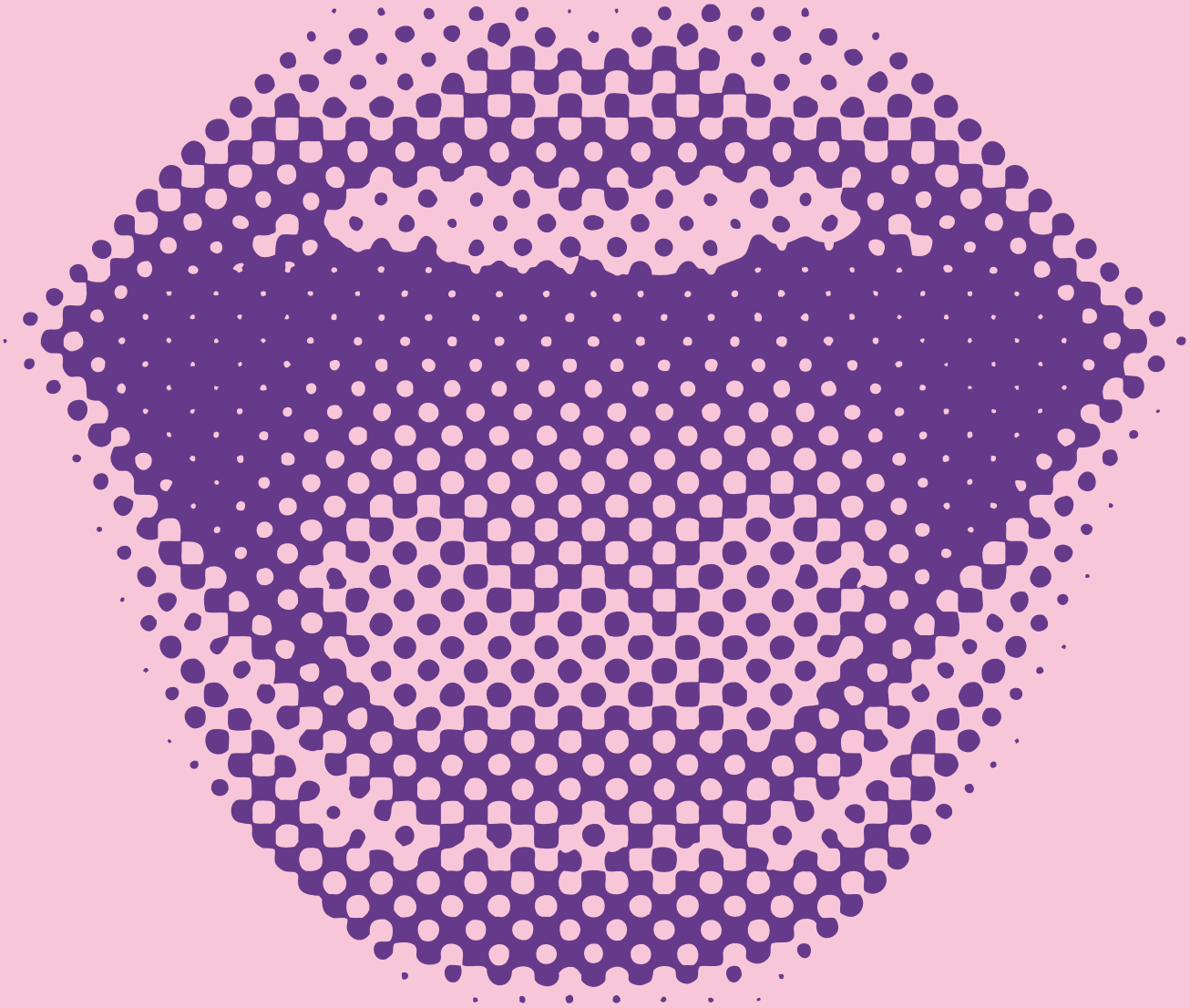
Can a game of 'vocal' charades act out the origin of language?



Plato as depicted in Raphael's masterpiece The School of Athens (1511). Plato contemplated the origins of words. In the painting, he is shown gesturing, which may reveal a clue to the puzzle.

Marcus Perlman explores whether a game of 'vocal' charades can shed light on how our ancestors created the first words.

Sometime during the course of human evolution, a group of our ancestors created the first language, with a grammar and a repertoire of words. Scientists don't know very precisely when this happened – different theories posit that it could have been tens of thousands of years ago, or even as far back as a million years ago or more. One's estimate, of course, hangs on one's definition of terms like 'language', 'grammar' and 'word'. But whenever it was, we know that at some point in our history, humans must have created new, meaningful elements from scratch – *de novo*. How did we do this? How did our ancestors create the first words?

This is a question that cuts to the core of language, and it is one that scholars have pondered for thousands of years. Almost 2,500 years ago, the Greek philosopher Plato contemplated the origins of language in his dialogue The Cratylus. In this work, the characters Cratylus and Hermogenes engage Socrates in a discussion on the essence of words. They argue back and forth whether names for things are natural, with an intrinsic "truth or correctness in them", or whether they are conventional – an arbitrary agreement between a particular group of speakers.

In much of modern linguistics, scholars have – for most intents and purposes – taken the answer to Plato's question to be 'conventional'. De Saussure's fundamental axiom of the arbitrariness of the sign has ruled the day. Consider, for example, the observations of Charles Hockett in his seminal 1960 article on the origin of speech. "In language the ties are arbitrary," he writes. "The word 'salt' is not salty nor granular; 'dog' is not 'canine'; 'whale' is a small word for a large object; 'microorganism' is the reverse." In a later article, Hockett explains why this is so: "When a representation of some four-dimensional hunk of life has to be compressed into the single dimension of speech, most iconicity is necessarily squeezed out. In one-dimensional projection, an elephant is indistinguishable from a woodshed." Thus, words are arbitrary "perforce", because the medium of speech has little potential for iconicity. That is to say, speech sounds cannot resemble or depict their meanings, and therefore, words are doomed to be arbitrary. (There is, at least, one apparent exception to the arbitrariness of spoