterning of light on the retina in that dance (Gibson, 1979; O'Regan and Noë, 2001).

The methods employed by the first group are instructive. Visual experiments are typically conducted in front of a screen. The subject is told to sit still, and stare straight ahead. To ensure that they are "behaving correctly," a fixation cross is placed in the middle of the display. This is where the subject is told to look. Given the assumptions of representational cognitive psychology, this all makes sense. Movement complicates the business of seeing, if seeing is considered as the scanning and interpretation of images. But to those who understand vision as a kind of activity done by a moving subject, it appears bizarre.

It is not only cognitive psychologists who understand vision in the rather restrictive sense that imagines images beamed into the brain. Most of us are familiar with optical illusions. Every year there is a competition for new variations on this theme, and sometimes even genuinely new illusions result. However very many of the visual illusions work only if the viewer follows the instructions and stares at a fixed point, usually located at the center of the screen. Once the eye is allowed to wander, the illusion goes away. Nobody seems to doubt that the visual system is being tricked by such illusions. But is it not the viewer who is being subtly tricked by being instructed to suspend normal vision (active) and to hold the head in a pathologically still state? I am reminded of the anamorphic artworks of Felice Varini (Figure 17) in which a specific intelligible form comes into being only when a single eye is held steady at a fixed point in isometric space, and it disintegrates into diverse pieces as the viewer moves. This allows the coherent perceptual gestalt to be photographed from that single point, as a camera does indeed capture the play of light from just such a fixed point. But visually equipped animals move all the time. Birds and lizards nod their heads to generate the kinds of change necessary for vision. We dodge around, advancing and retreating, ducking and stretching, to observe and understand. When something demands our attention, we find the optimal distance and angle to view it at by means of a continuous stream of movement, and once done, we move again to whatever next catches our eye. And all the while the eyes are active, saccading, trembling, gliding, and locking onto different landmarks in succession.

Vision and space

*Figure 17: Artwork by Felice Varini. Top: The coherent perceptual form, as captured by a camera. Bottom: Disintegration of the form when viewed from any other position.*

Anamorphs are images that are intelligible only when viewed from a specific point in 3D space. Readers may be familiar with the anamorphic elongated skull in Holbein's famous 16th Century painting of The Ambassadors. Varini reduces the image to almost nothing, in order to emphasize how the visual gestalt transforms with movement, becoming a coherent unity only from a static viewing position. (Image reproduced courtesy of the artist.)

It is not only anesthetized cats and human subjects in visual experiments that are stripped of their active subjecthood, and forced to stop moving. The vast field of neuroimaging has been developed under similar strictures: Subjects in scanners are not only isolated from any meaningful world, they are instructed not to move. Movement in almost any brain imaging paradigm creates poor quality images, and so subjects are told to sit still. Only for the still subject can we put the pieces together that are necessary to construct the Cartesian theatre assumed to mirror, but not be, the world.

This elimination of movement is not innocent. Although frequently motivated by technical constraints (fragile electrodes, imaging requirement, etc.), it is taken for granted that the subject who normally moves can be studied when not moving. This view is at odds with the enactive view of subjecthood we have been bringing to bear on joint speech, in which a subject pole arises through the collective activity of speaking together. The subject of joint speech is not co-extensive with any person; it is necessarily collective. We started by noting the manner in which received approaches that take some essential psychological self as a given have failed to make joint speech a topic of inquiry, and that such approaches must necessarily have difficulty with the very idea of collective subjects. Now we have found a related characteristic of such approaches: They exclude movement. So it seems only appropriate now that we should turn to joint speech considered as a special kind of coordinated movement.

## Joint speech as synchronized movement

The definitions of synchronization and of behavior introduced at the start of this chapter are intended to fit together to allow us to identify purposive activity that is done by multiple people at a time. There are many activities that can be described as synchronized in the manner adopted here, though our pragmatic choice to tie the definition to observed movement will rule out some. Thus, we can have no index of synchronized thinking, unless someone comes up with a way to agree on what an observation of thinking might be. However we still have available to us such activities as marching, dancing (for some kinds of dance styles), rowing, and there is a small suite of sports in which synchronization becomes an end goal in itself: synchronized swimming, trampolining, and diving. Perhaps the reader can add to this list.

When we survey all these activities, it becomes apparent that joint speaking occupies a unique position. For all other synchronized activities, the

synchronization is achieved by relying on one, or both, of the following characteristics: There is a strong beat providing a public timing signal, and/or the activity is greatly constrained by physical properties such as gravity, viscosity, or elasticity. Thus in marching, we have a regular beat, often accentuated by the voice, for example in singing a military cadence. Dancing has an obvious beat that allows anyone to join in, as we noticed before. There is a beat in the music that inevitably accompanies synchronized swimming too. Frequently that music is piped underwater to act as a scaffold for the joint activity. Rowing is regular in this fashion too, but it is also an activity that is strongly shaped by the drag of the oar in the water, and by the mass of the oar itself. In synchronized trampolining and diving, we find gravity and elasticity providing non-negotiable strong limits on what kind of activity is possible, and when. Trampolining adds a quasi-regular beat through the vertical bounce of the trampoliner.

Despite the support provided by a regular public beat and/or strong physical constraints, many of these synchronized activities take a great deal of practice to master. On a recent trip to an army barracks, I had the pleasure of watching a phalanx of relatively new recruits practice their marching around a courtyard. It was difficult. While I watched, they managed to march into a wheelbarrow, and then into the corner of a building for which they were suitably reprimanded by the officer in charge. Discussing what I had seen with the officer later, I asked him what he saw as the purpose of marching drill, which continues to be done by virtually every army, even though nobody has marched into war for over a hundred years. His candid reply was enlightening. “We don’t want them to think for themselves,” he said. “We want them to think as a group.”

Unlike marching, speaking in synchrony with another does not require practice. It seems to come quite easily to most speakers, even in the unnatural setting of the laboratory. Furthermore, although some kinds of joint speaking do make use of a regular beat, of repetition, and of overtly musical elements, these are not required for synchronization. In a typical synchronous speech task, the text is new to the speakers, there is no regular beat in the speech, and there is no repetition. Speech is also distinguished by being a highly coordinated form of activity that is done almost without contact with the physical environment. Of course there must be air, but the business of moving muscles in a precise fashion is all done behind the lips. Unless the speaker is chewing gum or smoking a pipe at the same time, the organs of speech are almost internal, and certainly not constrained in the way in which the body of a trampoliner or a rower is.

And so joint speech immediately poses an interesting problem for students of movement: How is such highly synchronized activity possible without either of the two features that normally provide the scaffolding needed to align movements across individuals? I do not have an answer to offer, but the field is wide open for investigation. However some observations made in the last chapter will probably be important: In order to join in with other speakers, one must of course know, and be willing to speak, the words. This is quite different from the lightweight business of joining a dance or a march. Speakers share more than a beat. They share a willingness to utter together. And, of course, as speakers of a common language, they share a baseline of motor skills appropriate to the task.

Joint speech, studied as synchronous speech in the laboratory, throws up another interesting observation. Under these specific conditions, attempting to remain in synchrony with another speaker, it sometimes happens that both speakers simultaneously and abruptly stop speaking (Cummins, 2014b). Their speech thereafter becomes entirely uncoordinated, and laughter usually happens on both sides. The abrupt cessation can happen in mid-syllable, and would be most peculiar if it happened to a single speaker speaking alone. In that unlikely circumstance, we might perhaps look around to check for the presence of a sniper. But with two yoked speakers, it happens quite frequently (Figure 18).

A speech error unique to synchronous speech

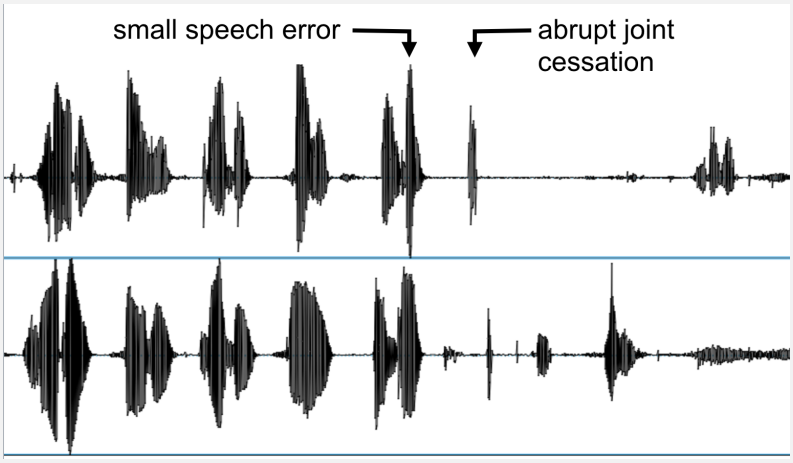


*Figure 18: Abrupt and simultaneous cessation of speech by synchronous speakers.*

Speech and language scientists like errors, and they particularly like errors that might be informative about the underlying processes that give rise to speech. For example, Spoonerisms ("queer old dean" for "dear old queen") have provided endless material for arguing for the existence and sequencing of particular kinds of processes during the planning and execution of speech movements. In the case of joint speech errors, with simultaneous abrupt cessation of speech in two speakers, we have a highly context-specific error that can tell us something about the manner in which two such subjects are linked. How should we understand it?

A useful analogy presents itself in the three-legged race. In such a race, the two runners are, of course, perfectly capable of running on their own. However during the race, their legs are literally bound together, which constrains the running somewhat. Under these conditions, the running pair is in a more precarious position than one of them alone would be. Alone, I can rapidly compensate for an unexpected bump, a slip, or a wobble. Together, if something unexpected happens, it is quite likely that the whole pair will come crashing to the ground. The coupling reduces the ability of each member to respond autonomously, and that makes the resulting pair brittle.

This seems to capture well what is going on when such errors arise. For some reason, for example a small error by one speaker, uncertainty is introduced, and as a result, the dyadic, or paired, system which was previously evident is now nowhere to be seen as both speakers stop simultaneously. Formally, we say that the two component systems–the speakers during the production of unison speech–are coupled, or non-independent while they are synchronized. That which one speaker does has immediate effects on the other, and the two speakers are in *continuous real time reciprocal interaction* with each other. In fact, the coupling among speakers is straightforward evidence for the transient existence of a synergy arising through the coupling among speakers. The synergy exists for exactly as long as the speakers are yoked together, and ceases to exist once they are no longer coupled. Understanding that coupling now becomes an immediate concern. On what basis are these speakers coupled? What is the nature of the feedback that closes the loop between them?

Coupling between two systems, each of which is going about its own business, has been a matter of interest to scientists for centuries, since Christiaan Huygens, the Dutch astronomer and polymath first took to his bed with the flu in the middle of the 17th Century (Bennett et al., 2002).

**Synchronization as coupling**

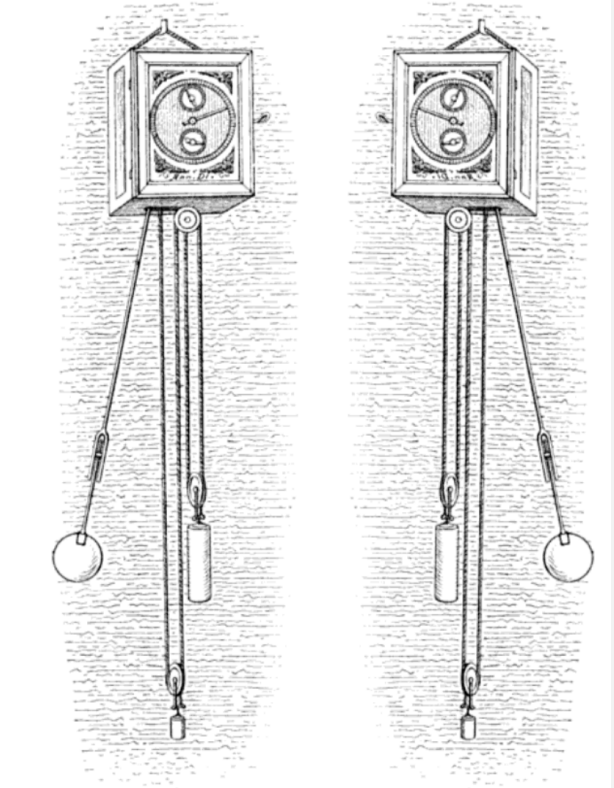

*Figure 19: Two of Christian Huygens's clocks illustrating an anti-phase relationship (or "odd sympathy").*

Coupling, or entrainment, happens whenever two oscillating systems are capable of influencing one another. The influence may be very low-energy, and it will be more effective as the resonant frequencies of the two oscillating systems are more similar. Such coupling is found in both animate and inanimate systems. Dynamic systems theory provides a rich mathematical toolbox for describing such systems.

Huygens, along with his many other accomplishments, had invented the pendulum clock, which represented a considerable advance in timekeeping technology at that time. In bed with the flu, he is reputed to have amused himself with two such clocks mounted on a common housing. The pendula of the two clocks were found to display non-independent motion. Specifically,

they displayed a syncopated relation whereby one bob was halfway through its cycle while the other was just starting. He called this relationship an "odd sympathy." Today we would describe this as an anti-phase relationship (in-phase would describe a related stable relation in which each bob was at an identical point in its cycle as the other at all times).

To amuse himself, Huygens repeatedly interfered with the clocks using a long stick to retard the motion of one or other pendulum somewhat. As he did so, he noticed that the two bobs eventually settled back into the stable anti-phase relationship. This interdependence among the pendula, Huygens concluded, was enabled by the transmission of very small vibrations through the common beam supporting the clocks. If the clocks were hung far apart from each other, synchronization did not occur. Mathematicians have since greatly elaborated upon Huygens's initial account, and the mathematics of coupled oscillating systems has burgeoned into a large field with results applied to many kinds of physical, biological and social phenomena.

If vibration was the basis by which the two clocks became coupled (or *entrained*, if we may bring another term of art to bear), what then is the basis for coupling among speakers during a synchronous speech experiment? It might be that some easily identifiable characteristic of the sounds produced may serve as a link between them. For example, the speech signal is characterized by a relatively slow-moving intensity envelope, which increases and decreases with most syllables at a rate of about 5 times per second. Perhaps this intensity envelope is what links the two systems? Or the speech signal also contains pitch information based on the vibration of the vocal chords. Either of these two signal attributes might here play the role of vibration among linked clocks. This can be tested (Cummins, 2009).

We previously observed that subjects can synchronize with a recording of speech by another person, even if this synchronization is not quite as precise as that obtained among live speakers. This allows us to make a variety of alterations to the recorded speech and examine whether they affect the ability of subjects to speak in time with the resulting altered signal. We tried many manipulations, including replacing the pitch contour with a flat monotone, using just the amplitude envelope, synchronizing with hiss-like sounds, and more. What we found was that the more we degraded the signal, the worse performance was (i.e. the less synchronous the resulting speech was), but performance was not to be explained in terms of any single acoustic parameter. Intelligibility was the principal factor: If the speech was intelligible, subjects could synchronize with it. When we altered the signal so much that it was no longer intelligible, the results were all poor, but there were hints that a combination of an appropriate intensity envelope, along

with an underlying speech-like source, could support a small degree of synchronization. All in all, this set of experiments served to demonstrate that coupling between speakers was not to be attributed to any simple physical link. Rather, synchronization demanded that the signal be interpretable as speech, and preferably intelligible speech.

We need to be aware that the phenomenon we are studying (joint speech) does not translate entirely into a laboratory situation. While synchronous speakers in the lab achieve very tight synchrony and demonstrate strong coupling among speakers, this is not a good characterization of joint speech as found in everyday practices. If we listen in to the crowd reciting the rosary, the voices are not particularly well aligned. It is also emphatically not the case that a mistake by one speaker is in danger of bringing the whole collective into difficulties. Prayer and protest chants, football and school chants all draw on the ability of speakers to speak in unison, but they do so in a manner rather unlike that which we observe in the lab. Another important difference is that joint speech in everyday life is motivated. Speakers care greatly about what they are uttering, and the act of uttering commits them in ways foreign to the artificial bounds of the phonetics laboratory.

So joint speech must be of interest in the study of movement and in the study of behavior. Precise synchronization without a beat or physical constraints serves to make it unique among synchronized behaviors. In the laboratory, two speakers who are precisely coordinated with each other constitute a dyadic synergy whose existence is precarious, and which can fall apart when perturbed, as indexed by a unique kind of speech error. There is a lot of science as normal to be done here, to better understand how speakers can effortlessly couple with each other when speaking in unison. And there is much food for thought as we relate the empirical study of movement to the interpretive study of behavior.

# Chapter 6

# Joint Speech and the Brain

Discussion of brains will be challenging. Brains are complex. From a human perspective, they seem to be the most complex thing in the known universe, but there is no metric that can substantiate that claim, nor is there a way to disentangle it from the very concerned perspective of a human. In critiquing the psychological subject and the manner in which it appears in science, there will be a great deal of difficulty when it comes to brains, because brain science has almost all been done squarely within the tradition that understands people as having discrete minds, as being agents whose agency is independent of context, and as having experiential lives that demonstrate a unity and continuity from birth to grave. Within this tradition, furthermore, subjectivity is understood as arising causally from the activity of the brain. So from a contemporary neuroscientific perspective, there can be no collective subjects, because there are no collective brains. There can be no collective subjects, because subjects are experiencers, and experience is the kind of thing that goes on individually, one experience per person, and each utterly distinct from all others. There can be no collective subjects because subjectivity is almost the same thing as consciousness, and we all seem to know that brains are what make us conscious. No matter how dissatisfied we might be with the self-aggrandizing certainty of the Cogito (Descartes' "I think, therefore I am"), it would appear that (almost) everybody within modern Western society is convinced that they live inside their head, that their thoughts, fears, dreams are all going on within the skull, that their world of experience is discretely different from everything else. Without a doubt, the approach taken in this book so far will encounter resistance from contemporary neuroscience. On the upside, perhaps consideration of joint

speech might help in developing rather more varied and pluralistic accounts of subjects and values. Perhaps neuroscience itself might even contribute!

In this context, we need to consider, and then strongly reject, the pronouncements of another Nobel Prize winner. Francis Crick, together with James Watson, Rosalind Franklin and others had uncovered the double helical structure of DNA, a towering achievement in the biological sciences, and so in 1962, he and Watson, along with Maurice Wilkins, was awarded the Nobel Prize in Physiology or Medicine. Ever the hard-nosed scientist, Crick was particularly critical of Christian beliefs, quipping that "Christianity may be OK between consenting adults in private but should not be taught to young children." Later in life, Crick turned his attention to Cognitive Science, or to Cognitive Psychology, for the position he took makes the two indistinguishable. In his book "The Astonishing Hypothesis," Crick asserted:

> "You," your joys and your sorrows, your memories and your ambitions, your sense of personal identity and free will, are in fact no more than the behavior of a vast assembly of nerve cells and their associated molecules. (Crick, 1995, p. 3)

Even among neuroscientists, Crick's position represents something of a materialist, reductionist extreme, but it clearly articulates one entrenched orthodox position. It also carries authority, not least because of Crick's prior work in biology. Furthermore, reading reviews of the book in which this bald assertion is made, one discerns a lot of sympathy among practicing psychologists for the view espoused. For example, Glyn and Brelstaff asked: "So how astonishing is his hypothesis? ...most readers of Perception [a major psychological journal] will find it [...] to be much the same as the working hypothesis that they employ daily in the laboratory...the Astonishing Hypothesis is so plausible that it should not be called astonishing" (Glyn and Brelstaf, 1994). Rall noted "To many scientists, the astonishing hypothesis (AH) is not astonishing at all; it is precisely what we think, or so we say at first glance" (Rall, 1996). It is interesting that Crick saw his work as a programmatic attempt by science to do away with the frippery of the soul, yet in the identification of the person with the brain, it would appear, to me at least, that he is extending the questionable tradition of placing the soul at the heart of scientific psychology. I doubt he would concur.

Before progressing, it will be useful to distinguish between what I like to call "wet neuroscience" and a rather different enterprise that goes by the name of "cognitive neuroscience." By wet neuroscience, I mean the scientific

field within the broader domain of physiology, that looks at the organ—the brain—in much the same way as one might look at the thyroid, the liver, or, indeed, the heart. It examines its structure, its parts, their connectivity, and their role within an economy of the body. Wet neuroscience is species-neutral as it adopts pretty much the same approach to the brain of the human as it would to the nerve net of a jellyfish or the ganglia of an earthworm. When wet neuroscience speaks of function, it is well aware of its commitments: The brain stem plays a role in keeping respiration going and in stimulating the heart to beat, for example. Functions are referred to the continued integrity of the body, and nothing else.

Cognitive Neuroscience, on the other hand, takes the concerns of cognitive psychology and attempts to cash them out in terms of brain activity. It is here that we find talk of the brain "recognizing," "deciding," "controlling" and the like. This enterprise is absolutely not species neutral. It seeks to understand the human brain using the concepts of psychology. It frequently makes use of a suite of concepts including "beliefs," "intentions," and "desires" to account for the observed behavior of a person. Cognitive neuroscience takes the brain to be the seat of consciousness and with that, it is the place where experience "happens." Wet neuroscience does not speak of experience.

It is probably clear that the author of the present work is not going to be happy with a cognitive neuroscientific account of the person that assumes a singular mind arises from the activity of a single organ, such that the "external" world is only indirectly experienced by the inflow of information through the senses, where it is crafted into an interior model of a transcendent exterior reality. This entire metaphysical picture is a direct descendent of the Cartesian/Kantian approach to mind which was forged in a strongly Christian crucible. It insists on regarding each person as entirely autonomous, and acknowledges no subjectivity or agency other than that of the individual person, considered as removable from any context. We will leave the elaboration of this theme, and the consideration of alternative views, to the final section of the book, as this middle section is intent on keeping a focus on joint speech and the kind of things we can learn from it as we pursue "science as normal."

Given the large gap between the conventional view of the brain and the challenges we meet in considering genuinely plural or collective subjects, a bit of a digression is necessary at this point. It should be possible to make some observations that can tease out the role of the brain in human activity, even perhaps in joint speech, despite our necessary departure from an orthodox interpretation of the brain as the meaty instantiation of mind,

as Crick would have it. I will begin by arguing that experience, consciousness and the like are not condemned to a hidden interiority and are not necessarily to be attributed to brains. This seems like a minimum of work necessary before we tackle joint speech and the brain directly. I will also point out how the manner in which much discussion of brains is conducted is rather obviously little more than lightly retouched soul-talk (Reed, 1997). I will then return to our empirical focus as I describe some recent work in which we find that even conventional contemporary cognitive neuroscience can see some quite remarkable phenomena when people speak in unison. In Part 3, I will flesh out some theoretical arguments that might make an astonishing counter-hypothesis somewhat more plausible, namely that you are most definitely not your brain, or any set of neurons.

## The Brain in the Vat vs. The Brain in the Ass

Philosophical arguments rarely make the mainstream. Descartes "I think, therefore I am" is a rare exception. A much more recent thought experiment, proposed by the American philosopher Hilary Putnam, elaborates on the concern that Descartes had that he might be utterly mistaken about his perception of the world by being misled by his senses. This "Brain in a Vat" argument found a dramatic and effective realization in the first of the three Matrix films. In a dystopian future, humans are being farmed for their "bio-energy" (not a terribly well thought out concept) by machines. To this end, they are "grown" in pods, in which they lie inert and isolated. However they thrive nonetheless, because each individual experiences a simulated reality entirely unlike the physical situation of being a farmed object in a pod. Thanks to an electrical/digital connection to a large computer (the details are necessarily a little sketchy), each person-in-a-pod has the experience of being an autonomous, active person roaming freely in the world of 1999. The world is of course simulated, and this is possible because the connector provides just the right "inputs" in response to "outputs" in order to sustain the illusion. This is all good fun, and allows the film to explore parallel realities in a manner that might (who knows?) have delighted Descartes himself.

But is it plausible as a thought experiment? If we ignore any and all technical details, the scenario provided only makes sense if the experience of an individual person depends upon the inputs to the brain, provided from the outside, where these inputs are dependent, in large part at least, upon outputs from the brain. The brain, on this view, is an input/output system,

and the world provides the inputs, while outputs arise, inter alia, from the volitional action of the person whose phenomenal world is generated by the brain. A recent take-down of the brain-in-a-vat hypothesis goes into rather more detail than we will need here (Thompson and Cosmelli, 2011). In essence, the argument asserts that the split between brain and world that allows talk of inputs and outputs, is a non-starter, because simulating the "inputs" necessary to mimic the flux of energy at the interface between nervous system and world would require simulating the whole body, and then the set of relations the body is in with its surrounds, which pretty much negates the premises on which the thought experiment is based. Separating brain from body and then from world, as an input/output account requires, is not possible in principle, as everything we know about brains and bodies suggests that the lived experience of any creature arises through coupling among systems: including neural, endocrine and immune systems, such that the body is continuously and reciprocally coupled to its environment. Once more, the importance of real time reciprocal interaction becomes insistent as we study the relationship between subject and world. As the authors put it:

> We've given reasons to think that the body and brain are so dynamically entangled in the causation and realization of consciousness as to be explanatorily inseparable. (Thompson and Cosmelli, 2011, p. 28)

The interested reader is pointed to the full paper for the gory details. Here, in place of this venerable discussion, we might put forward an alternative, unpublished, and as yet unfilmed, hypothesis that likewise serves to interrogate our intuitions about the relationship between brains, experience, and worlds. I call it the "Brain-in-the-Ass" hypothesis. It starts with the commonly held conviction that we are "in our heads" somehow, or at least that the subjective experience of an embodied being has a center, or origin, and that that lies in its head, somewhere behind the eyes. This thought experiment requires rather less suspension of disbelief than the brain in a vat one. Instead, we consider a relatively trivial (by comparison) rewiring of the central nervous system in which the brain itself is located outside the skull. To keep it safe without its bony helmet, let's put it in that other well padded sanctuary, the buttocks. On this hypothesis, the optic nerve still runs between the occipital lobe of the brain and the eyeball, it just has a longer path; the auditory nerve still runs from primary auditory cortex to the inner ear, and so on. Under these conditions, I contend there is no reason to suspect that there would be any substantial change to the character

of experience, the nature of consciousness, or the sense of "being inside my own head."

In a brain-in-the-ass body, I would still use my head and neck muscles to orient my head. The complex suite of muscles that cradle and direct the eyeball to allow me to look here and there would still be around the eye socket. In this case, the visual experience of the world would not change at all, as the visual sense of being centered somewhere behind the eyes arises from the orientation of the eyes and head towards the world, and not from any neural activity. (I am here partitioning experience into the visual and the rest, which one might object is a post-theoretical dissection that is unsustainable. But I am not claiming that experience actually decomposes in this manner, just that much of the sense of having an experiential center arises in part from the anatomy and physiology of the eyes and muscles of the head.)

But there are other reasons that we feel our conscious experience lies in the head. Consider those inner voices. Whether it is just you silently enumerating your intended shopping purchases, or inwardly reciting a prayer or mantra, we all "hear" silently, and our best guess for where those voices might be seems to be in the head. Now consider what it is to speak. When you speak, you cause a pattern of vibration in the air that is available to others, but you also cause your skull to vibrate, and you hear your own voice primarily through bone conductance. (This marked difference in the perception of your voice compared with everyone else's voice, is part of the reason why people typically get such an unpleasant shock when they hear their voice from a recording for the first time.) Assuming that an inner voice is very similar in origin to an overt utterance (an assumption we explored in consideration of the voice as it relates to both speech and music), it seems unsurprising that your experience of speaking silently to yourself should bear a strong resemblance to the situation of hearing yourself speak out loud. This relation between hearing and the head, though, is based on the location of the ears, the conductance of vibration through the skull, and the location of the larynx in the neck. It does not depend in any way on the location of neural activity, and so in a brain-in-the-ass body, inner voices will still be in my head. The location of the brain is once more a detail of implementation; it does not determine or even influence the character of the experience.

Young infants first engage with their worlds through the mouth, in suckling, but also in exploring the world through touch and taste. In common with speaking, this may serve to fix the head as the locus from which we engage with the world. Later in life, it is our faces that others orient them-

selves with respect to. But once more, it is the body and its engagement with the world that seems important, not the location of neural firing.

And so the brain-in-the-ass thought experiment might serve some purpose, to help us to distinguish between the veridical interpretation of the head as an important locus in engaging with the world in various ways (hearing, seeing, tasting, suckling, facing others), and the rather mistaken notion that experience might lie in the brain itself. Shift the brain, and nothing happens. Just be careful sitting down.

## The blue brain and the soul

Each of us, every individual, finds themselves at the center of their own world. You look from somewhere, you negotiate your position with respect to your surround, and above all, you feel and care. Things matter as they relate to you. Objects nearby are more salient than those far away. Distant events seem to matter less than those near you. A million people killed in a distant land, or at a distant point in time, can scarcely compete for your attention with your thumb after you hit it with a hammer. This perspectival structure is necessarily true, not only of human beings, but of all autonomous living beings. It describes the geometry of being, but not a separation into an interior and exterior realm, or a division into a subject who stands in opposition to objects. Existing in relation to a world is not a Western or a Christian notion; it is a bare fact of life, or, better, of living.

But with the profound change in sensibility ushered in during the 16th Century, where an older view of a hierarchical cosmos was replaced by what I have called isometric space and time, it became necessary in a European context to provide some account of what this perspectivalism is, and how it relates to the common sense notion that we inhabit a shared world. It is out of this foundation that the contemporary account of experience was built, first by dancing around the soul, then by building scientific accounts that were significantly informed and shaped by Christian theology. Everything we have just discussed is compatible with a strong association between an individual person and a point in isometric space located somewhere behind the eyes. However that point is nearly always moving, unless the psychologist has been at work with anesthetics, head clamps, and fixation crosses. That point is not a cavern, a hidden interiority populated with images and representations. Rather, immanent world and subject arise in a continual dance.

The single point associated with the subject is beautifully described by T. S. Eliot in his 1935 poem "Burnt Norton" as "[T]he still point of the turning world." But pulling back and seeing the context in which this evocative view is made, we find a remarkable view of time located at the razor edge of a single moment, the present:

> At the still point of the turning world. Neither flesh nor fleshless;
> Neither from nor towards; at the still point, there the dance is,
> But neither arrest nor movement.
> And do not call it fixity,
> Where past and future are gathered. Neither movement from
> nor towards,
> Neither ascent nor decline. Except for the point, the still point,
> There would be no dance, and there is only the dance.
> I can only say, there we have been: but I cannot say where
> And I cannot say, how long, for that is to place it in time.

To which we might just add the question raised by W. B. Yeats in "Among School Children": "How can we tell the dancer from the dance?"

Sometimes poems help where our words are stuck. Images can do that too. I would like to return to the contrast between wet neuroscience, concerned with the brain as organ in a body, and cognitive neuroscience, which by and large conceives of the brain as the locus at which experience arises, and for whom the brain is the first port of call in tying the concepts of cognitive psychology (memory, perception, attention, etc.) to the physical world.

If we conduct an image search based on the single term "brain," images similar to the blue brains on the left of Figure 20 abound. On the right hand side, we see a real brain. The contrast is striking. The sanitized, toilet-cleaner-blue brain bears no relation to a body at all. In many such images, lines radiate outwards into the cosmos, an image entirely familiar from sacred art, where it connotes revelation, immediate contact with the transcendent Godhead, or visitation by the Holy Spirit. A large Swiss research project housed in EPFL in Geneva is called The Blue Brain Project. How did this remarkable stance with respect to the fleshy gelatinous organ come about?

**Blue brains**

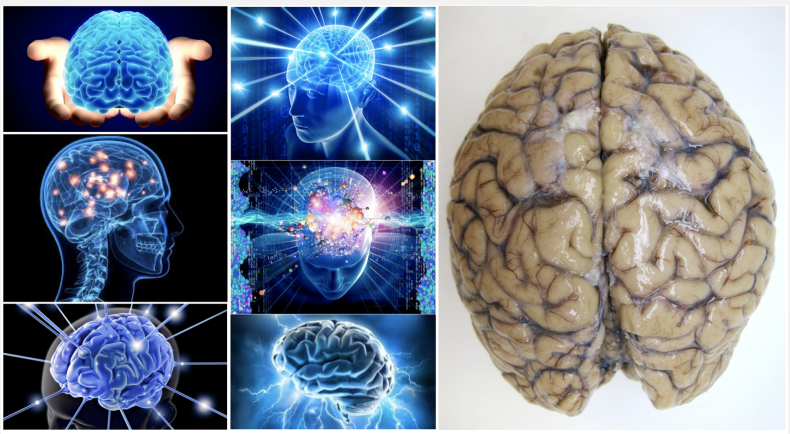

*Figure 20: Left: The brain of cognitive neuroscience. Right: A brain.*

Searching for images of brains returns innumerable pictures of shiny clean blue objects, frequently radiating outwards, or superimposed over pictures from the Hubble space telescope. The lines emanating from the idealized brain would not be out of place as a halo or conduit to the heavens in religious imagery of the Middle Ages. (Image credit, wikicommons user Jensflorian. CCA-SA 4.0 International.)

One way to consider the story is the path taken historically by science. This starts with the attempt to understand Man's (yes, again, Man) role in the universe, from the firm conviction that Man is categorically distinct from all other animals and life forms, endowed with rationality, and possessed of a rich interior life populated with images, plans, perceptions, and memories. Man was fixed within a hierarchy above the animals and below the angels and the divine. As this starting point came into contact with the body, it was clear that the brain is of supreme importance, and so the brain is where these defining features of Man were sought. Over many years, and especially with the development of a theory of evolution, it became harder and harder to resist the obvious commonalities with animals, and so the rarefied intellectual gifts of Man had gradually to share space with brain bits devoted to more mundane activities such as chewing, locomotion, and sex. Happily, the human brain does have a rather salient feature that puts just a little distance between it and the brains of the great apes: The human brain is distinguished primarily by the relatively large size of the pre-frontal cortex,

located over the eyes, and behind the forehead. Contemporary cognitive neuroscience still seeks to locate such capacities as planning, remembering, reasoning, etc. in the brain, and the pre-frontal cortex is large enough to suggest that the human brain might still survive as the meaty implementation of the Cogito, the soul or the mind.

An alternative trajectory is provided by the process of evolution itself, working forward from the earliest proto-brains to the present. This is a story that arrives at homo sapiens, not at Man. Not all animals have nervous systems. Those that have the simplest nervous systems are very unlike us indeed. These animals are jellyfish, sea squirts and sea slugs, worms, and the like. In such animals, the role of the nervous system in the economy of the body, i.e. its physiological function, is fairly clear. It subserves locomotion (Keijzer et al., 2013). That is, it acts as a mediating organ between the patterns of change on sensory receptors, and the pattern of activity in muscular effectors, without any necessary interpretation in terms of input and output. As Keijzer and colleagues put it: "Nervous systems are foremost spatial organizers that turn large multi-cellular animal bodies into dynamic self-moving units." Movement is again key in the reciprocal interaction of an organism (one kind of subject, one kind of self) with its surround. But animals with simple nervous systems cope with simple environments. Animals with more complex nervous systems manage to get by in increasingly more complex environments. As they do so, the density and richness of the interneurons–those that are not directly connected either to receptor surfaces or to muscular effectors–increases dramatically, giving rise to the local clumps that are first ganglia, and later brains. Despite the yawning gulf between the body and world of a sea squirt and that of a human, there is available to us a view of the human central nervous system that is continuous with our understanding of the more simple cases. As Thomas Fuchs puts it: "The brain is not the sole producer of the mind, but rather mediates and regulates the cycles of sensorimotor and social interactions with the environment that underlie our conscious experience."[1] This looks far more like a task suited to the meaty fleshy organ found in skulls than to the cosmic blue shiny things so popular in the world of images.

[1]Quote from a talk given at the "The Future of the Embodied Mind," eSMCs Summer School 2011, San Sebastián, September 2011.

## The neuroscience of joint speaking and the self-other distinction

We seem to have strayed way off topic. But perhaps not so much. The blindness of contemporary science to joint speech has been attributed to its insistence on a single kind of subject, the minded or ensouled subject of psychology. Consideration of the manner in which the person is understood in relation to the brain makes this very clear. But what would a neuroscience of joint speech look like? Could we make that coherent?

There is little prior art, but there is some. A recent neuroimaging study by Kyle Jasmin, Sophie Scott and others, located mainly at University College London, shows how contemporary cognitive neuroscience too can find something of interest when people speak with one voice (Jasmin et al., 2016). In order to tell this story, it will be necessary to describe some basic features of the imaging technique used, which is functional magnetic resonance imaging, or fMRI for short. This is the technology so beloved of the press, that seems to generate images of brains with colored areas "lighting up," which are popularly, if thoroughly inaccurately, reported as evidence that one or other area in the brain is performing one or other function. It is not only the public at large who are routinely misled by these evocative images. Scientists too, even cognitive neuroscientists, are apt to misinterpret, or over-interpret such images. The fact that the technology required to produce such images is enormously complex, with very many stages between the process of image acquisition in the scanner and the resulting picture of a colorful brain, makes it all the more important that we approach such work with caution.

The device employed is an MRI scanner (Figure 21). Many readers will have first hand experience of this contraption, as it is routinely used in many areas of medicine. An MRI scanner can produce high quality detailed three-dimensional images of anything inside its central bore. It resolves soft tissue details, and unlike X-rays, it is non-invasive, with no risk to the person in the scanner (as long as they have no metal implants, piercings, or the like). The prohibition on metal parts arises because the doughnut-shaped ring around the central bore contains a very strong magnet that rotates at high speed. It is strong enough to rip out any metal pieces implanted in the body.

An unlikely place for joint speech

*Figure 21: An MRI scanner. (Image credit: Wikicommons user KasugaHuang, CCA-SA 3.0.)*

The subject lies on the central bier, and is inserted into the bore. As so often in psychological studies, this impedes movement. In fact the subject is not allowed to move, or speak, while scanning is ongoing. The subject is not only tightly enclosed in the tube, she is also completely alone, and the rotation of the magnet makes a frightful racket, even in the best machines. All this sounds like an unpromising starting point for investigating joint speech, but bear with me. As the magnet rotates, the magnetic field interacts with hydrogen molecules to generate a signal from which the detailed image is produced. If that sounds like a summary, hand-waving, description, that is because it is. The physics underlying modern neuroimaging techniques is complex and quite unknown to many of the scientists using such devices.

Functional magnetic resonance imaging, or fMRI, is a means of exploiting a rather different signal generated by bloodflow alone, but captured with the same device. It is a relatively fast acquisition process, with one scan taking between 3 and 5 seconds, whereas the detailed MRI image takes up to 15 minutes to generate. In an fMRI experiment, there are usually at least two experimental conditions. The experiment will begin with the slow

acquisition of an MRI image. This is necessary because brains differ greatly across individuals. Like fingerprints or iris coloration, each brain displays a unique pattern of folds, ridges, and valleys. Thus the anatomy must be captured before we can move on to the physiology of the brain in action.

Then the subject is scanned in a rapid fashion during each of the two or more experimental conditions. The signal acquired is generated by oxygenated blood, and not by the brain itself. In essence, the question asked at any point is whether the flow of highly oxygenated blood differs between the two conditions or two sets of conditions. In order to obtain reliable data, it is quite normal for each experimental condition to be repeated many times over. This may mean looking at a similar picture, or listening to a similar sound, over and over. Obviously such a procedure can capture nothing of the novelty of the first viewing of a picture or the first hearing of a sound. It can capture only differences in blood flow that are reliably found on each presentation of a "stimulus." It is worthwhile to consider what a crude instrument this is. Unlike the popular reports, the fMRI image does not capture brain activity, i.e. the firing of nerve cells in the brain. Instead it captures an increase or decrease in bloodflow in one condition compared to another. The relation between bloodflow and brain activity is poorly understood. There is a sort of unspoken working hypothesis that greater neural activity will result in greater bloodflow, albeit with a significant lag in time between the neural goings-on and the measured bloodflow. Bloodflow is recorded approximately 5 seconds after the experimental manipulation, such as presenting a picture to look at. This means that the temporal resolution of the method is very poor indeed. It can make only the crudest of distinctions in time.

With all these caveats, it may seem that there is no way to employ this technology to examine joint speech. In part that is true. The urgency, the participatory urge, that accompanies prayer, protest, and sports chanting cannot be transported to the inner tube of the scanner. But in studying synchronous speech, we knowingly forfeited these essential aspects too to examine something of the mechanics, as it were, of speaking in unison. Sophie Scott's expertise in speech research becomes relevant here in the design of an experiment that manages, despite all these limitations, to demonstrate that speaking in unison with a real person with whom one is in real time reciprocal interaction (again!) has demonstrable and somewhat surprising effects within the brain. This work was carried out as a major portion of Kyle Jasmin's Ph.D. thesis under Sophie Scott's supervision.

In the experiment to be reported, there were 6 different conditions, allowing for multiple comparisons of regional oxygenated bloodflow across pairs of conditions and pairs of conjunctions of conditions. The subject in the

scanner was equipped with metal-free headphones and a microphone so that they were in audio contact with the experimenter. In each condition (except the silent one), subjects spoke (or heard) a sentence, which takes about 3 seconds, and they then remained still while a quick scan of bloodflow was done. The sentences were short coherent but unmotivated texts presented on screen. The lag between event and subsequent changes to bloodflow is here exploited to allow us to look at the effect of mechanically speaking, despite the prohibition on movement during the scanning process itself. The six conditions were:

- REST. The subject does nothing but lie still.
- SPEAK ALONE. The subject spoke a given sentence out loud.
- LISTEN ALONE. The subject heard a recording of the experimenter speaking a sentence.
- DIFF LIVE. The subject spoke a given sentence, while they heard the experimenter speak a different sentence.
- SYNCH REC. The subject spoke a given sentence in unison with a recording of the experimenter speaking the same sentence.
- SYNCH LIVE. The subject spoke a given sentence in unison with the experimenter who spoke the sentence at the same time.

Crucially, subjects were not told about the difference between the last two conditions; that is, they did not know that recordings were involved. Each subject was debriefed after the experimental session and none of them were aware of the difference between speaking with a recording and speaking with a live person. They all were under the impression that they were speaking with a live person only, never with a recording. This distinction is very important in interpreting the results, so we also ran a behavioral experiment in more optimal conditions in which a subject (not in a scanner) spoke in unison with either a recording of an experimenter or with a live experimenter, and we verified that if this is done carefully, subjects cannot tell the difference between the two.

Speaking in unison is not just speaking + listening

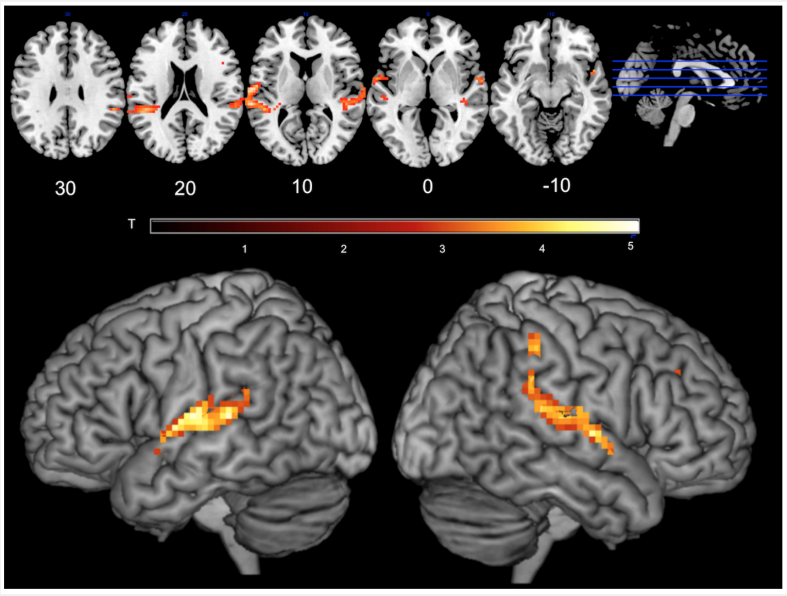


*Figure 22: Comparison of speaking in unison, on the one hand, and the combination of speaking alone and listening alone, on the other. Reproduced from Figure 1 of Jasmin et al. (2016). The colored areas are those which have a greater amount of highly oxygenated blood in the first condition (joint speech) compared to the second (speaking+listening).*

One way of looking at joint speech is to regard it as nothing special, and one way in which one might adopt this skeptical stance is to note that, yes, in speaking in unison one is both speaking and listening simultaneously, but that description exhausts what one might say about joint speaking. On this hypothesis, there is nothing more to say about what is going on in joint speech. Given the above experimental design, we can test this hypothesis directly by combining observations from the SYNCH REC and SYNCH LIVE conditions together, and comparing them with observations from the SPEAK ALONE and LISTEN ALONE conditions. The skeptical hypothesis here suggests that there will be no important difference in comparing these two groups of conditions. The results showed substantial differences however. Figure 22 shows those areas that jump out in this comparison. Given the thrust of the present argument, I will be extremely cautious in interpreting such fMRI data, but to a cognitive neuroscientist, the regions that stand out here are quite coherent: They are considered to be involved in

the perception of sounds generally, and interestingly, they show up on both sides of the brain, although a classical view of brains and speech suggests that in listening or producing speech, most of the action is to be found on the left side of the brain. Without over-interpreting the data then, there is clearly something of interest here for cognitive neuroscientists.

But a rather more interesting comparison calls for our attention. When we contrast the final two conditions, in which subjects spoke in unison with either a live voice or a recorded voice, we find a marked difference, as shown in Figure 23. Recall that subjects themselves were entirely unaware of any difference here, and did not know that recorded voices were used. The difference we see thus has to do with the distinction between synchronizing with an inflexible model (the recording) or with an accommodating partner (the experimenter) who can sensitively respond in real time. The latter condition thus features that hallmark of collective action, real time reciprocal interaction, and this is apparent in the regional bloodflow observed in the brain.

Liveness matters

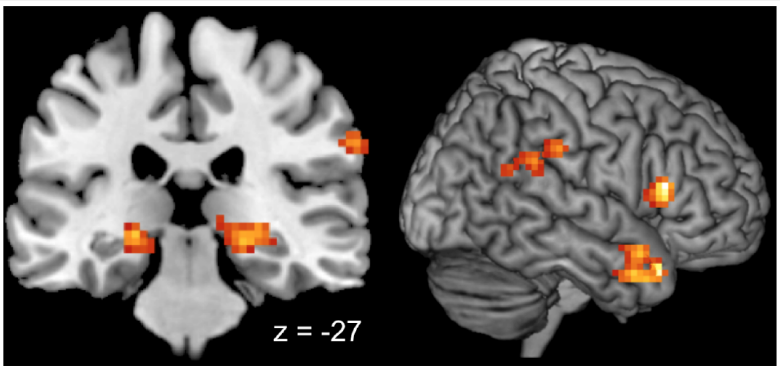


*Figure 23: Increased regional blood flow when speaking in unison with a live voice (SYNCH LIVE) compared with speaking in unison with a recording (SYNCH REC). Reproduced from Figure 4 of Jasmin et al. (2016). The left panel shows a vertical slice through the middle of the brain parallel to the plane of the shoulders. The right image shows the exterior of the right side of the brain. Colored areas have increased oxygenated blood in the first condition compared with the second.*

What should we make of this? Obviously the thrust of the argumentation throughout this book would caution against over-interpretation of anything

in the brain, for we have little idea of how to approach such data without projecting our pre-theoretical assumptions about minds and agents onto an innocent lump of meat. But that does not mean that we can say nothing. For one thing, it is noteworthy that the observed effects lie primarily in the right hemisphere. Aficionados may already be wondering about the little area towards the front of the right hand side of the right hand panel in Figure 23. If the corresponding area on the left hand side of the brain were to feature in any experiment that involved speech production, we should be unsurprised. On the left hand side, this region is known as Broca's area, and its critical involvement in processes of speech production is one of the oldest findings in the field. To see the corresponding area on the right hand side become distinguishable during the real time interaction of joint speech with a live person suggests that there might be an overall alteration to the very well known asymmetrical activity of the brain which we associate with speaking "normally." In joint speech, that is, when we speak, not as a speaker speaking to a listener, but as participant in the collective activity, the asymmetrical distribution of activity in the brain looks rather different. This is indeed worth following up.

In the analysis done by Jasmin et al., a further finding arises that is also thought provoking. A comparison was made that pitted the live synchronization condition against both speaking along with a recording (SYNCH REC), and speaking one sentence while the experimenter spoke another (DIFF LIVE). These are the three conditions in which there are two voices in play simultaneously. This served to identify an area in the right temporal cortex, which appears as the lowest of the colored regions in the right panel of Figure 23. Focusing just on this area, activity there was examined also in the condition where the subject spoke on their own (SPEAK ALONE) and when they listened to the experimenter (LISTEN ALONE). Previous work had established that this particular area reliably displays different kinds of activity depending on whether one is speaking or one is listening to someone else speaking. The conventional interpretation is that one's own voice is, sensibly, treated as different from everybody else's voice, and that this distinction shows up in this region. It makes sense for the brain activity associated with hearing one's own voice to be different from that associated with hearing anybody else's voice. For one thing, as we have noted, you hear your own voice in a different manner, largely through bone conductance. Then there is the rather obvious fact that you hopefully know what you are saying in a manner unlike how you listen to someone else. What the examination of this region demonstrated was that when speaking alone, when speaking while the experimenter speaks a different sentence,

and when speaking with a recording of the experimenter, this region behaves in one and the same way. This is the activity we associate with speaking normally. It behaves very differently when listening to the experimenter speaking (LISTEN ALONE). But the punch line is that when speaking in real-time reciprocal interaction with a live speaker, this region displays activity that resembles listening, and not speaking. To put it in rather more dramatic terms, a known signature of a distinction between self and other, based on the voice, is suspended when speaking in unison, such that one's own voice is no longer clearly distinguished from the voices of those around one.

Despite the enormous limitations of fMRI experimental methodology, despite the crudeness of the signal recorded, and despite the constraints of the scanner, this experiment demonstrated that unison speaking with a live partner is interestingly different from speaking alone, from speaking with a recording, or from the conjunction of speaking and listening. The observed characteristics that single out the live joint speaking condition point to an alteration in the signature of brain activity that normally serves to differentiate self from other. And all this goes on without any awareness of the experimental subject themselves.

Reading brains is undoubtedly hard, and I have refrained from any claim that any single region of the brain bears responsibility for any specific behavior or function. However, as long as we take care not to project onto the brain, it seems the brain may speak back to us, by drawing our attention to qualitative differences that go along with being together with others in real time. I cannot but see this as exciting work, despite the challenges, and it suggests many kinds of experiment one might do to tease these issues apart. For now though, we can note that joint speech is once again amenable to empirical study using the scientific methods and paradigms of the day, and when so studied, it has fascinating characteristics that demand further investigation.

# Chapter 7

# Language and the Voice

## Example 8: Commemorative silence during a talk

A MOMENT'S SILENCE, PLEASE. I am giving a talk to an audience in a University in the South of France. During talks on the topic of joint speech, I frequently find myself wishing I could get the academic audience to engage in some chanting or collective recitation, but I am acutely aware of how uncomfortable that would make people feel. Academia is a precious place in which individuals are free to entertain the oddest of ideas as long as they play by the rules. But chanting seems to require submission to some temporary ad hoc ideology, which would violate the local spirit. I know my fellow academics are not going to join in any meaningful, and hence potentially controversial, chant. And to chant a facile phrase together ("John kicked the ball," perhaps?) seems to miss the point. Playing video clips that record chanting in remote locations is instructive, but a recording is precisely that: a still, unresponsive trace after the event. Chanting seems to demand liveness and reciprocity.

It is April 2016, and France is still suffering after two atrocities, one in January 2015, when gun-wielding fanatics slaughtered the cartoonists of Charlie Hebdo magazine, and a second in November, when similarly aligned deranged individuals shot many dozens of people at a rock concert. In order to bring home the point that joint speech must be understood on its own terms, and not merely as a collection of individual voices, I am going to use an example that works when performed collectively but not at all if done individually. I will argue that the commemorative silence held after such awful events is itself a form of joint speech. Being silent on one's own is nothing. Being silent with common purpose together is something,

something important, participatory, and collective. It is a limiting case, in which joint participation is everything, and the words are reduced to a bare minimum, to nothing. To give the words content might induce dispute, disagreement; it might fail. Silence alone cannot be disagreed with.

Something happens as I play a video clip of a moment's silence held before the start of a football match shortly after the November attacks. I only intend to let it run for maybe 10 seconds to make my point. But the point makes itself, and with force. I find myself unable to interrupt the video, and the audience displays no sense of impatience. The point is made quickly, but the commemorative silence itself turns out to be the one kind of joint speech that can be transmitted, more or less faithfully, over a network. There is no sound, so there is no signal to be delayed. There is no voice spoken, and thus nothing to respond to, to align with, or to resist. The silence captures us all.

## What is language?

The terms "speech" and "language" get bandied around a lot. In an everyday context, they are sometimes interchangeable, and rarely problematic. You speak to me "in" language. Or perhaps you "language" me in speech? Let's not try to reorganize the English language more than we must! It has become common in some rather refined theoretical circles to speak of "languaging" rather than language (Maturana and Varela, 1987), thereby emphasizing that it is an ongoing (and reciprocal) form of activity that is under discussion, rather than an independently existing thing or system. As sympathetic as I am to this reconsideration of how we approach the topic of language, I will refrain (for now!) from insisting on that particular novel term. However in common with those who use it, I think it important that we interrogate the word "language" critically to see how it is used, what it picks out, and what it misses.

Let us discuss both everyday usage, and the manner in which the everyday term finds application in science. In informal use of the word language, we speak usually about an activity, a form of communication, in which two parties are distinguished, as speaker (or writer, sender) and a listener (or reader, receiver). That list of terms already points to a source of possible confusion. The word seems to indifferently describe a process of message exchange using writing, and a process of message exchange using the voice, or hands, or flags. There appears to be a complete separation of message from medium. In either case, the sender has some message or content that she

wants to transmit. To do so, the message must be expressed (or encoded) as a sequence of marks on paper, of sounds generated by the mouth, or of signs traced in the air by the hands. It is not necessary that the sender and the receiver be in the same place at the same time. The message, if faithfully encoded, will survive transmission, as the code captures the necessary information or distinctions that allow the receiver to recreate the underlying content. We do so much of this kind of communication that we need a single term to describe it, even if the means of encoding, transmission, and decoding are themselves always changing. Recent internet-enabled practices of communication illustrate the plasticity and responsiveness of such systems, as emojis, acronyms, and even purely image-based elements become swept up in the ongoing and intensifying business of message exchange. This, I hope, corresponds roughly to what the reader also thinks of as language, more or less, and with many possible qualifications.

Without wanting to change or modify everyday usage, we might note in passing how much this commonsense view of language is shored up by a background in which minds are discrete, thoughts are private, and experience is solipsistic. The activity of one actor (speaker, writer) is distinct from and separable in space and time from the activity of the other (listener, reader). The Cartesian cogito has become the wallpaper before which our understanding of ourselves is situated–not seen, but providing texture to all our statements. Everyday usage will not be overturned easily, nor should it be. But it would be useful to have a larger set of concepts available to us to discuss our interactions. The shortcomings of the commonsense view become more apparent when we consider how they have contributed to and in turn drawn from the more formal study of language within science.

Our brief encounter with the academic field of linguistics introduced some of the concerns of scientists in this area. To most linguists (i.e. scientists who study language, not people who know many languages), language is a kind of system that needs to be studied on its own terms. It can be demonstrated to display a great deal of regularity, in the manner in which sounds or letters or signs are combined and sequenced, in the way larger elements such as words or phrases are themselves combined into ever larger units such as conversations, novels, and speeches. Exploring, documenting and understanding this system is much of what linguists do. In many respects, they are concerned with the business of encoding, and with the particular nature of the code. Much of their primary data comes from the intuitions of native speakers who pronounce judgment on questionable sequences such as “John kicked the ball” (grammatical, unsurprising), “The ball kicked John” (grammatical, surprising though), and “kicked ball John

the" (ungrammatical). The term "grammatical" is most at home in the domain of syntax, which is concerned with sequences of words, as above. But a similar exercise takes place in the domain of sounds, to allow "pipe" and disallow "pfiff" within the language of English, for example. The existence of regularity is attested by the consistent intuitions of native speakers. Where regularities obtain, scientists are apt to try to uncover "natural" laws to explain them.

There is nothing surprising here, except perhaps, for what is left out. And we can see most clearly what is left out if we remember why we might be interested in studying language in the first place. Something happened to our species as it evolved. A mere 5 or 6 million years suffice to take us back to the last common ancestor of humans, on one side, and our chimpanzee and bonobo cousins, on the other. That is very little time. If we examine changes in the body, there are some that are rather obvious: We became almost hairless, and we adapted to walking exclusively on two legs. If we compare brains, we find remarkably little difference. Human brains are bigger relative to body size–considerably bigger, and slightly differently shaped, as humans have particularly enlarged frontal lobes (the bit of the brain situated above the eye sockets). But there are no cell types found in the human brain that are not already found in the brains of the great apes (which include gorillas and orangutans as well). The basic structures are all the same, and we know of no major differences in connectivity either. The rapid increase in brain size is quite recent indeed, having happened in the last two million years. As we study something we call language, it is worth recalling that we are trying to understand an innovation that transformed us and our worlds. We are not trying to understand a change in brains.

It is hard to look back at the history of language, its evolution or development or whatever it was that happened. Language use itself leaves no physical traces. We do not know what kind of language Neanderthals might have had. We do not have a date for the origin of language. Something happened within the last million years or so, after our big brain expansion and it changed us. But as we try to peer back, we must remember that many things that have happened in the meantime have irrevocably altered our manner of thinking about language. It should be clear that emojis (the little graphical symbols so beloved of smart phone users) are probably not going to be relevant to studying the history of language. They may change things in the future, but we should probably ignore them as we try to figure out what happened over the last million years or so. But emojis are by no means the only innovation that we need to undo, as it were, to peer backwards. Writing is the big development that we need to address. Writing is

not very old itself. The first forms of writing were used for keeping lists, and they are about 6,000 years old. Writing spread slowly; it was confined to an elite educated few until very recently, and it served only a very limited set of purposes. It is only really since the development of the printing press and movable type that widespread literacy could make its mark, and that is a mere 500 years old–no time at all, when we are trying to look back a million years; and certainly not something that defines the species.

And what a 500 years it has been, for some of us at least. It is within that time frame, and within the somewhat restricted social world of the literate, that the whole of modern science emerged and that the nation state with its various kinds of authority and law were founded. These developments have not been shared by the whole of humanity, but by a geographically and economically restricted minority. They have since then been filling libraries with documents of all kinds. And these texts have changed us. They have changed how we think about our communication, and about the world. They have made possible the notion that there are facts, or propositions, that can stand freely, unvouched for. "The Earth revolves around the Sun," for example, is the kind of thing that can be written down, and can thereafter exert an influence on readers entirely divorced from the conditions in which it is produced. The independence of the proposition from the conditions in which it is uttered is an innovation made possible by writing.

Because we are so used to writing, it is something of a challenge to see what writing does not capture. For what it does not capture has nevertheless been there all along. Not only in the last 500 years, but in the last million. In almost all of that time, a sentence, or utterance, was spoken by someone in a specific context, and it was gone as soon as it was uttered, unless it was taken up by another and uttered anew. It was uttered in the presence of others to whom the speaker stood in a meaningful, ethical relationship. As it was uttered, the speaker faced, or did not face the listener, and that was meaningful. The eyes did or did not meet. Gestures were or were not present. The voice had a tone. The listener grunted or nodded in encouragement. When writing appeared on the scene, some aspects of this kind of face-to-face interaction were peeled off; specifically, those aspects that could be described within a coherent system, and captured by a small set of categorical oppositions. The structuralist approach to language that arose in the first half of the 20th Century forced a split between the linguistic and non-linguistic; the linguistic was those aspects of communication that described systematic categorical oppositions, distinguishing one word from another. But those elements that admitted of systematization were no more than a small subset of the features of face-to-face spoken and embodied

communication, and they were precisely those contrasts captured by writing. Much of what we try to capture by the term "language" is thus massively determined by the relatively recent practices of the literate.

Joint speech does not belong on the page. It belongs to the vastly greater stretch of human history and development in which communication was primarily oral (Ong, 1977). Within cultures that are still predominantly oral, repetition is not considered tedious; it is a means by which memory is forged, and by which the effect of an utterance can persist beyond the single event of its uttering. In oral cultures (to use Ong's phrase), formulas allow for specific kinds of utterance to persist, to be re-used. These are the pieces from which liturgies and rituals are built. Oral communication is participatory and necessarily involves both speaker and listener who are present to each other in real time. The dialogical preaching of Dr. Cosby (Example 3) is obviously best seen as embedded within an oral tradition. It cannot be properly captured in print or script.

The thing that linguists study when they study "language" has been restricted for the most part to those aspects of human communication that survive the transfer to the written page. So in comparing "John kicked the ball" with "ball the John kicked," the sequence of words is successfully captured. But John is missing, as is his ball, as is any context in which such a trivial and useless sentence might be uttered. "Language" has come to be seen as a system divorced from life, disembodied, impersonal. But this is not any kind of account of the processes that happened to our species since the last common ancestor of humans and chimpanzees. To understand that, we need to peer back to a time before emojis, but also to a time before alphabets, before handwriting, indeed before it was possible to make an assertion that could stand on its own, unsupported by its utterer.

## Monological vs. dialogical sense-making

Not all approaches to language have fallen in meekly behind Saussure, and later Chomsky, in treating of an abstract system, in which specific "sentences" are studied independently of their authorship and context of generation. If the reader is versed in contemporary linguistics, many alternative stances, and with them alternative research agendas will come to mind. But we cannot attempt to be comprehensive when our territory is so large, and we must content ourselves with some fundamental distinctions. One such distinction comes from the school of Mikhail Bakhtin (1895-1975) who worked in Soviet Russia. In keeping with most work in language studies, he

studied texts, but he did so with a fine ear to the voice, and to the importance of both the act of writing or speaking, on the one hand, and making sense, on the other, such that speaker and listener, or author and reader, become entangled with each other in a kind of dance. Bakhtin's work also illustrates the necessity of casting our net wide as we critically consider the strengths and limitations of a science that acknowledges only one kind of subject, the psychological subject. Although his published work is ostensibly in the field of literary criticism (his main works discuss the writings of Dostoyevsky and Rabelais), his stance and that of the school that emerged around him, is of relevance to the most basic discussions of voices, subjects, objectivity and truth.

An untethered text

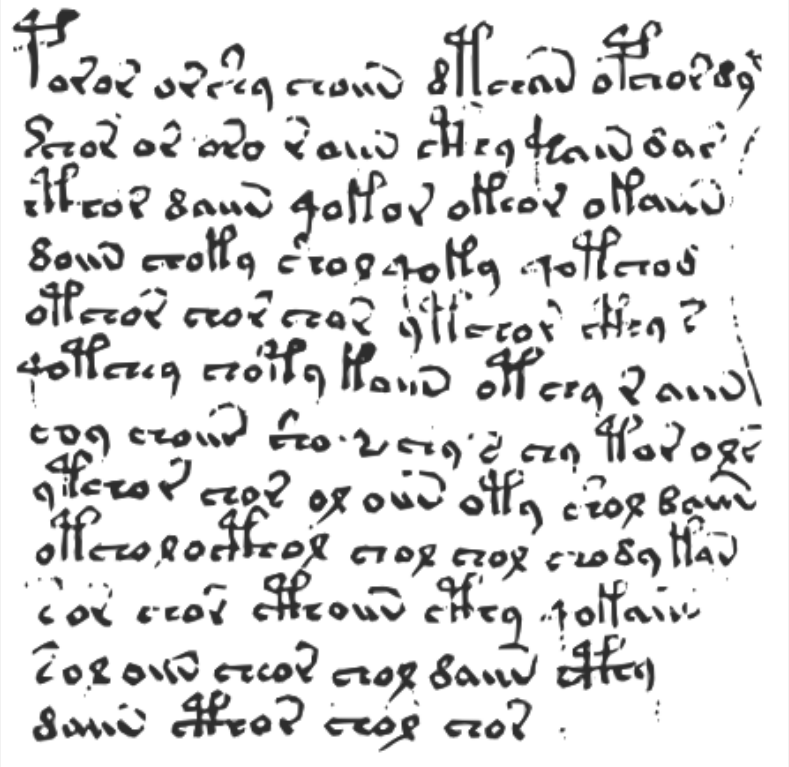

*Figure 24: A page from the Voynich manuscript, which remains today stubbornly uninterpretable.*

We tend to think of words as having meaning. When confronted with a text such as the Voynich manuscript, however, we are forced to recognize that meaning inheres in the interpretive activity, or sense-making, done by a reader. Mere marks on a page mean nothing on their own. They form a bridge, linking the sense-making of the author with that of the reader.

Consider if you will the Voynich manuscript (Figure 24). This is a text, hand written, probably from the early 15th Century, written in a script that has never been deciphered. As far as we can tell, this is not a hoax, a forgery, or an exercise in futility, though it is hard to be sure. It seems that the author had some intention when writing the manuscript, of which about 240 pages remain. What remains is literally meaningless, not because of any flaw in the writing, grammar or style, but because there do not exist readers who can bring the meaning to life by reading it. From the illustrations, we can reasonably guess that its topic lay in the world of medicine and plants, but the text itself tells us nothing. Readers make meaning from texts, and to suggest that meaning resides in the texts alone seems a bit odd when we consider a residue such as the Voynich text. The voice of the author has fallen silent, though the text survives.

Bakhtin introduces the notion of dialogism, which brings to the fore the manner in which meaning is not a fixed quantity inhering in a text, but is something that is constantly made, by listeners and readers, in interaction with speakers and writers. It is a back and forth, a process, in which each utterance or sentence builds on what has just gone and influences what is just to come. We might stick our necks out and say that each utterance is associated with a different "subject," for with each, the common ground shifts, the unspoken assumptions move, and the space of possible future interactions changes. Thus, in "The Problem of Speech Genres" (Bakhtin, 1987) he suggests that the basic unit of speech communication is not the sentence, but the utterance, whose boundaries are marked by a "change of speaking subjects" (ibid, p. 81). On this view, the subject belongs to the utterance. It is not a monolithic entity existing before during and after the act of uttering. It is certainly not a mind or any kind of psychological subject. Subjects are less than fixed. Within an individual, even, we might identify multiple voices, not all in agreement with each other. One does not have to suffer from dissociative identity disorder (formerly multiple personality disorder) to acknowledge that each of us conducts dialogues within

ourselves. We are sometimes surprised by the words of our thoughts, as if they were spoken by an unfamiliar voice.

Bakhtin's notion of the dialogical nature of language, while applied to the novel, the play, and the poem, is most at home in conversation, where the obligatory grounding of one utterance in the context-bound flow, in the immediately previous utterances, and in the shifting common ground of the participants is most obvious. It stands in opposition to the idea of meaning as fixed, as captured by text, and as generated by a monological, Cartesian, subject. Such a radical view is inexpressible within the received vocabulary of the cognitive psychologist, for whom there is only a single subject, coextensive with a discrete mind. And it is not available within a classical linguistic analysis that examines sequences of words torn from the context in which they are spoken.

The shifting subject behind the voice is thematised too by Stephen Connor in his account of the history of voices coming from unexpected places (Connor, 2000). Although we might think of ventriloquism as a form of children's entertainment today, its history suggests something much more important, and at times unsettling. Tracing a rich history from the Delphic oracle, through demonic possession in the Middle Ages, and the origins of ventriloquism as a performance after the Enlightenment, Connor never loses touch of one basic characteristic of voice: A voice belongs to a subject. The original insight of Aristotle in De Anima (Book 2, Part 8) that "Nothing that is without soul utters voice" seems to stand behind the obligatory attribution of subjects behind voices that emanate from places voices should not be. When a voice speaks from a crack in the rocks, from the belly or the genitals, our immediate concern is not with the odd manner of sound production, but with the identity of the assumed subject who speaks. We immediately ask "who?" not "how?"

As we try to envisage language as it existed before the advent of writing, we encounter other obstacles. The disembodied voices of the Oracles, the demons, and the ventriloquist dummies have lost something of their power as more recent technologies have provided so many ways in which voices can be separated from speakers. The near simultaneous invention of the telephone and the phonograph (1876 and 1877, respectively) gave us voices without speakers, and speakers without co-present listeners. Radio, television, and lately, speech-enabled devices telling us "you have mail" or "mind the gap" have further desensitized us to the power of the voice to invoke an obligatory subject. If Siri has a soul, it is a thin and watery one indeed.

If voices conjure up subjects, then the shifting voices in a dialogue generate shifting subjects. Each utterance serves to briefly enact a perspective on

the world, and those perspectives align more or less in conversation, align more fully in liturgy, and align completely in the chorusing of the Credo (Figure 4, p. 36). These subjects are not fixed entities in opposition to fixed objects. They arise and die in the act, in the performance. They constitute ephemeral perspectives, sustained by the act of collective performance, and resonating in the lives of those who took part. If we are to understand how language arose, what it was that happened between people, that so profoundly altered the trajectory of our species, we need some means to address this head on. Speaking as one is efficacious in helping to bring into being a collective perspective, and the essence of any coherent social organization or ideology is the capacity to adopt and rely on such a shared perspective. The collective subject of joint speaking can draw our attention to the importance and familiarity of this collective side to our being, urging us to attend to the rituals, the group activities both rehearsed and spontaneous, that underlie all human being.

## Joint speech and speech acts

What can we do with joint speech? If joint speech is not in the business of message passing, what does it actually effect? It certainly functions differently from our conversational talk, our lectures, and our silent inner thoughts. Inevitably, the words we speak in unison are not crafted by us. Indeed, Rappaport (1999) has taken the non-novelty of the words and acts in ritual to be an important characteristic when he defines "ritual" as "the performance of more or less invariant sequences of formal acts and utterances *not entirely encoded by the performers*" (Rappaport, 1999, p. 24, emphasis mine). The persistence of rituals through time demands repetition, not authorship. Likewise, in the more volatile worlds of protest or the football terraces, the chants are repeated, and the speakers do not create the words they speak. Even on those singular occasions such as the swearing of an oath where joint speech is used purely instrumentally, the words spoken are authored elsewhere, designed to suit institutional purposes, not to express the opinions or emotions of the speaker. This necessarily means that the ends to which joint speech can be turned are different from those of dialogue, monologue, or thought. If language considered from a conventional perspective is about the passing of encoded messages from one Cartesian mind to another, what is joint speech about?

Joint speech, I have argued, gives rise to, or enacts, a kind of collective subject, but this subject is not to be confused with the subject studied within

scientific psychology or with the construct of the individual mind. The words may be fashioned by persons unknown, or at times long in the past, but in their uttering, these words can still have demonstrative effects, that bring about something in the world, beyond simply the transient enactment of a collective subjectivity or a shared perspective on the world. Collective intentions and desires can certainly be expressed, as when we together appeal for mercy or the fall of the regime. Certain kinds of performative acts can also clearly be accomplished, such as when we express generalized assent. For example, in many parliament situations, some votes are taken by collective voicing of "aye" or "nay," and on the basis of the perception by an official that one is louder than the other, a motion may be passed or struck out. But what about some of the other things we uncontroversially achieve with our voices in everyday settings?

Chanting can actually be used to address another. Because the originating subject is collective, the other is usually also collective. In so doing, it is not unusual for chant to cause offense, despite the fact that the words are not crafted spontaneously. Football chants are frequently used to express an identity explicitly. That is often most readily achieved by identifying an "other" to whom the collective can stand in opposition, and the words chosen, in keeping with football's ethos, are antagonistic, provocative and not infrequently humorous. Offensive name-calling is thus clearly possible. Large political gatherings, such as conventions, may likewise give rise to partisan and contentious chanting serving to make manifest different subgroups. Where multiple groups are present, groups may even dialogue with each other, creating a collective conversation. A recent atheists' convention in Australia, for example, drew a crowd of Wahhabi (fundamentalist Islamic) protesters, who launched into a call and response chant listing the names of prominent "new atheists" thus: Call: "Christopher Hitchens," Response: "Burn in Hell," Call: "Richard Dawkins," Response: "Burn in Hell." The chant seemed to amuse the atheists who gathered to watch, but they soon organized themselves into a counter-chant of "Where are the women?" creating a back and forth that entertained both participants and passers by.

But if causing offense is something that works in both joint speech and in conversation, lying is an act that seems to be improbable, or perhaps impossible, when speaking collectively. I have in mind here the common-or-garden lie, in which one person deliberately conveys erroneous or misleading information to another. To define a lie, I have to make use of the message-passing metaphor, and we have observed that this is not a helpful metaphor in understanding joint speech. If no messages are passed, perhaps then it is rather obvious that no lies can be told. But I think the apparent

absence of lying as a function of joint speech deserves a little more unpacking, for it must, of course, touch upon the notion of truth, and the notion of truth must be addressed as we consider moving from a strict subject/object divide to a different form of negotiated consensus. As Rappaport (1999, p. 133) observes, statements and descriptions report existing states of affairs, while performative acts realize states of affairs. Joint speech is intrinsically performative. It does not produce statements or descriptions. It enacts, or brings into being, through doing. This seems to remove it from the symbolic order, or the systematic world of language classically considered. Truth is typically assessed by comparing statements with some independent state of affairs. "The rabbit is in the hat" will be judged true or false depending on whether or not the animal in question is located in said hat. But "The people demand the fall of the regime" does not admit of this separation of word and world. The act of uttering makes the content necessarily true. The correspondence theory of truth, then, does not seem appropriate to the collective utterances of joint speech.

Rappaport (1999) provides a much fuller consideration of the relation between rituals and truth. He notes that sacred truths, which may be enshrined in scriptural authority or affirmed in ritual, invert the normal relationship between statements, the world, and the truth. Whereas statements about the world are normally judged to be true or false with respect to the state of the world, sacred truths, by being uttered in liturgy and ritual, are frequently unquestionable, and so states of the world are judged to be "true" or "false" depending on whether they correspond to the utterances. This represents a complete inversion of the more familiar relationship between the world and statements. He notes that states of affairs in the world that depart from the liturgically established order are called by words indicating a lie in both Zoroastrian (druj, Persian) and Vedic (anrta, Sanskrit) traditions (Rappaport, 1999, p. 133). Elsewhere (p. 359 ff.) he argues that the broad notion of an underlying order governing the affairs of the world and its inhabitants is common to many traditions. In Greece it is the term Logos, which we are making free use of in this work. But the ancient Egyptians had a similar notion of Ma'at, pre-Islamic Persia had Asha, Vedic India had Rta, Mexico (the Nahuatl culture) had Nelli. Order and the notion of truth are linked in every case. Consideration of just how joint speech, and its associated rituals, liturgies and practices ground the unquestionable, the necessary and the inevitable may require us to tolerate a considerable shift in our vocabulary for speaking of truth and necessity.

## Example 9: The nightingale of the revolution

HAMA, SYRIA, JUNE 27TH, 2011. The market square is packed to overflowing. Crowds throng from side streets. It is evening. A local poet, disc jockey, and fireman, Ibrahim Qashoush is leading a chant, using a microphone. There are few instruments; most prominent is a drum that picks up the beats of the often-repeated chorus "Yalla erhal ya Bashar," which translates roughly as "It's time to leave, Bashar." The crowd claps with vigor. The chant is infectious. The audience allows Qashoush to improvise two-line verses, and they all join in the chorus, singing the four beats with joy.

This is the brief period in which it seemed that a popular uprising in Syria might overthrow the brutally repressive regime of Bashar al-Assad. Autocratic governments had fallen in Tunisia and in Egypt. There were stirrings elsewhere, in Bahrain, in Libya. Naively, optimistically, the press spoke of an Arab Spring. Qashoush's chant had spread from his hometown of Homs, and each evening, the crowd grew. The uprising was real, and it was gathering strength.

Of course it all went to Hell after that. I write this in 2017 where Syria has descended into an abyss with no end in sight. On July 3rd, 2011, Ibrahim Qashoush was dragged from the Orontes River, killed by thugs from the regime. Science may be blind to the power of the collective voice, but Al-Assad was not. In order to make it perfectly clear what they objected to, his thugs neatly cut out Qashoush's voice box and filmed it beside his corpse, framing the wound so that there could be no doubt about the message. That voice could not be tolerated. Let us move on to less distressing application of the voice.

## Example 10: Happy birthday

THIS SITUATION IS SURELY FAMILIAR. A family crowd, a birthday, a cake. Candles are lit. Someone among the assembled is sufficiently confident to begin singing "Happy birthday to you . . . " Everyone joins in, but not all with equal facility. Unless we are in the singing country of Wales, there are probably several people out of key. There may be even no agreed key. Timing is off. The whole is something of a musical disaster, but it was never intended to be a musical performance. It is a different kind of performance, one that has found uptake around the world. Sometimes translated ("Tanti

auguri a té . . . '), sometimes left in English, as I have seen in China, Brazil, and elsewhere.

Our concern with joint speaking ensures that no fine line can be drawn between speech and song. Here we have strayed far into sung territory, but the ritual like nature of the singing, the lack of musicality, and the performativity all serve to make it a useful source of insight into joint speech practices. What can we learn?

Look at the eyes. During the singing, the eyes are free to wander. There is a central focus, usually a cake, but there is no obligation to look there, and the eyes are free in a way they are not in conversation. Individuals glance every which way, now at the birthday child, now at each other, the cake, the floor. It is as if the eyes are off duty. There is no need to negotiate the common ground while we find ourselves within the space of ritual. Sometimes the absence of something is more telling than its presence.

## Gaze and speech

Once we have chosen to examine vocal behavior independently of writing, of recordings, and of technological mediation, there are a lot of new things to observe that are not to be found in grammars or lexicons. The fact that utterances are fleeting becomes important. How else are we to understand the repetition we reliably find? The ephemeral nature of the utterance makes the interpersonal context in which speaking happens important. Presence, or liveness, infuses all spoken words, giving them a necessary connection to the shared ground we occupy, and making them inherently dialogical or even political. The musical elements of the voice stand out, and become an integral part of the act. Gesture, or to use a rather misleading phrase, body language, necessarily accompanies speaking.

Beyond this catalogue, which could certainly be extended, there seems to be a necessary connection between the eyes and the voice that we must examine if we are to make sense of what is going on in speaking. Among mammals, gaze is usually aversive. Dominance is signaled by meeting gaze, submission by avoidance. With the apes, though, gaze becomes much more interesting. Perhaps it is partly because of the position of the infant when nursing: An ape cradles its young, and eye contact between mother and infant is both inevitable and irresistible. Gaze among apes and humans is neither simply affiliative or aversive; it is, rather, a currency or medium used to regulate our social interactions. We don't talk about it much, but we understand the power of gaze without any further elaboration. Stare long

and hard at someone you don't know, and you will quite likely be punched for your trouble. The "contact" we make with the eyes is as effective as bodily contact. Meeting the gaze of another is fraught with opportunity and threat. A beggar in the street will frequently try to establish eye contact, because when that happens, there is a sharing of a moment, a commonality, that makes an appeal for alms far more compelling.

**Gaze during a two party conversation**

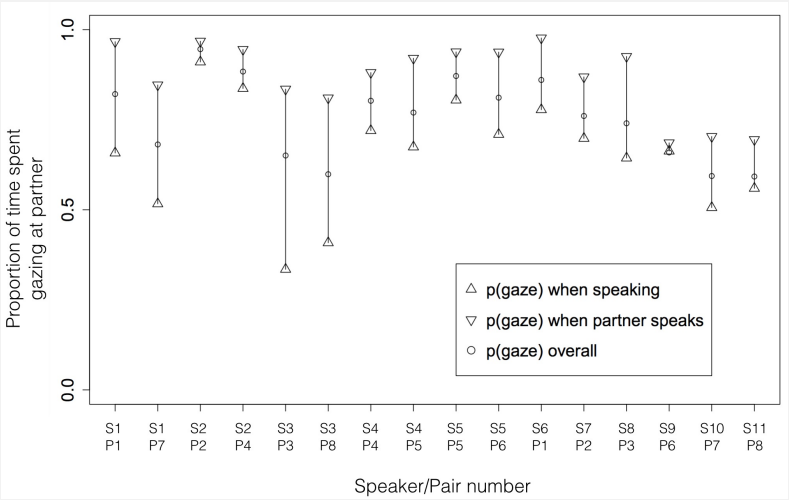


*Figure 25: Proportion of time spent gazing at the conversational partner within a dyad as a function of who is speaking. Adapted from Cummins (2012).*

Plot of the proportion of time each speaker spends gazing towards the face of the fellow speaker in a two-party conversation. The amount of time gaze is directed towards the partner is higher when listening than when speaking for every participant.

Of course we regulate our interactions using the voice as well. It is thus no big surprise to find that the eyes and the voice work together closely when we speak (Cummins, 2012). For example, in a dyadic, or pairwise, conversation, the speaker, having taken the floor, will usually look away from the person spoken to, while the listener will usually look towards the speaker (Figure 25). Gaze patterns arising as one speaker relinquishes the floor and the other takes over are more complex, but part of the reason

conversations work as smoothly as they do is because gaze is co-orchestrating the interaction. The way that gaze works within a conversation is vastly more complex than such a small finding might suggest. The relationship between speaking and looking is not a mechanical one. The eyes act in concert with the voice, linking us reciprocally to the other, binding us. If the conversational utterances are the punches of a conversation, then the eyes might be thought of as the fancy footwork of the boxers.

### Liveness

*Figure 26: Freddy Mercury of Queen leads mass synchronization at Wembley Stadium during the Live Aid concert, 1985.*

Liveness is not an all-or-nothing thing. "Is it Memorex or is it live?" was an old advertising slogan that asserted a simple distinction between a live performance and a recording. But liveness comes in many degrees. We may participate fully, as when we attend a concert (but perhaps the sense of liveness among the musicians who are playing is even greater), but we may also enjoy watching a "live" performance at home. Even if we watch a recording of a "live" concert years after the fact, we will distinguish that from an edited film that never bore a moment-by-moment correspondence to events in any time line. In an

age in which social media allows a new kind of mediated interaction, not live, but not quite not-live, the role of temporality in interaction will need to be given careful consideration.

Interesting, then to contrast this audio-visual fabric of the conversation with the way the voice works in joint speech. If there is a leader, as in our preaching example, or the human microphone, or countless others, then the eyes will be trained on the person who leads. The collective subject has a center. A birthday cake may function in this manner too. When we examine the relation between the audience and Freddy Mercury, lead singer of Queen, onstage at Live Aid in 1985, we see a crowd brought into a coherent focused whole. Mercury leads a chant and an entire stadium responds in rapt attention. The words are irrelevant, but they are known to everybody. ("All we need is Radio Gaga" can hardly be construed as particularly meaningful.) The power of the event comes from the way in which the individual disappears and is subsumed into the whole. The shared gaze is undoubtedly an important part of this process. That particular performance has often been applauded as the greatest live show in the history of rock music. The liveness is absolutely central; this is not something one watches, but something one takes part in. . The individual is temporarily subsumed within the superordinate event. The performance bears an uncomfortable similarity to other great spectacles of mass synchronization, as in the Nuremberg rallies of the 1930's or in contemporary displays in North Korea, with the dictator at the center.

But gaze at the leader in joint speech is not dialogical as it is in conversation. There is no sensitive moment-by-moment regulation of the direction of vision. There is, effectively, no eye contact. When we move now to situations of chorusing, when all present recite a known text, without any leader, we find that the eyes are remarkably free suddenly. If we all proclaim with one voice, there is no "other" to receive our gaze, to stare back. This is quite unlike conversation, for in chorusing, we are not negotiating. We are not dynamically creating, defending, extending and relinquishing common ground. We stand together, projecting outwards with our voice. This analysis seems to work for the Credo, but also for Happy Birthday. Being together is not always such an earnest business.

## Language evolution revisited

Recognition of the importance of the eyes in conversation may allow us to reconsider the mystery of language "evolution." Recall that it was consideration of the remarkable difference between humans and apes that caused us to put clear blue water between our concerns with joint speech (as old as humanity) and the concerns of contemporary linguistics (heavily influenced by writing and texts). As we continue to think about how the voice has functioned over the course of human history, and, more importantly, pre-history, it is worth bringing the eyes into the conversation. For here, unlike in the brain, there is a clear biological difference between humans and apes. When we reconsider the manner in which eyes and voice work together, there is one superficial change within the human lineage that may have had profound consequences.

The eyes have it

*Figure 27: Top: Human eye; Middle: Ape eyes; Bottom: Ape eyes as portrayed in Planet of the Apes (2001).*

The human eye has a white sclera. This is the "white" of the eye, and it is a recent innovation. Our ape cousins have dark eyeballs. If you have a white sclera and a colored iris around a black pupil, the whole becomes a stark signal that announces the direction of gaze, and hence conveys the focus of attention. When two chimpanzees interact, one will quite likely be sensitive to what the other is looking at, but the means by which this is achieved seems to be mainly based on the orientation of the head, and not the eyes (Okamoto-Barth et al., 2007). Chimpanzees and bonobos are more likely to follow gaze signaled by head orientation than gorillas, who in turn are more likely than orangutans. But the white human sclera provides us with a far more informative signal about the direction of attending of our fellow human. This has led Michael Tomasello to suggest that the small biological change, from dark to white sclera, may be a key innovation in the story of how our species came to be so wrapped up in a world of shared

attention (Tomasello et al., 2007). The cooperative eye hypothesis points to the role of the eyes in publicly signaling gaze direction as a possible basis for the development of richly cooperative, shared and social human worlds.

Infants are sensitive to the direction of gaze of their caretakers. More than that, within the first two years of life, infants progress from being aware of the direction of gaze of the other, to following that direction, and soon thereafter, to influencing and manipulating the direction of gaze of the other (Carpenter et al., 1998). From birth, infants are swaddled not only in cotton and linen, but also in copious folds of joint attention. When we pay attention to the same things, our worlds align. An external threat will be recognized and evaluated in the same way by both parties. A novel opportunity will appear to be solicitous to both. To pay attention jointly is to share a perspective. It is to dance together, and not in competition.

If Tomasello and colleagues are right, they may have identified the most important evolutionary change that facilitated the development of language. Not the language of the linguist, with its insistence on messages and codes, and its denigration of the medium of interaction, but the language of the voice, which can piggy back on the situation of joint attention, using sounds to further pick out common topics, to point and to signal a common stance with respect to things. The link between the voice and the eyes may be considerably closer than we can easily recognize, in a technologically mediated world of disembodied, soulless voices.

If these considerations should prove to be a fruitful way of considering how the voice and the eyes work together in allowing us to continually and reciprocally influence each other, then they provide a rather different view of what it is that served to so radically alter our species. The notion of common ground that we have been pursuing would seem to be far more naturally at home in patterns of interaction that rely on the dynamics of joint attention than in the abstract, unsituated world of text and messages. And once more we would be led to the extreme commonality established by speaking as one; freeing up the gaze during chorusing, whether it be the Credo or Happy Birthday, frees all the participants from the tussle and negotiation of conversation, allowing a common stance, or origin, from which to face the world.

## Joint speech as an object of study

In the past few chapters, we have dropped in on several highly specific locations within the scientific world. We met the experimental paradigm of

synchronous speaking, where participants read texts given to them by an experimenter with the injunction to read in synchrony; even in this anemic experimental setting, we saw that some matters of interest to several different scientific specialties became apparent (Cummins, 2003a, 2009). The speech produced in this manner has some interesting phonetic characteristics; most notable is its lack of variability, or perhaps we should say, its radical determination by the context of its elicitation (Cummins, 2004). There is considerable "science as usual" to do here. The phonetician can further examine the manifestation of speech under these circumstances, perhaps expanding the inquiry to look at the articulatory movements of synchronized speakers as well as the sounds they produce. One could further pursue the perception of speech by multiple speakers, and one could vary the context of elicitation in many ways, e.g. by introducing rate or intensity variations, examining the link between synchronization and familiarity, or dialectal distance. In short there is much work to do here for phoneticians.

Movement scientists are also presented with an interesting case study that is singled out from all other synchronized activities by its lack of scaffolding by either a regular beat, or by strong inertial, elastic or gravitational constraints (Cummins, 2011). This raises questions that can be pursued experimentally, and the results will add to our understanding of coordination within and among individuals as a function of task demands, context, and more. Once more, there is work to do.

The neuroscientific findings of Jasmin et al. are particularly tantalizing (Jasmin et al., 2016). The observed characteristics of cortical activity in the case of live reciprocal interaction are substantially and intriguingly different from the activity found when speaking along with a recording, or when speaking or listening as an individual. This is not straightforward to accommodate within existing cognitive neuroscientific models, but there is such a great deal of variety, and so little stability, in models within this field, that the finding becomes a valuable landmark that subsequent studies can do much to elaborate upon. It is eminently plausible that subsequent findings might constructively contribute to the more established results within the field. The apparent alignment of the neuroscientific results–which suggest an alteration in an important marker of the boundary between self and other–with the familiar experience of choir singers of an altered relation between self and other while chanting, provides encouragement that this brain-based approach might prove to be useful in the interpretation of individual experience. The stark difference between brain activity when participants were in interaction with a live person compared with activity arising from speaking

along with a recording may be of use in formulating the future agenda of what has been called "second person neuroscience" (Schilbach et al., 2013).

**A social psychology experiment**

*Figure 28: Bob, on a unicycle on a tightrope, is displayed on a large screen. He is in constant danger of being unseated by incoming tomatoes, generated at random. Participants have clickers with which each can administer a small nudge right or left. The collective goal is to keep Bob upright as long as possible. Adapted from Von Zimmerman and Richardson (2015).*

There is one small study within the field of social psychology that examined the effect of a little chanting (of lists of words) on subsequent group performance in a collaborative task (Von Zimmerman and Richardson, 2015). The task involved the projection of a large image of a man (Bob) on a bicycle balanced on a tightrope. Every now and then tomatoes appeared from the right or left, striking him and making him wobble. Groups of about 30

people watched this, and each could administer a small compensatory nudge to Bob using a clicker. The desired result of keeping Bob on his bike (or the inevitable end state of watching him fall off) was an emergent property of the actions of many individuals working together, without any central locus of control. Results obtained indicated that the groups who chanted in synchrony before the task did better overall, i.e. they managed to keep Bob on his bike for greater levels of tomato intervention, than those who did not chant together. The chanters functioned better as a collective agent. This is a small finding, situated within social psychology, whose methods and insights do not travel particularly well to other areas of science. But it does add to the growing list that suggests that joint speech makes a fine object of study from many points within the varied landscape of current scientific specializations.

To add, incrementally, to established findings within a specific scientific field is potentially useful, but it is not headline news. The marked absence of the thematization of joint speech within all potentially relevant scientific fields suggests that there is more going on here than that. It is notable that when we considered joint speech as a form of language behavior (and now the term "languaging" might begin to appear more reasonable), it did not simply add to findings within the established discipline of linguistics. Consideration of joint speech as a central, ancient, and important part of human vocal expression caused us to question the boundaries of language as an object of study (Cummins, 2014c). It suggested that the emphasis of contemporary linguistics on modality-independent formal properties of sequences of tokens within an abstract system may not be the most insightful way to study language, i.e. to understand what it was that so differentiated our species from our fellow apes, giving rise to unheard of levels of coordination within a shared culturally-tinted world. It alerted us to the fact that a message passing metaphor does not cover all that one might need to cover in studying language, and may, in the case of joint speech at least, be positively misleading. It allowed us to make highly suggestive links from joint speech studies to additional work within developmental studies and primatology, that suggested that the voice and eyes interact in the support of joint attention, thereby allowing a common orientation towards the world. This is not simply incremental. It is to use joint speech as an instrument, rather than merely as an object.

Ancient joint speech?

*Figure 29: Kesh Temple Hymn written in cuneiform script.*

This is an inscribed tablet from Sumeria (present day Iraq) containing the text of the Kesh Temple Hymn. It is probably the oldest piece of literature in the world. This tablet, from the Walters Museum, Baltimore, MD, dates from about 1600 BCE, but examples of the text from as far back as 2,600 BCE have been found. The text was thus in use for about 1000 years. The text itself consists of a series of 8 verses, each ending with a common chorus, which reads "Will anyone else bring forth something as great as Keš? Will any other mother ever give birth to someone as great as its hero Ašgi? Who has ever seen anyone as great as its lady Nintur?" The verse-chorus structure is, of course, still very popular today in song and music. Importantly, the use of identical lines, repeated at the same place in the structure at the end of each verse, suggests that this may document archaic joint speech, used within a liturgical structure. Joint speech is thus probably older than written literature, and has played a role in founding societies for a very long time indeed.

## Joint speech as an instrument

Many of the instruments of science serve to quantify observed phenomena, events and objects, we encounter. A ruler, a thermometer, or an accelerometer produce quantitative data, the bread and butter of science. But many of the instruments of science do something rather different. They allow us to make observations by directing our attention to phenomena that might otherwise be missed. The telescope and the microscope enlarge the scale of our vision by making the tiny and the distant perceivable. They do not make measurements directly, though once they have served to frame an object of study, measurements of many kinds can be made. We noted earlier how time-lapse and high-speed photography enlarge the temporal window within which we perceive events, making slower and faster happenings visible. The MRI scanner and the cyclotron of the nuclear physicist expand the range of phenomena we can observe beyond any simple extension of our sensory systems.

Joint speech may act as an instrument, as well as an object of study in its own right. When a phonetician uses joint speech to rein in the wild variability of speech in order to make some specific features manifest, joint speech is a tool, not the thing being studied. When we attend to joint speech and thereby recognize commonalities, rather than differences, between speech and music, our observations have opened new avenues to explore that were not previously available. When we thematize joint speech and use it to point out the limitations of a conventional message passing view of language, new questions arise for empirical investigation that were invisible before. This more resembles the role of the microscope, which frames our vision and allows interesting phenomena to grab our attention, than the thermometer, which provides a quantitative readout that is of use after phenomena have been identified, observations framed, and hypotheses possibly phrased.

It is my suspicion that joint speech has a great capacity to act as an instrument in this way. As an empirical phenomenon, in most cases, any two language-enabled observers will be able to agree when they are observing an instance of joint speech. There are certainly grey areas, but there are very many more clear instances. This allows for common observation, which is a theme we will develop further in the final section. Just as a microscope brings the rich world of single-celled animals to our attention, without telling us how to sort out the pathogens from the friendly organisms, or telling us how to classify and order our observations, so the instrument of joint speech, turned on human social practices, brings domains of collective intentionality into view. It motivates the conjoint study of ritual, liturgy, protest, and the

affirmation of identity on the football terraces as specific kinds of activity that share important empirical features, and that thus may be studied together. It does not tell us how to do this, nor does the bare fact that people are engaging in joint speech tell us anything much about what is going on.

The domains that are singled out by the framing device of joint speech have not previously been linked, because such activities have appeared to provide thin gruel for scientific inquiry. The structure of a liturgy would be understood to be of interest perhaps to religious studies scholars, or other experts in the domain of culture, but culture is usually treated as if it were outside the pale of scientific inquiry altogether.

However, it seems clear that the invisibility of joint speech practices for the business of scientific inquiry is better understood as a reflection of a blind spot within science itself, specifically when it comes to the notion of the subject. I have suggested that the manner in which the subject is treated within the human sciences, life sciences, and cognitive sciences is unduly restrictive. By attributing subjectivity, intentionality, and agency to a single kind of entity only, an individual human body that is equated with an individual human person, we miss a great deal. Furthermore, this specific commitment turns out to be entirely optional. If we pull back a little, it itself takes on the appearance of a contingent culturally specific practice that bears the marks of its own historical development in a West European, post-Renaissance, post-Reformation, Enlightenment, Christian setting. As with any complex field of human endeavor, there are many influences. And without a keen awareness of those influences, and how the current practice of the sciences of the living came to be, it is easy to think that things must be so and could not be otherwise.

But now a much larger picture comes into view. In thematising joint speech, we find ourselves confronted with the limitations of a specific view of the person-as-subject upon which much of the human sciences have been built. When we exercise a critique of the subject we are asking questions about personhood, subjectivity, and identity. The critique that starts within a relatively narrow area of science suddenly bursts its banks and the questions multiply, and become political, moral, and theological. We no longer stand before an intellectual puzzle; we become enmeshed in arguments that show no respect for the boundaries that tame our world. They concern us, but in many different ways. From contemplation of the stars as unassailable referents, objective, distant, inert, we are pulled back to earth, where we live, whoever "we" are.

# Part III

# Beyond Science as Normal

When we illuminate the difficulties of trying to conduct science in an objective fashion suited to astronomy, but our subject matter is now living beings, we need to develop an appropriate suite of concepts, a technical vocabulary we can bring to bear. A contrast is drawn between a positivist, state-based, Parmenidean kind of account and a presentist, dynamic, Heraclitean variety. This distinction allows us to employ the novel and developing suite of concepts grouped under the heading of enaction which, it is suggested, should enrich and enlarge our scientific accounts, rather than overturning them.

Finally we return to the theme of what we might learn as we study joint speech, and the manner in which disputes arising from the conflicting claims of the natural, civil and traditional sources of order might be negotiated. The study of joint speech now appears as one way of cautiously developing accounts that lean on plural accounts of who "we" are, in differing circumstances at differing times. It admonishes us to be cautious about premature generalization, especially with respect to the human species, and reminds us that we each speak from a specific cultural and historical perspective.

# Chapter 8

# An Enactive Starting Point

## A disclaimer

Joint speech is ubiquitous, it is interesting, and there is much to be found by studying joint speech using the conventions and assumptions of contemporary scientific practice. If this small work were to stop here, its work would be done, and there would be a rich empirical agenda to be pursued by many. But as we have topicalized joint speech, at every turn we have run into problems of subjects who struggle to be recognized, described, or even seen when we conduct science in a specific kind of objective key. In this last section, I will attempt to contribute towards the development of a language suited for addressing subjects, but sufficiently robust for use in empirical inquiry. I will fail. The issues here require us to pull back a great deal and to consider what we mean by science, by facts, by reality, but above all to ask frankly, brazenly, who gets to insist that they have the authority to make pronouncements that must garner assent? Any attempt to contribute here must be tentative. When we delineate the shortcomings of a specific kind of inquiry, that might be perceived as a sleight, for the claims of any scientific discipline constitute the basis for exerting a kind of secular power. Attacks on the claims to authority of secular institutions are ten-a-penny (The people demand the fall of the regime!). Indeed, joint speech is a medium through which such objections are conventionally made manifest. Attacks on the claims of religious authorities, or other custodians of the dictates of tradition and culture, are also entirely pedestrian, at least in a secular Western context (c.f. the Church of the Flying Spaghetti Monster). But principled concern about the kind of certainty generated by empirical scientific inquiry tends to be restricted to a relatively small group

of insiders–philosophers of science, social scientists who study science itself, or humanists of various stripes who address the practices of science from a great distance, without themselves conducting scientific inquiry.

I will begin by outlining a few linked concepts at the heart of a relatively recent kind of thinking from the philosophy of biology and the philosophy of mind. Indeed, the enactive tradition explores the notion that many of the traditional questions that revolve around subjects, subjectivity, agency, mind, consciousness, value, and experience, are best framed within a biological, rather than a neurobiological/psychological, framework, that is, as they relate to the goings-on of the living, including cells, organisms, plants, and collectivities of many kinds. This technical language is still being developed from many quarters, and my own use of enactive concepts is slightly idiosyncratic, but it should become clear that the language of enaction seems well suited to the discussion of the kinds of subjects we encounter in joint speech studies. Equipped with an enactive perspective, many of the traditional concerns of the psychological and social sciences appear rather differently, and in the final chapter, we will pursue some of the issues that arise as we move from traditional understanding of representation, minds, and being to enactive understandings of presence, intersubjectivity, and becoming. Finally we will try to gather all these threads together constructively to suggest how one might approach the boundary issues that arise from the colliding pronouncements of science, civil authority and religion/tradition. Doing so means groping, cautiously, towards an altered conception of how science might work for us, for subjects such as us, for some value of "us," to be continuously negotiated.

## Challenges

As we have surveyed joint speech from many different angles, it has been necessary to recognize some limitations with scientific practice as it is applied to the goings on of the living. Joint speech, once topicalized, brings an enormous elephant into the room. It becomes necessary to confront the presence of many kinds of subjects. Any unquestioned assumptions we might have that the practice of science exists to establish objective facts that are independent of all subjects, that depend on no value system or concerned perspective, become untenable under these circumstances. A view from nowhere is not in our gift (Rorty, 1979; Nagel, 1989). We must perforce ask what kind of scientific vocabulary we can use that is not entirely beholden to the individualist presuppositions that arose in a specific historic,

cultural, and theological context in Post-Enlightenment Europe and that now lie at the heart of contemporary human, social, and life sciences. To pursue this line of thought is surely in the interests of the scientific endeavor itself, which aspires to avoiding unacknowledged ideological or non-rational assumptions in its statements. The great scientific enterprise has also historically sought to make statements of maximal generality, going beyond the local and contingent to the general and principled. With such goals, occasional correctives are going to be inevitable, even desirable.

We have met the familiar Cartesian subject, who is nowhere to be found in the observable world, but whose existence we have great difficulty in doubting. As the inner workings of the cogito have been fleshed out by psychologists, providing a home for perceptions, memories, beliefs, dreams, and desires, this creature of the imagination has become what many of us consider to be our "self," the "I" that persists from birth to grave. We have looked in passing at how this particular kind of subject arose within a specific social, intellectual, and cultural context that many–though by no means all–of us see as our home. This is the post-Enlightenment consensus that birthed the notions of human rights, one vote per person, individual accountability, and a neat separation between the three strands of logos–natural law, civil law, and the dictates of religion and tradition. While there may be important aspects to our lives and our being that are not readily approached within a Cartesian framework, the starkly individualist assumptions that have grown around this notion extend far beyond the concerns of cognitive scientists, linguistics, or psychologists. A great deal depends on how we conduct this discussion, and important established structures will not be overthrown simply by turning our collective empirical gaze on the practices in which joint speech is embedded.

But we are also not condemned to a single, positivist, account of the world within which all our assertions of fact must fit. When we approach the topic of the subject, broadly construed, we will find ample reason to interrogate the distinctions between the three branches of logos. Unlike the scientific examples we met in Part 2, this discussion is not science as usual. It is inherently political, and the manner in which we approach the discussion will necessarily lead us to ask about the origin of various claims to authority. If natural law is beyond question ("we don't negotiate with gravity"), then statements that describe natural law must command assent with even greater force than those that stem from government or cleric. But if the boundary lines are rather less fixed, then any positivist claim that things simply are so-and-so and not other becomes, at the same time, an assertion of a kind of authority.

Much of the compelling authority of science seems to come from the indubitableness of mathematics, and the strong belief that if we understand our world and the things in it in the correct mathematical way, some of the certainty of mathematical deduction will be transferred to the statements we can make about the world. This of course demands that we identify a specific correspondence between the elements of a mathematical statement, and things, processes, or relations in the world. To a first approximation, where the statements we wish to make concern the motions of inanimate objects under idealized conditions, this appears to be warranted. However that specific kind of mathematicization is suitable only for describing the motions of inanimate objects under idealized conditions. It is good for domains in which mechanical explanation exhausts what we might want to describe. In a Newtonian mechanical description, the positions of things and the changes of those positions provide the aspect of the world that is mapped into the statements of mathematics. Position is expressed in spatial coordinates of course, and, when discussing matters of existence, each such spatial record is mapped also to a unique time stamp. So we can say that the cat is on the mat at 3 p.m., more or less. We can say nothing about any experience of the world by the cat, or its motives for being there, mind you.

Newton's clockwork world demands that we treat space and time as isometric containers. Things happen "in" space and time, indexed by their three spatial and one temporal coordinates. An unimaginable amount of effort has gone into making this language work for us. This is the work done by the scientific institutes and authorities who have standardized such measures as the meter and the second, freeing them from the messy entanglements of politics. When a standard measure is a part of the king's anatomy, as the cubit, foot or inch once were, space and politics are more deeply entangled, and coordination among distant people becomes more difficult. In order to extricate the measurement of distance from such concerns the meter was brought into being by definition in 1795 as the length of one ten millionth part of the distance from the North Pole to the Equator. This remarkable shift brought the non-negotiable dimension of the Earth into the definition of a standard length, and it gave rise to a flurry of activity to try to ascertain how to measure such an ideal unit. Surveys had to be made that tried to measure as accurately as possible the distance between certain landmarks in distant cities, allowing a retrofitting of the result to the very notion of measurement itself. This was a complicated business that demanded some simpler system for sorting out the mundane concerns of everyday measurement. It would not be practical for every road builder or architect to have to measure the globe themselves. To this end, a series of prototype rods was

produced, each trying to improve in accuracy on the previous ones. Copies of the authoritative meter bar were made and distributed internationally, but the process of copying is error prone, and improvements in the tools of the physicist gave way to a replacement definition, now expressed in terms of the distance travelled by light in a vacuum over a fixed and very short interval. The speed of light thus came to be the gold standard "non-negotiable" yardstick when measuring distance. Measuring the speed of light, is not a simple matter either and it is grounded firmly in the observations we can make of the stars.

The quantitative measurement of time, of course, was always rooted in the observations of the heavens. The predictive powers of science were forged in the provision of advance notice of eclipses and of the wanderings of the planets against the fixed and presumed eternal backdrop of the stars. As with the measurement of space, the story of how we quantify time is one of gradual standardization, achieved through great efforts, and culminating in a kind of practical consensus that allows us earth-bound creatures to coordinate our activities more and more effectively, relying implicitly on the impersonal regularities of the heavens and the development of agreed practices that allow such observations to be turned to practical use.

It would be almost conventional at this juncture to contrast such "objective" measurement procedures and standards with something called the "subjective" experience of time and space, which are rather different. Subjectively, the story goes, time seems to pass at different rates depending on our state of arousal, the flow of attention, and so on. Subjectively, space extends from our embodied center out, first into a space that is within our grasp and that depends on our orientation, then to the near-distant where events and things are potentially, but not actually, in our grasp. From there, there is a distant field where things are simply "far away," and we do not clearly and immediately see differences between something 10 miles and 10 million miles distant. This stark contrast between subjectivity and objectivity is familiar, and it is built upon a single notion of subjectivity that has been allowed into science, and that is attributable to a single kind of subject only, the skin-clad body that is identified with the person; the person, in turn, becomes intelligible only if we attribute to it an inner world that is something like the cogito.

In what sense is the first, isometric, objective view considered more real than the second, personal, view? It is certainly the more useful view when we are navigating the globe, launching rockets into space, building cities, and arranging to meet for dinner. It should not be necessary to argue that these are fine ends, and the standardization of the meter and the second

have furthered them incomparably. Indeed, I scarcely remember how we navigated cities before the advent of the smartphone and Google Maps, which are both only a few years old. Those who study science itself will opine that the isometric view is a wonderful construction, but will insist that it is a construction nonetheless (Galison, 2004; Latour, 1999; Shapin and Schaffer, 1985). The means by which it was constructed extend over many centuries, involve thousands of actors, and have resulted in a set of processes that command some kind of assent. This particular kind of mathematicization has been wonderfully useful for coordinating our activities. But it has not been wonderfully productive in supporting any account of agency, of the person, or of the processes of the living. Indeed, the overly strong commitment to this particular kind of objective picture as if it exhausted reality has led to a problematic enshrinement of the corresponding view of the Cartesian subject, or the individual mind. Those activities that prove so very useful in coordinating our activities (agreeing on how to measure time and space) may end up inspiring a kind of fundamentalism with respect to the person and the collectivities that make up our lived worlds by banishing all subjectivity to the hidden interiority of the cogito.

## Two Greeks

Two landmark figures are useful here, as they provide contrasting and non-overlapping views of how we might frame our accounts of reality. As landmarks, they allow us to orient, sort, and arrange our ideas. As historical figures, they are dim indeed, each of them having written sparsely, in verse, and most of those lines having got lost along the way. No matter; it is as landmarks that we need them, not as authorities. Other accounts are possible, even necessary, but the contrast we obtain by bringing these two specific ways of approaching our being in the world is, I hope, what we need here.

Two landmark figures

*Figure 30: Parmenides (left) and Heraclitus (right), as depicted in Raphael's School of Athens.*

The first point of reference we need will be associated with the pre-Socratic philosopher Parmenides. Parmenides provides a reassuring, but nevertheless perplexing, account of existence. Existence is his main theme. This exists, that exists, and that which exists could not not-exist. Parmenides gives us a picture which in modern garb would be a static four-dimensional block of space time, within which everything that has happened, that is happening, or that will happen can be indexed. To many people, "existence" is precisely what they mean by reality. If it exists, it

is real; if it doesn't, it isn't. This can be mathematized readily by using the law of the excluded middle, which allows an assertion that something, p, exists or not, but there is no third option available. Existence and truth are here intimately linked. Of course, the law of the excluded middle stems from the domain of logic. It is thus a mathematical statement that may, or may not, map usefully to some things in the world. It has no empirical content in its own right.

The mathematization of mechanics that indexes positions, velocities and times, fits in nicely here. So do many other ways of drawing mathematical descriptions of our world. The natural language of many branches of science is provided by dynamical systems theory (DST). The name is unfortunate. It is not a theory, but a branch of mathematics that may be applied in describing many kinds of things and processes. In describing some actually occurring process dynamically, we must provide two things. The first is a description of the state of the thing to be described. Within a Parmenidean kind of account this would mean that we can come up with numbers that capture the relevant aspects of the thing to be described at a specific time. This is a state description. It serves to describe a momentary condition at a specific time. The second thing we must come up with is an equation that describes how this state changes over time. This is a differential equation, and is sometimes called a dynamic. I will spare the reader any worked examples, as an equation in an otherwise readable book tends to sour the mood—but also because I wish to keep this overview of how mathematics relates to the world maximally general and abstract. If we were dealing with a mechanical description in a Newtonian sense, our state description would be a set of numbers describing the position and velocity of individual elements over time, and Newton's three laws of motion would provide the dynamic. But we might be describing the reaction of two chemicals, in which case the concentration of each at a given moment might provide a state description, and equations would describe how changes in one relate to changes in the other. Or, to move to a slightly less material example, we might be modeling the economy, in which case the amounts of money in various locations, the number of employed persons, and the various entries in an accounting ledger at a specific time might provide a complex state description, while economists would vie to provide equations that would predict change in this state.

We could go on. The important commonality here is that existence in this kind of framework means that you can describe how a thing is at a specific time. In a lawful universe, then, that state description will change lawfully. (The plausibility of deterministic change in the economy is left

to the reader to gauge.) Many activities in many scientific fields do exactly this. Whether we are predicting eclipses, measuring chemicals in a test-tube, or modeling stresses along tectonic plate junctions, the mathematization of reality takes this form.

There is a well-known fly in this ointment. In such a deterministic vision, the manner in which "change over time" is treated does not match our lived experience of time progressing from past to future at all. Time is treated as if it were a fourth spatial dimension, and there is no way to express the passing of time, or the process of change, expect by analogy to the curvature of shape. The temporal axis is equivalent to a fourth spatial dimension. This is an eternalist vision, suited to an all-powerful being to whom the future and the past are of equal status, for there is no way to express the notion of "now" in such a picture except as an infinitesimal. In an account of this sort, all of time is laid out before us, and it is not meaningful to distinguish between past, present, and future.

In order to come at things differently, we will turn to Heraclitus, a contemporary of Parmenides. We met Heraclitus earlier as one interpreter of the rich word "logos." For Heraclitus, all is flux, all is change. It is from Heraclitus that we get the notion of panta rhei, or "everything flows," from which it follows that you cannot step into the same river twice. If Parmenides is a philosopher of existence, Heraclitus is concerned instead with becoming. The notion of time appears completely differently in a Heraclitean framework. Where Parmenides lays out all of time before us, unchanging in itself, Heraclitus is a presentist, emphasizing the reality of the lived moment in which all that is comes into being. These are not two different ways of talking about the same thing. The Parmenidean and the Heraclitean frameworks are orthogonal rather than complementary. We cannot simply assume that insights and statements situated within one framework can meaningfully be translated into the other.

Existence and becoming are radically different ways of approaching that which we might want to call real. Neither can claim to be an exhaustive framework capable of encompassing the other. In various forms, all metaphysical systems seem to draw from both the notion of existence and the notion of becoming. Things that can be clearly described within one framework may be entirely inexpressible within the other. Yet within Western science, it is the Parmenidean approach that has been most influential, to the extent that it may be hard to see that it is just one, out of several approaches, that one can take. It is no coincidence that different religious and philosophical traditions have drawn in varying measure from these two basic stances with respect to time and being. The German existentialist Martin

Heidegger belongs squarely in the Heraclitean camp, as does the process metaphysics of Alfred North Whitehead, while his collaborator Bertrand Russell would be at home with Parmenides. Buddhist notions of the way in which things come into being are well aligned with Heraclitus, whereas Christian accounts of an earthly life, followed by an afterlife, draw more from a Parmenidean view of time (with the addition of free will, etc.). Because both philosophers have left so little in the way of textual particulars, we can use them to signal fundamentally different stances with respect to time (eternalist vs. presentist) and being (existence vs. becoming), without worrying too much about whether we have accurately represented the views of either. The actual opinions of the two men are largely irrelevant in this context, and are not clearly expressed in their small textual legacies.

## Enaction and The Embodied Mind

At this point, we are ready to take account of a remarkable volume that was published in 1991. "The Embodied Mind: Cognitive Science and Human Experience" was, in many respects, a highly unusual book (Varela et al., 1991). The three authors brought diverse expertise with them. Francisco Varela was a neuroscientist who had worked on an abstract theoretical model of the nature of life itself together with the Chilean biologist Humberto Maturana. The theory they developed, based on a central concept of autopoiesis, or self-production, sought to characterize life on its own terms, and with the added reflective dimension of including the observer in any characterization of the living. The theory of autopoiesis extends to matters beyond our present concerns, but a remarkable feature of the approach is the way in which it emphasizes the necessary link between the observer and the phenomenon being observed. This might remind us of the way in which the observer plays a role in framing any account of behavior. Movement becomes intelligible as behavior just as we frame it as arising from some goals, and these goals are imputed by the observer. This implicates the observer in any account of the behavior and ensures that the ensuing description cannot pretend to be bluntly objective in a mind-independent sense, as if that which is described simply exists, in the spirit of Parmenides.

Varela was also a practicing Buddhist. Evan Thompson was a philosopher who had for his whole life been steeped in both Western and Asian philosophy and his work still represents a unique fusion of insights from traditions across the East-West axis. Eleanor Rosch was a psychologist famous for her ground-breaking work on categorization. She was also a practicing

Buddhist. Buddhist philosophy formed a central part of the account of experience provided in the 1991 book, which very carefully sought to introduce central notions of Mahayana Buddhism to a Western scientific audience, and on that basis, to arrive at a way of approaching the puzzles of experience that avoided some of the pitfalls of traditional accounts that insisted on a single subject/object dichotomy.

Much of the book was concerned with bringing a radically different perspective to the consideration of lived experience. Where Western psychology and (to a lesser extent) philosophy of mind relied more or less on a separation of mind and world, linking subjects to their world only through the medium of representation, the authors advanced a very different view that saw subjects as continuous with, and inextricable from, their worlds. The mixture of biology, neuroscience, and Buddhist metaphysics was, if nothing else, unique in the already multicolored landscape of theories of minds, experience, and behavior. Its concern with the analysis of experience lay closer to the phenomenological tradition, especially that of Maurice Merleau-Ponty and Hans Jonas, than to orthodox positivist positions within cognitive psychology.

At the heart of the book's argument lay an approach to the way we experience ourselves and the world that is strongly Heraclitean in character, and that makes most other approaches appear positively Parminedean by contrast. For reasons that go far beyond our concerns in this book, the authors strive to find some way of avoiding a traditional dichotomy between a form of realism that insists that things simply are so-and-so, in a purely objective key, and an unsatisfactory alternative which is an idealism in which experience is given priority and the reality of the material world seems to be threatened. The approach taken will turn out to be helpful in the task before us, of finding a way to treat of many kinds of subjects and many kinds of associated worlds. Both of those positions, which we might glibly label no-mind and all-mind, are problems that arise when we approach subject/object relations as Parmenideans, trying to establish that-which-exists, and thereby assuming a single kind of subject/object split. The alternative that is put forward is to come at things as Heraclitus might, and to observe that human experience is grounded in the present, in which a coming-into-being happens in which both subject and associated world arise together and in conjunction with each other. This is not an easy idea to express in language more usually suited to discussing matters of existing than becoming. We will flesh it out shortly, with no more regard for fidelity to the original 1991 book than is necessary as we try to develop a language suited to the issues of subjecthood that arise in understanding joint speech practices.

The book took off slowly. Initial reactions were more perplexed than hostile. Daniel Dennett, a central mainstream figure within cognitive science gave it a cautious but rather positive review (Dennett, 1993). He noted that the authors had reached more or less the same conclusion as he, that there was no essential psychological self, no unified mind, present and persisting from birth to grave. Mind-based accounts arose from narratives we tell ourselves after the fact. He also noted that the authors had bent over backwards (perhaps too far, he suggested) to avoid seeming to dismiss other perspectives within cognitive science. Rather, they sought to position themselves carefully with respect to very many other researchers and theoretical positions. Such respectful treatment of others is not the norm in science.

It is 25 years now since The Embodied Mind first appeared. A new version has just been published to celebrate the anniversary. The ideas within the book have given rise to a burgeoning field knows as *enaction*, and the central concerns of enaction have found rich interplay with several other approaches to understanding minds, experience, and behavior. Many of these supporting approaches make use of the term embodiment, though they do so in a variety of ways. For our purposes, embodied approaches to cognitive science will be those that assume that experience arises as a function of an active subject interacting with a meaningful world (for some yet-to-be clarified sense of both subject and world). This is a stance that is entirely at odds with more orthodox approaches that assume that experience arises in the brain, and that the subject is something that exists independently of its world (and vice versa). I like to contrast the two approaches thus: The Parmenidean account describes a world that exists. Somewhere in that world (where?) mind-or-consciousness-or-experience is assumed to be found–often described as "in the head" of the person. The Heraclitean approach starts with the indubitable reality of present experience, and then seeks to characterize both subject and world, both knower and known, by extrapolation from the lived center.

Somewhat surprisingly, overt discussion of the Buddhist content of the seminal 1991 book has been almost completely absent as the ideas have found further development within cognitive science. Yet it is the injection of a non-Christian sensibility that gives this work a lot of its force, allowing it to distance itself from the problematic assumption of individual separate minds (invisible, unobservable) exerting complete control over bodies which interact with a mindless world. In place of this familiar scenario, we get a language with which we can describe a triadic relationship of subjects, environments, and the observer, that will be of use in trying to understand what is going on in joint speech and its associated activities.

## Of systems, worlds, and observers

A self-organizing system

*Figure 31: An eddy. (Image credit: Wikicommons user Shutinc, CCA-SA 3.0.)*

Let us start in truly Heraclitean mode. All is flux. Everything is changing. Yet in all that change, some things seem to persist. Rocks stubbornly remain rocks for as long as we watch them. They too will pass, of course, but while we observe them, they remain fixed. For our purposes, they simply exist. But some things that are dynamic and changing also seem to persist over time. Looking at the flow of water in the river, we see an eddy forming, a little whirlpool sucking in water, leaves, and twigs at the top of the stream, and spitting them back out at the bottom. This eddy is a dynamically individuated system. Prior to its formation, there was just water, all changing, all flowing. Once it forms, we can distinguish between the system and the rest of the water. The eddy is recognizable against a backdrop of the surrounding water, but it is not independent of that background. Its existence does not appear to us to be like that of the rocks. It exists as long as the conditions that gave rise to it exist (flowing water, perhaps a particular obstacle like a branch or rock) and while it exists, it engages in exchanges with its surround, sucking in water, and spitting it back out. This is a useful starting point. The notion of a dynamically individuated system

that persists as long as specific conditions are given will be useful. But the eddy does not yet have the appearance of a subject. It does not act on its own behalf, as far as I can see (Cummins, 2014a).

Now we move to a somewhat more complex situation, that of a single cell (Cummins and De Jesus, 2016). The example of a single cell swimming in a neutral medium towards a food source is the most discussed illustrative example in the enactive literature. We will use it to draw out just those features we need for our purposes here. Much like the eddy, the cell is a dynamically individuated system. It consists of a bunch of chemicals mutually interacting, and it swims in a medium that also contains chemicals. We have no difficulty in distinguishing the cell from the background, mainly because there is a physical barrier, the membrane, that carves up the space into cell and environment. But on closer inspection, the membrane itself is somewhat porous, so that it alone cannot completely underpin our distinction between cell and background. Of greater importance is the identification of the group of chemical reactions going on within the cell, those that keep it going from moment to moment, ensuring its continuation as a dynamically individuated entity. This suite of chemical processes is really what makes the cell what it is. We could exchange all the material components of the cell, including the molecules that make up the membrane, and we would still have the same cell. Indeed, in the course of its life, it is highly likely that every atom and molecule of the cell will be refreshed, swapped, or altered, so that the cell cannot be said to be co-extant with any specific set of matter. We understand the cell better if we attend to what is *happening* (Heraclitus) and so much to what *exists* (Parmenides).

The chemical reactions that make up the cell allow a distinction to be drawn between the system that is the cell and its environs. Some components of those reactions have their origin outside the cell. This will include the glucose that stands, in this illustrative example, for all food and externally sourced material that provides input to the system. In Figure 32 we see a schematic representation of the processes contributing material from outside into the cell, producing waste products that emanate from the cell, and those mutually linked process that sustain its continued existence as a cell. Many components are generated within the cell, and they contribute to reactions that generate other components within the cell, so that we can identify a circular network of linked processes, consuming some matter from outside (A), excreting some waste materials (E), but otherwise self-contained (B, C, D). The cell here is a system, distinguishable from a background, but not separable from that background.

The cell considered as a self-organizing system

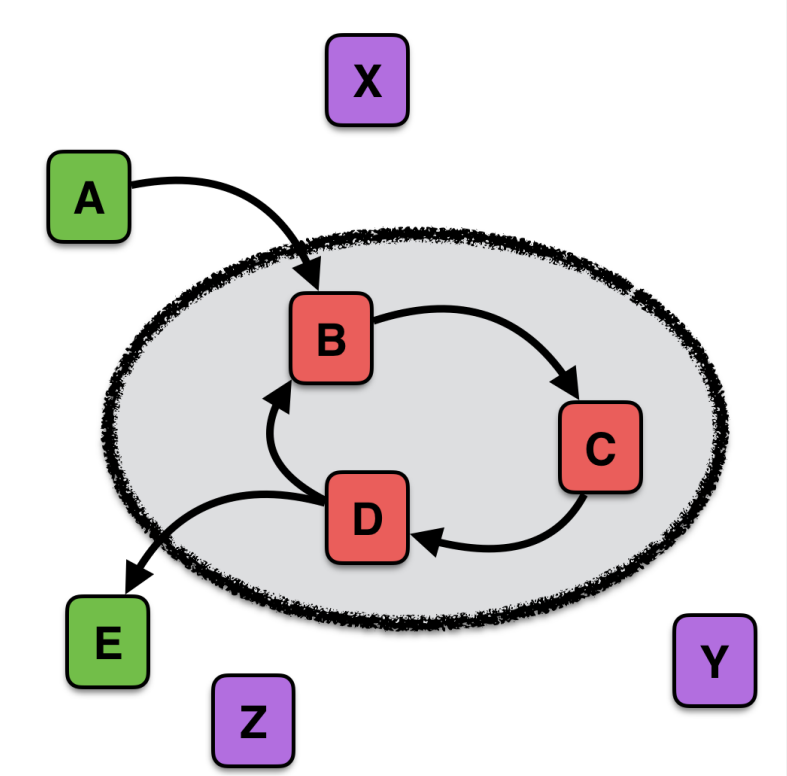


*Figure 32: Chemical processes within the cell (B, C, D) and in regulated exchange with the surround (A, E). Chemicals X, Y and Z do not take part in the cell's metabolism.*

As well as recognizing the domain of the system, further observation of the cell will reveal that it is sensitive to some aspects of its environment, generally considered, and insensitive to others. As we develop this argument, we will let chemical A stand for glucose, presumed to be necessary for the continued metabolism of the cell. But there may be countless other chemicals, X, Y, and Z, to which the cell is insensitive. We need to partition the entirety of the environment of the cell that we can distinguish between those bits that are relevant to the cell (including A), and those that are irrelevant (including X, Y, and Z). We might do something similar with respect to the reader of this book, who sits (perhaps) in an environment in which the play of visible light has a manifest influence on their behavior.

The same environment may be interpenetrated by shortwave radio waves, which are just another chunk of the electromagnetic spectrum, but one to which the reader is entirely insensitive. In recognizing the system as a distinguished domain, we also pick out a milieu, or environment-of-relevance, that is inseparable from the system. They co-define each other, in much the same way as a porcelain bowl defines not only an interior space, but also an exterior. Inside and outside go together, just as system and milieu do. (The term "milieu" here is somewhat idiosyncratic, but the literature is inconsistent here. Where we speak of system-and-milieu, we might also find organism-and-environment, or many other lexical choices.) As we apply this vocabulary to subjects of various kinds, it will become clear that we cannot ever coherently remove a given dynamically individuated subject from its milieu or context.

In order to continue its existence, the cell must be active. It must engage in the kinds of regulated exchanges with its milieu that allow its continued existence. This activity is intelligible (to us) as serving the interests of the cell. The cell thus necessarily appears (to us) to be agentive. In the absence of such activity, the system will cease to exist, which is literally a matter of life or death for the cell. Such activity is called sense-making within the enactive literature. This provides a nice lexical way of avoiding the traps of describing perception and action as if they were separate or separable activities, or as if all subjects were individuated multicellular organisms. This single term, sense-making, serves to describe the self-serving activity of the system (here, the cell) without the unfortunate psychological commitments that talk of perception and action would necessarily entail. In its sense-making activity, we might say that the cell enacts, or brings forth, a meaningful world. This is not an objective world, but an encounter, by a subject, of a meaningful domain of phenomena. If the cell is swimming up a gradient of dissolved glucose, the glucose gradient is inherently meaningful when considered from the perspective of the cell. From a more disinterested perspective, e.g. one we might adopt as scientific observers, glucose is just another chemical, ordered among many others. But from the perspective of the cell, the glucose gradient is of significance to its sense-making activities. When we speak of "enacting a world," that "world" is necessarily laden with significance for its associated subject.

This form of cautious perspectivalism is at the heart of enactive description. The observer differentiates a system from its background domain. Recognition of the system as agentive makes it intelligible to the observer. The agentive activity of the system is sense-making, and in its sense-making, the system enacts a meaningful world.

It is important to note several things as we adopt this kind of language. Firstly, we have not imbued the cell with a mind, with sentience, or with consciousness. Those three terms belong in a different conversation. When we say that the sense-making activity of the cell enacts a world, we are pointing out that its activity becomes intelligible to us observers to the extent that we recognize the perspective of the cell and the web of significance it casts upon the world as it swims. The cell, on this account, is a subject of the kind that we will need to be able to identify as we proceed. Its activities can be described as serving a function, to the extent that we have framed our observations carefully, such that we are describing functions pertaining to the continued viability of the cell.

The cell and its milieu are non-separable in principle. They are distinguishable, but non-separable, just as the eddy cannot be taken out of the river. This might give us pause for thought. One assumption that goes with the conventional apportionment of agency and subjectivity to a single individual only is that this individual retains her subjecthood irrespective of context. This is the justification for taking the person out of their lifeworld and testing them in the white room laboratory of the psychologist. Such an approach to the person belongs squarely in the Parmenidean camp, as it tacitly assumes an essence to the person that is non-relational, and that can thus survive the displacement from one context to another.

With the treatment of the subject in an enactive framework, such dissociation is impossible in principle. Subject and world are co-defining. They also both change over time, in an interdependent way. This strong temporally extended link between subject and world has come to be called structural coupling within enactive approaches. In the Heraclitean world of becoming (or co-dependent arising, in Buddhist terms), entities are relational. The subject and associated milieu have no independent existence, but come into being together through activity, through the agentive actions of the subject. Changes to one will entail changes to the other. Should the cell develop a new kind of chemical sensor, for example, its milieu will change accordingly to include the newly sensed chemical.

(On a side note, this strongly biological picture has provided a very insightful way to consider the processes of evolution that differs somewhat from the story told in textbooks. In place of the adaptation of organisms to the conditions in which they live, the organisms and their conditions of living are seen to co-evolve. It is trivially true that most of what we consider the terrestrial environment has been produced by the activities of the living. This extends to the soils that cover the ground, the gaseous makeup of the atmosphere, and, of course, the patterns of climate. Interested readers may

follow these considerations within an emerging field known as the Extended Evolutionary Synthesis.)

**Self-organized systems enact a common world**

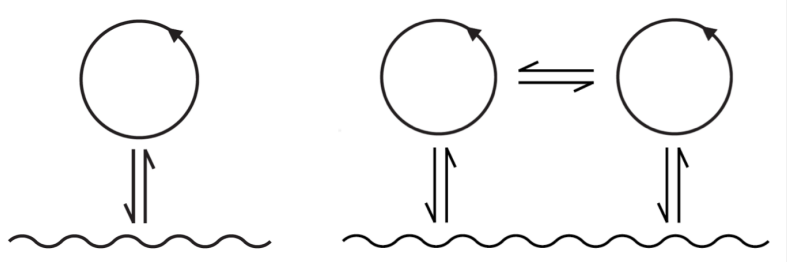

*Figure 33: Left: A dynamically individuated self-organized system perseveres as it engages in regulated exchange with its environs. Its self-serving activity is called sense-making. Right: Sense-making of one system may become entangled with the sense-making activities of others, leading to participatory sense-making, and to the enactment of a common world.*

But our concern here is rather more narrowly circumscribed. We seek a theoretical vocabulary that is adequate to the task of acknowledging, describing, and ultimately better understanding, the kinds of activities found wherever we find joint speech. To this end, we will extend the basic picture of sense-making in a straightforward fashion. Figure 33 employs a graphical shorthand introduced by Maturana and Varela to refer to the sense-making activities of a system (Maturana and Varela, 1986; Moore, 2016). On the left we see a single system engaging in regulated processes of exchange with its environs, which is the process of sense-making. We can say that in its sense-making it enacts a world, though we must be careful with our language, as the system and its "world" are mutually co-defining and non-separable. On the right, this basic picture is extended so that the ongoing sense-making activities of one system become dynamically entangled with those of another (or many others). In this case, the sense-making activities of several systems are strongly dependent upon each other. Indeed, such systems will be in continuous real-time reciprocal interaction with one another, which seems to be an appropriate way to frame the kind of mutuality we find in joint speech observation. We might speak here of the enactment of a common world.

This kind of entangled sense-making has become known as participatory sense-making (De Jaegher and Di Paolo, 2007). It provides a powerful way

to resist the solicitations of solipsism that follow when we consider minds as separate and discrete. If the ongoing regulated exchanges of a system with its milieu leads to the enactment of a meaningful world, then the joint sense-making activities of multiple systems will lead to an inseparability of their enacted worlds. The sense-made will be collective, and it will be necessary for us to recognize both the level of the individual systems, and also the superordinate level of the set of systems whose sense-making activities are so interlinked. Furthermore, as the two or more systems change over time, their changes will also be non-independent. They will exhibit structural coupling with each other, and not just with their immediate environs. A familiar picture used to illustrate this kind of entanglement between the activities of many agents and their worlds is provided by the notion of a "path of desire," or "desire line," which arises when many individuals walk across land on which no paths are laid out. Each crossing will slightly affect the land, adding to an emerging trail, and as the trail emerges, it in turn will represent a more likely path for subsequent individuals.

Anchoring the subject in a shared world

*Figure 34: A Path of Desire created by many individuals with similar sense-making propensities.*

Getting around is a form of sense-making common to most life forms. When many individuals interact with their environment in similar ways, the environment itself will take on properties derivative of those activities, leading to further entanglement between agents and their (shared) world.

## Making use of enaction

The enactive vocabulary may be employed in many ways. We should not consider it to be a suite of concepts suited for pursuing science in a strictly objective key. When we observe the stars together, our perspective is entirely common, and our perspective is the only one that matters. As we observe the stars, they do not observe us back. A Parmenidean framework is appropriate, as we can get by just fine with a strict subject/object divide. But this kind of approach has necessary and important limitations. As we turn to the goings on of the living, the triadic structure of an enactive account–system, milieu, and observer–provides us with tools that may go some way to overcoming such limitations, by facilitating the development of appropriately bounded consensus in acknowledgement of the many perspectives involved.

There is no route from existence to meaning or significance when we conduct our business in a Parmenidean manner, seeking to pin down that which simply is. This is the old Humean adage that one can't get an "ought" from an "is" (the fact/value distinction). The disinterested stance achieves so very much precisely because it refrains from indulging in joint consideration of multiple perspectives. If the worlds we inhabited were exhausted by this kind of scientific activity, then science would be intrinsically uncontroversial. It would establish the mythical view-from-nowhere, nailing down matters of fact, and it would necessarily command assent.

Such a positivist agenda would aspire to providing a view-from-nowhere, which suggests an image or picture revealed by the practice of science. But, shifting metaphors, we might also consider such pronouncements as utterances. Pronouncements, as the word implies, emanate from one who gives them word. We may fix them in text or in image, but the underlying sense is one of words spoken, and the counterpart of a view-from-nowhere must surely be a voice-from-no-one. It is here that joint speech observation might give us pause for thought.

As we, as scientists and investigators, use the definition of joint speech to frame our observations, we are repeatedly drawn to human practices that serve to establish common ground. Voicing in unison is a part of such practices. But everywhere that such speaking goes on, we find other means by which common ground is established too. We find the activities with which we first coordinate our activities in time, building rituals to mark both transitions and recurrences. We find assemblies gathered with common purpose, acting towards shared goals. We find the many ways in which we transiently come together, relieved of the tension of negotiation that characterizes dialogue and conversation, voicing as a single subject. In short, we find a window into the practices that allow us to recognize our common being, defining the "we," and the common world of that "we." What we do not encounter is a voice-from-no-one.

We might crudely summarize the position we have reached as follows: Science done in a positivist, Parmenidean manner seeks to establish facts that must command assent. Such a view of things demands a cosmological or metaphysical picture in which objective matters of existence are established without a residual doubt, but in this picture subjects (along with their notional minds and consciousness) are nowhere to be found. This approach to science has had a great deal of success, but it has relied implicitly upon a notion of individual identity and individual autonomy that is deeply problematic. Such individualism arose in a Western, modern, post-Enlightenment context and it shows. While we have many reasons for cherishing the fruits of such a scientific tradition, we have a great need also to move beyond its limitations. To assert that this view of the subject can demand assent in the same way as our models of the wanderings of the planets is to claim a premature unification among the individuals partaking of the conversation, as if they shared common ground without any need for negotiation. It is to repeat the universalism of the Colonialists and the Crusaders, and it will meet with resistance. Resistance will come from those who are not included, with whom no common ground has been established. Such resistance will not only be from other humans. The domain of the living is occupied at every point by subjects. Animals, plants, and the intricate intertwining of the entire biosphere call to us to negotiate our positions, to settle down together. A science rooted in a Protestant view of the autonomous individual, extending to a human species cast as qualitatively different from all others, is insufficient to assist us in our human lives, and entirely inadequate for addressing, as we must, issues pertaining to our coexistence with the many kinds of subjects distributed over the surface of the Earth.

What do we want from our science? Gadgets are nice. Improved health, longer lives, all accrue. Hurrah. But we also want insight, understanding, and a greater sense of recognizing our continuity with the world we live in. We want explanation that satisfies. The demand for explanation means we need to be able to answer "why" questions. Why did this happen? Why does that persist? Why can we not do the other thing? Why questions, as Richard Feynman astutely observed, require that they be answered within a framework in which we simply allow some things to be true. Without framing our explanations within a framework in which some entities are simply taken for granted, our questions regress infinitely. To every answer, we find another why. Physics does not answer our why questions, it merely provides us with descriptions of observables, from which we might predict future observations. Yet once we adopt this framework or that, we find that we have been presumptuous. Perhaps the entities we take for granted are problematic for our interlocutor. Perhaps your world contains spirits and ghosts that do not exist in mine. If that is the case, then my answers to your why questions are going to be profoundly unhelpful. Perhaps your world contains minds, selves, souls. If so, they will act to buttress explanations of a certain kind. But those explanations have a restricted dominion. They extend precisely as far as the common ground extends.

An enactive, Heraclitean approach to science might serve to help us here. As we recognize the plurality of subjects we live among, and that we even embody, we come to see that agreement is achieved by negotiating and curating that common ground carefully. To this end, we need to improve the manner in which we artificially separate politics, religion and science. The artificial barriers we have drawn up there equip us with wonderful tools for exerting mastery over inert materials, but they leave us stranded when it comes to disagreements. We long for a science free of the trappings of tribe and cleric, but we simply can't have that, and no amount of positivist insistence can change that.

Such an approach is not available to us yet. In the final chapter, I will consider what kind of themes arise if we adopt this radical realignment of our forms of inquiry and agreement. As the Heraclitean and Parmenidean approaches to empirical inquiry do not align or simply map from one to the other, the adoption of a novel Heraclitean perspective will make many familiar issues from the life, social, and psychological sciences recede into the background, to be replaced by themes that were invisible as we worked in a simply objective, Parmenidean, key. The invisibility of joint speech as an object of inquiry was our starting point. Perhaps by addressing it,

we might come to find new ways of integrating empirical inquiry into our collective forms of organization and being.

# Chapter 9

# Making Use of Joint Speech Observation

## Example 11: Extinction

FOR OVER 10,000 YEARS, the Yaghan (also Yamana) and Selk'nam (also Ona) people inhabited the southernmost archipelago of South America, known to us as Tierra del Fuego, or the land of the fire. The name comes from the first visual impression of European settlers who observed very many fires where the indigenous inhabitants lived. It is cold there, with mean daily temperatures ranging from about freezing point in winter to around 10° C in summer. So long were they there that they seem to have acquired small genetic changes that helped them to stay warm. At any given time, there were probably only a few thousand of each ethnic group.

10,000 years is a very long time. It makes for interesting comparison with the roughly 500 year interval in which modern science arose. It speaks of a stable form of life, unimaginably different from my own. What kind of world did they inhabit? How did they regulate their affairs? What entities did they live among? There is little evidence. By the time these peoples came to the attention of Western anthropologists, they were being exterminated. Gold and land were the attraction, and ruthlessness was the order of the day. Between about 1870 and 1950 the two peoples were almost completely wiped out. Bounties were set on these indigenous inhabitants, so that any settler producing the ears or hands of a dead Selk'nam native would get a reward from the incoming commercial organizations. What violent extermination did not complete, imported disease, forcible resettlement and alcohol would finish off. Four Yaghan individuals were kidnapped by the captain of the

first Beagle voyage and taken to England to be "civilized" and presented at court. One died, but the remaining three shared a return trip on the second voyage of the Beagle with Charles Darwin, who observed a great contrast between their behavior and that of their kin in Tierra del Fuego. Of the latter, Darwin noted:

> I could not have believed how wide was the difference between savage and civilised man: it is greater than between a wild and domesticated animal, in as much as in man there is a greater power of improvement. (Darwin, 1909, p. 210)

The anthropologists who documented a little bit of the language and rituals of the original inhabitants were, themselves, missionaries. They interpreted the lives they observed through the categories of the European and the Christian. Colorful initiation rituals were documented. A small Selk'nam dictionary appeared in 1915. The missionaries made more progress with Yaghan. A little searching readily brings up links to translations of the Gospel of Luke, the Gospel of John and the Acts of the Apostles into Yaghan. A set of three archival recordings of the Selk'nam chants of a single person have been captured. This example is incomplete; its subject matter is gone.

## Boundary disputes

The genocide of the people of Tierra del Fuego happened a hundred years ago. The small set of artifacts and texts that remain cannot breathe life back into a dead community. A translation of a book of the bible into Yaghan tells us absolutely nothing about the people of the area, but it speaks volumes about the worldview of the settlers, missionaries, and others who wiped them out. When we look a little closer, of course, we find a more complex picture. The interests of the settlers were not the same as the interests of the anthropologists and linguists, some of who would have seen themselves as trying to agitate on the part of the aboriginal inhabitants. It seems plausible that Thomas Bridges, the missionary who lived among the Yaghan people and compiled an extensive account of some aspects of their language, saw his work as important in documenting part of a way of life that was vanishing rapidly (Bridges, 1933).

The manner in which Bridges approached his task of documenting the language is fairly unsurprising. He sought to capture the sounds of the language, by which he understood the distinctive sound units that one would map to letters, if the language were written. The languages were, of course,

not written, and Bridges had the task of shoehorning the somewhat indistinct and variable sounds he encountered into this literate straitjacket. That he had fond hopes for a civilized, literate, future for the natives is clear when he noted:

> In writing this language I have been much hindered by this indistinctness of pronunciation, being often doubtful which letter was nearest, and have often substituted these letters one for another, and again recurred to the first as nearest the truth. *No doubt when they learn to read this language, their pronunciation will be strengthened.* (Bridges, 1933, emphasis added)

Along with the sounds, we have considerable information about the words and their sequencing, i.e. the lexis and syntax, of the Yaghan language. We know nothing of how those words were used, in what contexts, to what ends. We have no way through the words to the life of the people who spoke them.

It is tempting to consign such horror stories to history. The obvious legacies of colonialism and missionary zeal can be found across the globe, but are usually spoken of in the past tense. Yet today, in 2017, it is still the case that most of what we know about endangered and extinct languages has been collated, ordered, and published by a Christian organization whose principal goal is the dissemination of the Bible. For thousands of endangered or extinct languages, those aspects of the language that are documented are precisely those that are needed to render the written text of the Christian scripture, regardless of the kind of life the unwitting recipients of such attention may lead or have led. The academic study of language has a peculiar history which is intertwined with a missionary Christian organization known as the Summer Institute of Linguistics (now SIL International). This organization was established in 1934, and was later headed by a highly respected language scientist, Kenneth Pike. Today, the organization boasts of a staff of over 5,000. Since the 1950's, it publishes the most authoritative information on extinct and dying languages, knows as Ethnologue. This is the go-to resource for scientific information about languages, as no government or funding agency could match the perseverance and zeal of such a faith-based enterprise. SIL International and its workers are the principal actors involved in documenting threatened languages. It is they that go into the field, spending months or years living alongside their informants, deploying methods of documentation that have been refined over decades to capture precisely those elements of the local speech necessary to allow the bible to be translated.

Here we have a dramatic confluence of the scientific, the religious, and the political, inextricably intertwined. Language as object for scientific study; language as defining characteristic of the only species equipped with rational souls; language as the vehicle for encoding the text of scripture, to be gifted upon an unsuspecting other. And so we find the obscenity of scripture, so beloved of the Europeans, encoded in words cribbed from a group who are being exterminated after 10,000 years of continuous existence. No doubt the Anglican hymns rang loudly within the little mission church in that God-forsaken neck of the woods.

Had the missionaries attended to joint speech practices of the local people, I doubt the locals would have been helped much. Robbed, hunted and infected, it would have been no great succor to know that one's chants and rituals had been seen, recorded, and transcribed. But there was no effort at all on the part of the settlers to understand the world of the local people, incomparably different from their own. The European world was the only world, as far as they were concerned, and the only way forward for the locals was "improvement," conversion, and domestication. The world of the settlers did not overlap with the world of the indigenous, except in its exploitation of the land and its resources.

We do not stand as the settlers did, before a territory we can simply appropriate, inhabited by savages whose redemption depends on the translation of scripture. We can no longer even confidently identify ourselves as a single people. The confident "We" of Victorian days is now fractured, blended, broken and beyond repair. There are many "we"s, and each of us belongs to many communities. We do not have good ways of indexing our plural communities and the mixing tendencies of globalization are no more reversible than the ingredients of a cake once baked. When a nation takes a census, they may ask about religious affiliation, ethnic identity (always a fraught term), language spoken (selected from a drop down list), and citizenship. These meager labels do not do a good job of illuminating the many complex groupings that we together constitute. Our groupings and allegiances themselves are constantly shifting, and global demographics do not appear to be tending towards a stable equilibrium.

Perhaps here we may use joint speech as a means of identifying those human practices that reveal our allegiances, our values, and our many gods. Those occasions in which we (some "we") come together to chant or voice collectively are worth taking seriously. We do not have to travel as far as Cape Horn to be confronted with diversity. It is here, and we live among multitudes. The collectives we create, the collectives that create us, these can be illuminated, but we can do better than the threadbare categories of

race, gender, and creed. They may be revealed in part at least by examining the identities enacted though speaking as one.

The Heraclitean approach suggested by the enactive vocabulary may be of assistance here. It prompts us to seek to identify dynamically individuated entities that persist through their own activity in interaction with a specific kind of environment. To the extent that we are concerned with the affairs of language-using beings, any occasion on which joint speaking takes place looks like a good place to start. The singular role of joint speech in articulating, and bringing into being, enacted identities should encourage us to enrich our observation, e.g. through the strategy of thick description, so that we can identify the relevant couplings between collective and environment broadly considered. This may mean going far beyond the vocal activity alone, and drawing in observation of associated actions and gestures, attending to physical context, as well as to the many different forms of mediated interaction. A dynamically individuated system that enacts its identity as a collective entity is not divorceable from context, just as the observation of a thermostat in a climate controlled environment, or of a fish in a forest, would not reveal anything of their singular forms of organization. We will come with specific kinds of questions. What are the conditions that must be in place to allow the group to assemble? What activities are common, and what are differentiated roles? Under what conditions does the assembly cease? As medical science came to understand the perspective of the body and the conditions that allow it to thrive, so we may aspire to understanding other systems that we live among, and the processes and exchanges that allow them to persist, thrive, or fracture, wither, and die. We can learn to observe, but to do so, we must learn to avoid the blind spot we have identified, and the kind of positivism that rendered joint speech invisible in the first place.

As we adopt such a stance with respect to joint speech practices, the triadic structure of the enactive account must be carefully constructed. It encourages us to recognize the kind of collective made manifest through the practices in which joint speech is embedded (the "system"), in tandem with the manner in which such a collective regulates its being with its environment of relevance (sense-making in exchange with its milieu), all the while bearing in mind our status as observers, complete with our own biases and values. It challenges us to acknowledge and consider the perspective of the collective entity we observe, with its own normative perspective, which may be entirely at odds with any value system or cosmology we happen to bring to the table. This is considerably more cautious than the establishment of matters of "fact." The caution should be familiar to anthropologists, who have

spent decades extricating themselves from the arrogant presumptions of the curious European peering into wild and savage worlds. Anthropologists have learned to be careful, and to be sensitive to the strictures of their own grounding in specific cultures and tribes.

But the negotiation and caution of the anthropologist have traditionally been kept at arm's distance from the positivist aspirations of science. Where such uncertainty prevails, how can we get on with the important business of determining facts? This is the bind of the scientist who insists on working with a single kind of subject/object division. This is the reason the softer social and psychological sciences seem untrustworthy when viewed through the lens of the physicist or chemist, as if all scientists in all fields were engaged in one and the same enterprise, but some had more reliable measuring instruments than others.

The more hard-nosed scientific psychologist who tries vainly to develop measuring instruments that work reliably on all people is attempting the impossible, treating people as if they were stars, and failing. This unfortunate situation seems to leave us with a scientific enterprise that is incapable of addressing human needs, because it sees itself as excused from the need to negotiate, while being required to establish a singular truth to which assent must be given. The harsh rite of peer review prior to publication is seen as the price needed to establish a fact, canonized in the literature. With that the scientist's work seems to be done.

If the order that constrains our actions were neatly divisible into natural law, civil law, and the dictates of culture and tradition, such a stand-offish attitude would be understandable. But things are far more mixed up than that. The sorry example of the eradication of the indigenous cultures of Tierra del Fuego by a people who saw language as a vehicle for forcing the Bible upon them is an indictment, not only of the genocidal antics of the colonialists, but of the blindness that characterizes those who know they are in the right. The contemporary sciences of language, behavior, and the person are still drawn in lines that reveal a genesis in a post-Enlightenment Christian tradition in which intangible discrete souls/minds are assumed to animate individual autonomous bodies.

It is only if we insist on the separation of the individual from the collective, attributing to it a kind of independence no living subject has, that the notion of distinct domains of nature and culture becomes intelligible. The old fallacy of a nature/nurture divide has been run out of town by many thinkers, but it remains as part of the general discourse, and it seems entirely unobjectionable to many. Indeed, the venue which is most likely to insist upon an enforced separation between natural law, as evidenced by biology,

and the dictates of inherited tradition, as provided by parents, educators, and clerics, is in fact that third bastion of order, the court of law. If Jimmy, who did a bad thing, was simply born bad or morally deficient he will be regarded differently, and will be liable to different kinds of chastisement, than if the court finds he was behaving in the service of an odious ideology or a misguided belief system.

Two examples may serve to illustrate the inseparability of the domains of biology and culture, and with that, the challenges we must face as we conduct science with any kind of normativity in mind, sorting out good from bad, beneficial from harmful, or normal from abnormal. In the first case, let us take stock of the local population where you live. Among those folk, there may be a few oboe players, but not many. Most people, if they tried to play the oboe ("ill wind that no one blows good"), would find that they had to work hard to achieve even a modicum of proficiency. Some people, however, would turn out to have a special aptitude for the instrument. They would learn more quickly and with less effort than others. That is entirely unsurprising, and such variability will be found within any population for any arbitrary skill we may define; playing the oboe is really rather arbitrary and something we might consider belongs properly in the domain of culture, irrelevant to biology. Now imagine, however, that there arises a strong economic incentive to play the oboe; an incentive so strong that those who can play the oboe well get better jobs, live longer, and acquire vastly more resources than those who struggle or give up on the instrument entirely. Under these circumstances, there will arise a selection pressure that will favor those whose underlying biology just happens to be favorably configured for whatever it takes to blow on an oboe. This may lead to a reproductive advantage within the population, of course. But irrespective of whether that happens or not, there will now be a strong social pressure to identify oboe-playing skills, to teach them, and to ensure that everybody achieves what they can on the oboe. With that, a normativity arises, so that we would probably develop standardized tests of oboe playing ability. Those who fall outside some limits will thereby be considered to be failing. They may need remedial oboe schooling. We would do this with the honest best interests of the oboe players in mind. With the establishment of a norm, one necessarily brings into being a corresponding class of the abnormal. And so the economic turn of events will alter what we consider to be "normal" human development. Nonsense, you say, oboe playing is not important and never could be. But a very similar development happened once humans started to write things down. Initially writing was done by a few individuals who were regarded as highly skilled, but with the advent of the printing press (at the dawn of

"modernity"), reading and writing became skills that quickly became indispensible for many people, conferring enormous economic advantages. And we developed strong views about when a developing child ought to reach specific milestones, creating standardized tests, and thereby bringing into being the disorder of dyslexia. The biological configuration that suits literacy in one individual, and its counterpart that makes it difficult for another, is entirely arbitrary. Writing was not part of some God-given plan towards which the processes of evolution were marching. Evolution is blind. Literacy arbitrarily became important, and with that our view of what constitutes "normal" development changed, favoring some and disadvantaging others.

A second example is reported by Jeremy Berg (Berg, 2016), who retells a challenge put by Nobel laureate Michael Brown to a class: "How would you produce a new genetic disease in the state of Texas?" The answer he proposes has nothing to do with mutations, but with doors. If the local building codes were altered to prohibit doors higher than six foot, there would arise a clear medical syndrome we might call "bruised forehead syndrome." This syndrome would have a strong genetic component, being more prevalent in males than females, and it would be clearly heritable, as tall parents will have, on average, taller offspring.

## On universality

When Newton developed his magnificent mechanical account of the universe, his desire for simplicity and uniformity in explanation drove him to establish "Rules of Right Reasoning in Natural Philosophy." These insisted that one should "admit no more causes of natural things than such as are both true and sufficient to explain their appearance," and "to the same natural effects we must, as far as possible, assign the same causes," and "[t]he qualities of bodies . . . which are found to belong to all bodies within the reach of our experiments, are to be esteemed the universal qualities of all bodies whatsoever" (Quoted in Olson, 2004, p. 116). These admirably succinct principles allowed him to combine observations of the planets as they orbited the sun, of the moons of Jupiter and Saturn as they orbited their respective planets, and of our own moon as it moved around the Earth. It was a magnificent leap from these astronomical observations to bring the account to bear on the motions of apples and rocks that fall downwards when dropped. The shift from the heavens to the familiar ground on which we all stand was mandated by his aspiration to provide an account that was universally true. The success of such an audacious act of generalization was probably

the single most important factor in enshrining the science of physics as the repository of truth, and cementing the notion in the popular imagination that science uncovered natural law that provided an entirely new kind of certainty, superior even to the authority of the cleric or the sovereign.

But generalization is a dangerous tool that can easily be over used, and Newton's desire for explanatory parsimony does not thereby establish its necessity. The kind of truth that we can collectively arrive at when we gaze skywards needs to be distinguished from that which we can reach as we lower our gaze and look at each other. Observing events in the heavens is relatively straightforward. Equipped with similar telescopes and almost identical positions with respect to the stars and planets, any observation you may make can be communicated to me and I am free to repeat that observation. We literally stand on the same ground. Very little contemporary scientific observations of note are made with such little fuss and with such a prospect of consensus.

The experimental method post-dates Newton and caused great debate when it was introduced. It too sought to facilitate consensus for those cases where we cannot all simply make entirely comparable observations. Central to the adoption of the experimental method was the notion of witnessing, or joint observation (Shapin and Schaffer, 1985). If scientific equipment was hard to come by, expensive, and difficult to operate, then in order to justify its use in establishing some observation, that observation ought to be directly observed by several trustworthy witnesses. In this manner, consensus was arrived at within a collective. The Royal Society was one such collective, which was simultaneously an institution held in high esteem, and an insider's club distinguishing the witnesses who could be trusted from those that could not. Of course those very functions ensured that the Royal Society was, itself, never free of political, theological, and socio-economic entanglements. Such obligations must attend any institution founded by mere mortals.

Science blossomed. The community became huge. The collectives, each with their own highly refined and impenetrable form of specialist language, multiplied without bound. Experimental observations became increasingly difficult to make, and individual observations typically arose from a long and complex sequence of steps of which no single person could be master. The path from a question about the perception of a bell, to an image obtained in a brain scanner, for example, is tortuous in the extreme, and relies on many different communities of experts, most of who have nothing to do with the question being asked about the perception of the bell. Yet the reputation of science for establishing facts that brook no opposition has only become

stronger. The topics over which the pronouncements of science ranged extend now to every sphere of human activity. With that, the assent that astronomical observations once compelled have become transferred, bit by creeping bit, to observations that are, and must be, far less determinate. The unquestionable power that seems to lurk behind the pull of gravitation, a power supremely indifferent to human concerns, has been allowed to seep into pronouncements in the fields of biology, medicine, psychology, social science, and beyond. All of these extensions demand the recognition of many kinds of subject, but we find only one acknowledged: the individual organism, which in human affairs reduces to the skin clad body of the autonomous adult responsible citizen. What a load we place on its shoulders!

Taking stock of things in 2017 we see a fractured picture. Of course there are philosophers of science, and philosophers of every imaginable facet of human and animate experience, who resist any unthinking fundamentalism with respect to scientific pronouncements. Such critics are sensitive to the complex intertwining of the many strands that knit us together. But that kind of sensitivity is in short supply in an environment in which every second headlines announces "research has shown ..." or "scientists have shown that ..." Not only the public, but the figures of authority who make political, economic, and legal decisions affecting multitudes, are unlikely to be aware of the consequences of allowing the certainty of the pronouncements of astronomy to bleed into assertions about bodies, minds, clubs, nations, and families. But the incentive to argue that one's policy is "evidence based" is irresistible. Objectivity, if not curated with care, becomes an excellent tool with which to insist on political ends, and to silence opposition.

The subject/object divide enshrined in contemporary science, from biology through the psychological and social sciences, cannot be simply overturned, nor history rewound to approach such a distinction differently. The modernity we (some of us) have arrived at insists on a special, revered, role for the individual person, understood as co-extensive with a single skin-clad body, and assumed to be possessed of moral character, free will, and individual accountability. This unit is where the ethical buck stops, and it is the unit based on which economists spin tales of taxation and remuneration. But this monolithic account of the person is a creation, and needs to be seen as such. It cannot be treated as if its properties were comparable to the non-negotiable chemical constituents of the sun which simply have to be observed. This specific approach to the subject has its entailments, and they are neither universally shared, nor shareable. Furthermore, their limitations land us with specific kinds of problems that necessarily accompany such a narrow view.

The entire domain of mental health represents one area of acute human concern, that we acquire and shape when we apportion responsibility in this way and in this way only. The field of mental health has long grappled with the very obvious relational nature of many of its pathologies, where the problems and suffering encountered are clearly seen to lie in the web of relations an individual is enmeshed in. Its explanatory tools and conceptual stock are provided by this strongly individualist view of the person, with the result that the distinctions and categories of the field as embodied, e.g. in the Diagnostic and Statistical Manual, are constantly shifting, dissolving, splitting and aggregating, without any stable foundation. The field is riven with competing approaches that exhibit inconsistent stances with respect to biology and to interpersonal communication, as when psychopharmaceutical therapies compete with talk therapies. To be very clear: The problem is not that this modern view of the subject as monolithic, autonomous and self-contained should be held responsible for the suffering found in the domain of mental health. The issue is that any and every approach to the person will come with its limitations, its inadequacies, and its difficulties. Intolerable behavior and intolerable suffering arise in every society. The form they take, and the capacities ranged against them differ from here to there. In most parts of the world, perhaps all, prison systems and mental health institutions act as complementary institutions that work together, mopping up those cases that the vast majority cannot bear to live with or among.

The desire of Newton to uncover the "universal qualities of all bodies whatsoever" requires a health warning. The bodies he spoke of were inert material entities. They were not living beings, and most certainly not people. But the spirit of universalization extends far beyond the science of mechanics. It is a prominent feature of many varieties of Christianity and liberal humanism that accounts of something called "humanity" are sought in the hope of saving, helping, or more prosaically, understanding all those with whom one feels a commonality. This understandable sense of solidarity elevates the notion of the species, and homo sapiens in particular, to a very problematic throne. Which sciences can inform us about the members of that set that, for many, seems to define "us"? Here, I think, the challenges raised by the thematization of joint speech as an object of study may be of some assistance, not only to instill an appropriate kind of caution, but to also provide us with means for calibrating and enriching the generalizations we can make. Not that studying joint speech itself will provide the answers to such vexed issues, but that the strongly empirical study of this behavior and the suites of rich practices in which it is found might sensitize us to the need to recognize our pluralities, our fundamental differences, and perhaps

immunize us just a little bit from the seductions of over-generalization and the leap to pronouncements about a mythical being called "humanity."

## Science and the messiness of history

When we considered the explanatory load placed upon the brain, we considered two different historical trajectories. The first started with the rational Christian soul, which is not a construct that many scientists today would choose to buy stock in. The soul is the heart of Descartes' cogito, and every time we, as n-th generation post-Cartesian sophisticates, confront the problem of an enduring sense of self, our thoughts are still channeled within walls constructed from modern Christian theology, the Enlightenment, and Western modernity. This story tries to understand "Man," and seeks to separate his qualities from those of the beast and other life forms. Old fashioned nonsense, yes, but still clearly discernable in all the sciences of the human and the animate.

The other tack to take is strongly biological, and ends with Homo sapiens, not with Man. It is a historical account, no longer traceable in detail, in which contingency and accident play starring roles. On this account there is continuity with the rest of life, and we seek to understand a particular form of primate as best we can. This account cannot lean on a rational soul, or use rationality as a chasm between the brute and us.

Evolution is a queer process. Typically, students of the development of this species or that seek to understand the lineage of one species, or group of species, at a time. So we wonder about the relationship of birds to dinosaurs, or of *Homo sapiens* to *Homo heidelbergensis*. Like any historical narrative, the tracing of any such path selectively foregrounds some details and excludes others. One body part is understood to morph into another in response to some selective advantage or to selection pressure. From each snapshot view of a population, some will produce offspring who will produce more offspring, all subtly different, while others will end their lineage there and then. What such an account misses, indeed must miss, is the reciprocity built into the goings on within the biosphere. Each lineage evolves in conjunction with each other, and the whole evolving thing simultaneously gives rise to the "environment" to which, on the classical narrative, any given species reacts. The view of the biosphere regarded as an integral whole has been influentially presented by James Lovelock in his Gaia hypothesis (Lovelock and Margulis, 1974), now finding a new audience in those who consider

in depth the ongoing relation between the activities of some humans and the planet as a whole (Crutzen, 2006).

In the midst of this continuous activity in which life interacts with life bringing forth more life, patterns form. Some of these we distinguish as species, and we find this to be a useful way of demarcating one kind of entity from another. But species boundaries are less than fixed. Geographical distribution means that some species blend seamlessly into others without clear dividing lines. Reproductive criteria do not suffice to make species boundaries completely clear, as the mules, ligers and odd mixtures we find testify. Furthermore, for many species, the boundary we have just drawn seems to be arbitrary, as when we look at the symbiotic relationship of an alga and a fungus in a lichen, where neither "species" has any kind of independent existence, or we try to decide whether a siphonophore is a single animal or a colony. Evidence of the fusing and annihilation of species-like distinctions are provided by the mixture of DNA found in every eukaryotic cell, where some time in the distant past two organisms forged such a close symbiotic relationship that they became one, giving us the nucleated cell.

In order to conduct scientific activity of relevance to the one species we treasure so much, normative concerns must arise. We must take sides, and recognize that some processes, events, happenings that are not intrinsically good or bad in themselves (such as the extinction of all mammalian life, for example) are indeed very bad from the perspective we have adopted. And where normative concerns arise, we are necessarily faced with the treatment of subjects, for things only matter to subjects. But in restricting ourselves to a single kind of subject, the individuated person, we risk failing to recognize the many kinds of collectivities, and with that the many kinds of collective perspectives, that we co-constitute. Of these, the species collective is one, but there are very many others, and there are collectives we co-constitute that cut across these boundaries completely. The relationship of a shepherd to her dogs and sheep, for example, brings into being a specific kind of collective that has a specific perspective on things, that enacts its own identity, and for whom, as a collective, things matter.

So the problem of very many kinds of subjects, of a biosphere which is nothing other than the reciprocal interactions among very many subjects, each with their own perspective, making meaning in a manner appropriate to their own constitution, their capacities for making distinctions, and their capacities to act on their environment, goes very deep indeed. The simple schema provided by the notion of participatory sense-making serves to illustrate how the sense-making activities of many kinds of subjects, many of which are nested hierarchically within others, must give rise to a domain

of life in which no single subject can be extracted intact from its milieu. No subject can properly be regarded as separable. Subjects have autonomy, yes, but it is a specific kind of autonomy that comes from regulating the interaction with the environment of relevance. The individual liver cell is a subject with a milieu within the liver (though interpenetrated by fluids, chemicals, from the entire body). The liver itself is a subject within the body. The body, considered as a fleshy organism, is a subject within a specific kind of environment. For the body as organism, the quality of the air it breathes will be an important part of the milieu. But if we choose to view John Doe, not as a repsirator, but as a member of a lending library, we are speaking of a different kind of subject, with a very different kind of milieu, that includes such entities and relations as books, loans, and librarians.

The development over time of the biosphere is a historical process, arising from the reciprocal interactions of many kinds of subjects. This evolutionary churn of life gives rise to patterns, to forms and processes of relative stability, which we can discern. But just as the edges of a species are somewhat indeterminate, so the bounds of any pattern we recognize among the forms of life will quite probably be somewhat blurry. We find very many instances of regularity among the living, but regularity within biology is strictly not comparable to the kind of universality we find in the mechanical view of the world. In dealing with the living, we are dealing with the products of contingent historical processes, and thus regularity, not absolute determinism, will be the norm. The ambitions of the structuralists of the 20th century were grounded in the belief that we might identify self-contained domains that could be studied as if they were the products of natural law, rather than the result of contingent historical processes, which are, of course, much messier.

In the study of language, this gave rise to the notion that ultimately became the theory of universal grammar. From the kind of atomism introduced by structural linguistics, recognizing abstract building blocks such as phonemes and words, modern linguistics made ambitious claims that there were universal features of all human language that must be innate, because the environment of the child could never provide sufficient high quality information to allow language learning from scratch. We have already reviewed one consequence of this, which is the neglect of almost all interesting features of vocal interaction. Another unfortunate consequence is the widespread acceptance of the universalist claims of such a partition of the vocal sounds we employ into linguistic and non-linguistic features, with the fond hope that the universals of "language" might be systematic, sharing in the absolutist nature of gravitation rather than the messy contingency of biology. As the

study of human language has extended beyond its home base in the European sphere, the idea of any universals, even in the limited sense required for structuralist and generativist ambitions, has waned (Evans and Levinson, 2009). To the Summer Institute of Linguistics, it is very convenient to act as if "language" were a single thing whose joints and parts we might anticipate in kind if not in detail before we ever speak to a person from a culture we do not share. In this manner, the universal message of the Gospels can be straightforwardly dressed in different words and made available to others, whose personhood is arrogantly assumed to share in the single stamp of human nature, or Mankind.

Exactly the same kind of inappropriate generalization underlies the structuralist foundations of Anthropology of the mid 20th Century. Just as a belief in a single kind of human nature (that of Man) justified the expectation that all people everywhere would have a language made up of the same kinds of parts, so structural anthropologists conjectured that a universal human nature would result in specific kinds of social relations being manifest, so that moving from one culture to another would involve difference in detail, but not in essence. Specific sets of relations that characterized the family or clan would thus differ across contexts, but knowledge of many such sets of relations would equip one with the kind of knowledge necessary to interpret an entirely novel people. Human relations, like human language, were assumed to have a kind of essence that the anthropologist could uncover.

Such presumptions have fallen out of favor in anthropology. They seem to be on the way out in linguistics, at least in many circles, though the battles are still, it seems, being fought. But the universalizing tendency exhibited in both fields, and so at home in the European Modern worldview persists. Both structuralist programs seem to me to be founded on the thoroughly mistaken notion that a historical sequence admits of a determinate rendering, and can give rise to systematic regularities that can, in turn, be given a "factual" description on par with a robustly empirical account of the motion of the planets. This is not credible.

## The ground from which we speak

Joint speech has the potential to draw our attention to the many ways in which identities are enacted. It also provides a necessary caution against thinking of identity as monolithic, individual, enduring, essential, and entirely personal. In a pluralistic world in which people of many different traditions and origins mingle freely, trade, and live together, there is an

urgent need to move beyond a crude tribal framework in which individual people are "with us or against us," as an unfortunate political phrase has it. By bringing to our attention the many ways in which many kinds of allegiances are enacted, some transiently and some in very much more enduring form, it may be helpful in developing a consensus-based science that can be employed in the conduct of human affairs, without abusing the implicit authority of the pronouncements of science.

If we are able to resist a simple positivism, in the Parmenidean mode, the change to a presentist, Heraclitean perspective will bring some themes to the fore that have been rather neglected. The grand theme of representation, which addresses mediation in the relation of subject to world, underlies the venerable idea of communication as message passing, and the distinctness of subjects from their worlds. This is an entirely orthodox way of construing people and their interactions, and it is irrevocably Parmenidean in nature, with all the strengths and weakness that entails. In place of that whole edifice, a Heraclitean approach will bring to the fore questions of liveness and co-presence, drawing our attention to the manner in which real time reciprocal interaction among agents of various forms gives rise to collectivities, to enacted unities which arise under specific conditions, are maintained through regulated exchanges with their environment of relevance, and which will necessarily dissipate under specific conditions. With that, it might be possible to recognize the co-existence of incompatible value systems, and to appreciate when value systems are in conflict or collision.

There is a stark contrast between these two ways of framing ones observations and of moving towards statements that can garner consent. Neither approach will suffice. Perhaps we need different words for them, and should not lump both under a single notion of "science." If we have reached sufficient agreement, cleared enough common ground, that we can rely on the same set of entities we treat as simply existent, then we can reach very secure agreement. This is what Feynman means by answering a "why" question. It means we have explanations we can use. This will always be necessary if we are to use the results of our inquiry in a robust fashion, to inform laws or to contribute to the negotiation of various forms of authority. But it demands work in ensuring that we have that common ground, so that we can rely on the same set of assumptions. It means those truths apply *sensu stricto* only to those who share a specific kind of foundation.

Normativity and science are stuck with each other, as long as we wish to apply the scientific insights to ourselves and our world. In its brief history, science has never managed to extract itself from its genesis in a specific cultural, theological and political context. I see no reason to believe that it

will do so in the future either. But the consequence of this is simply that the written pronouncements of science, the journal articles and monographs, should not be regarded as scripture, set in stone, and established for all time. This would be, of course, a crude fundamentalism of a kind we have seen repeated in many religious and political contexts.

There is a further lesson to be learned from the study of joint speech, and the manner in which it reveals the centrality of uttering, rather than writing, in the interactions of language users. If journal articles are not scriptures, etching into the volumes an eternal and unmovable truth and thereby establishing an uncontestable view-from-nowhere, then might we also recognize that each scientist is a person grounded in many traditions and currents, and that the pronouncements of science can never amount to a voice-from-no-one either. The lamentable tradition of over-generalization might be improved upon. That would mean introducing something of the personal and contingent into scientific communication. It would mean being wary of premature generalization and developing a sensitivity to the political, social, and ethical consequences of any claims of fact. It would push back against a long tradition of rationality that believes that the faculty of reason necessarily arrives at eternal truths. It would suggest that science might see itself as constantly in need of renewal–not to reestablish the constant of gravitation, but to check the domains over which its many truths range.

# Bibliography


Austin, J. L. (1962). *How to Do Things with Words*. Harvard University Press.

Bakhtin, M. (1987). The problem of speech genres. In *Speech Genres and Other Late Essays*, pages 60–102. University of Texas Press.

Bennett, M., Schatz, M. F., Rockwood, H., and Wiesenfeld, K. (2002). Huygen's clocks. *Proceedings of the Royal Society A: Mathematical Physical and Engineering Sciences*, 458(2019):563–579.

Berg, J. (2016). Gene-environment interplay. *Science*, 354(6308):15.

Bernstein, N. A. (1930). A new method of mirror cyclographie and its application towards the study of labor movements during work on a workbench. *Hygiene, Safety and Pathology of Labor*, 5:3–9.

Bridges, R. T. (1933). *Yamana–English: A Dictionary of the Speech of Tierra del Fuego*. Missionsdruckerei, St. Gabriel, Mödling, Austria. Edited by F. Hestermann and M. Gusinde.

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., and Moore, C. (1998). *Social cognition, joint attention, and communicative competence from 9 to 15 months of age*. Monographs of the Society for Research in Child Development. JSTOR.

Carrier, D. (1986). Cogitamus ergo sumus. *The Monist*, 69(4):521–533.

Chemero, A. (2011). *Radical Embodied Cognitive Science*. MIT press.

Connor, S. (2000). *Dumbstruck: A Cultural History of Ventriloquism*. Oxford University Press.

Couper-Kuhlen, E. (1993). *English Speech Rhythm: Form and Function in Everyday Verbal Interaction*, volume 25. John Benjamins Publishing.

Cowley, S. J. (1998). Of timing, turn-taking, and conversations. *Journal of Psycholinguistic Research*, 27(5):541–571.

Crick, F. (1995). *The Astonishing Hypothesis: The Scientific Search for the Soul*. Simon and Schuster.

Crutzen, P. J. (2006). The "Anthropocene". In *Earth System Science in the Anthropocene*, pages 13–18. Springer.

Cummins, F. (2003a). Practice and performance in speech produced synchronously. *Journal of Phonetics*, 31(2):139–148.

Cummins, F. (2003b). Rhythmic grouping in word lists: competing roles of syllables, words and stress feet. In *Proceedings of the 15th International Conference on Spoken Language Processing*, pages 325–328.

Cummins, F. (2004). Synchronization among speakers reduces macroscopic temporal variability. In *Proceedings of the 26th Annual Meeting of the Cognitive Science Society*, pages 304–309.

Cummins, F. (2009). Rhythm as entrainment: The case of synchronous speech. *Journal of Phonetics*, 37(1):16–28.

Cummins, F. (2011). Periodic and aperiodic synchronization in skilled action. *Frontiers in Human Neuroscience*, 5. Published online, December 31, 2011.

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. (2014a). Agency is distinct from autonomy. *Avant: Trends in Interdisciplinary Studies*, 5(2):98–112.

Cummins, F. (2014b). The remarkable unremarkableness of joint speech. In *Proceedings of the 10th International Seminar on Speech Production*, pages 73–77.

Cummins, F. (2014c). Voice, (Inter-)Subjectivity, and real time recurrent interaction. *Frontiers in Psychology*, 5(760).

Cummins, F. (2015). Rhythm and speech. *The Handbook of Speech Production*, pages 158–177.

Cummins, F. and De Jesus, P. (2016). The loneliness of the enactive cell: Towards a bio-enactive framework. *Adaptive Behavior*, 24(3):149–159.

Darwin, C. (1909). *The Voyage of the Beagle*. Collier.

Daston, L. and Galison, P. (1992). The image of objectivity. *Representations*, pages 81–128.

De Jaegher, H. and Di Paolo, E. (2007). Participatory sense-making. *Phenomenology and the Cognitive Sciences*, 6(4):485–507.

Dennett, D. C. (1993). Review of F. Varela, E. Thompson and E. Rosch, *The Embodied Mind. American Journal of Psychology*, 106:121–126.

Deutsch, D., Henthorn, T., and Lapidis, R. (2011). Illusory transformation from speech to song. *The Journal of the Acoustical Society of America*, 129(4):2245–2252.

Dewey, J. (1896). The reflex arc concept in psychology. *Psychological Review*, 3(4):357–370.

Evans, N. and Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, 32(05):429–448.

Eysenck, M. W. and Keane, M. T. (2000). *Cognitive Psychology: A Student's Handbook*. Taylor & Francis.

Froese, T. and Di Paolo, E. (2011). The enactive approach: Theoretical sketches from cell to society. *Pragmatics & Cognition*, 19(1):1–36.

Galison, P. (2004). *Einstein's Clocks and Poincaré's Maps: Empires of Time*. WW Norton & Company.

Geertz, C. (1973). *The Interpretation of Cultures: Selected Essays*. Basic Books.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Houghton, Mifflin and Company.

Gilbert, S. F., Sapp, J., and Tauber, A. I. (2012). A symbiotic view of life: We have never been individuals. *The Quarterly Review of Biology*, 87(4):325–341.

Glyn, I. M. and Brelstaf, G. (1994). Reviews: The astonishing hypothesis: The scientific search for the soul, computational models of visual processing. *Perception*, 23:367–370.

Griffin, A. (2016). Mysterious chimpanzee behaviour could be 'sacred rituals' and show that chimps believe in god. *The Independent*. Retrievable online at www.independent.co.uk.

Harnsberger, J. D., Wright, R., and Pisoni, D. B. (2008). A new method for eliciting three speaking styles in the laboratory. *Speech Communication*, 50(4):323–336.

Held, R. and Hein, A. (1963). Movement-produced stimulation in the development of visually guided behavior. *Journal of Comparative and Physiological Psychology*, 56(5):872–876.

Hubel, D. H. and Wiesel, T. N. (1979). Brain mechanisms of vision. *Scientific American*, 241(3):150–162.

Huebner, B., Bruno, M., and Sarkissian, H. (2010). What does the nation of China think about phenomenal states? *Review of Philosophy and Psychology*, 1(2):225–243.

Jasmin, K. M., McGettigan, C., Agnew, Z. K., Lavan, N., Josephs, O., Cummins, F., and Scott, S. K. (2016). Cohesion and joint speech: Right hemisphere contributions to synchronized vocal production. *The Journal of Neuroscience*, 36(17):4669–4680.

Keijzer, F., Van Duijn, M., and Lyon, P. (2013). What nervous systems do: early evolution, input-output, and the skin brain thesis. *Adaptive Behavior*, 21(2):67–85.

Kelso, J. A. S. (1995). *Dynamic Patterns: The Self-Organization of Brain and Behavior*. MIT press.

Latash, M. L. (2008). *Synergy*. Oxford University Press.

Latour, B. (1999). *Pandora's Hope: Essays on the Reality of Science Studies*. Harvard University Press.

Latour, B. (2005). What is given in experience? a review of Isabelle Stengers' "Penser avec Whitehead". *Boundary 2*, 32(1):222–237.

Livingstone, M. S. and Hubel, D. H. (1981). Effects of sleep and arousal on the processing of visual information in the cat. *Nature*, 291(5816):554–561.

Lovelock, J. E. and Margulis, L. (1974). Atmospheric homeostasis by and for the biosphere: the Gaia hypothesis. *Tellus*, 26(1-2):2–10.

Marr, D., Poggio, T., Hildreth, E. C., and Grimson, W. E. L. (1991). A computational theory of human stereo vision. In *From the Retina to the Neocortex*, pages 263–295. Springer.

Maturana, H. and Varela, F. (1987). *The Tree of Knowledge: The Biological Roots of Human Understanding.* New Science Library/Shambhala Publications.

Moghith, K. (2014). *Hitafat Al-Thawra Al-Masriya (Chants of the Egyptian Revolution).* Supreme Council for Culture (Egypt).

Moore, R. K. (2016). Introducing a pictographic language for envisioning a rich variety of enactive systems with different degrees of complexity. *International Journal of Advanced Robotic Systems*, 13(74).

Nagel, T. (1989). *The View From Nowhere.* Oxford University Press.

Nichol, L. (2005). *The Essential David Bohm.* Routledge.

Okamoto-Barth, S., Call, J., and Tomasello, M. (2007). Great apes' understanding of other individuals' line of sight. *Psychological Science*, 18(5):462–468.

Olson, R. (2004). *Science and religion, 1450-1900: from Copernicus to Darwin.* Greenwood Publishing Group.

Ong, W. J. (1977). *Interfaces of the Word: Studies in the Evolution of Consciousness and Culture.* Cornell University Press.

O'Regan, J. K. and Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24(05):939–973.

Pettit, P. (2004). Groups with minds of their own. In *Socializing Metaphysics*, pages 167–193. Rowman and Littlefield.

Port, R. F. (2007). The graphical basis of phones and phonemes. *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, pages 349–365.

Rall, J. E. (1996). Review: The astonishing hypothesis. *The Journal of Nervous and Mental Disease*, 184(6):384–385.

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

Reed, E. S. (1997). *From Soul to Mind: The Emergence of Psychology from Erasmus Darwin to William James*. Yale University Press.

Rorty, R. (1979). *Philosophy and the Mirror of Nature*. Princeton University Press.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., and Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, 36(04):393–414.

Shapin, S. and Schaffer, S. (1985). *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life*. Princeton University Press.

Shapiro, L. (2010). *Embodied Cognition*. Routledge.

Sudnow, D. (1978). *Ways of the Hand: The Organization of Improvised Conduct*. MIT Press.

Thompson, E. and Cosmelli, D. (2011). Brain in a vat or body in a world?: Brainbound versus enactive views of experience. *Philosophical Topics*, 39(1):163–180.

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.

Vanrullen, R. and Thorpe, S. J. (2001). The time course of visual processing: from early perception to decision-making. *Journal of Cognitive Neuroscience*, 13(4):454–461.

Varela, F. J., Rosch, E., and Thompson, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. MIT press.

Vicsek, T. and Zafeiris, A. (2012). Collective motion. *Physics Reports*, 517(3):71–140.

Von Foerster, H. (2003). Responsibilities of competence. In *Understanding Understanding*, pages 191–197. Springer.

Von Zimmerman, J. and Richardson, D. C. (2015). Verbal synchrony in large groups. In *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*, pages 2523–2528, Austin, TX. Cognitive Science Society.

Wagner, P., Malisz, Z., and Kopp, S. (2014). Gesture and speech in interaction: An overview. *Speech Communication*, 57:209–232.

Warren, W. H. (2006). The dynamics of perception and action. *Psychological Review*, 113(2):358–389.

Weber, A. and Varela, F. J. (2002). Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the Cognitive Sciences*, 1(2):97–125.

Wei, W. W. (2002). *Ask the Awakened: The Negative Way*. Sentient Publications.

Wennerstrom, A. (2001). *The Music of Everyday Speech: Prosody and Discourse Analysis*. Oxford University Press.

Whitehead, A. N. (1920). *The Concept of Nature*. Cambridge University Press.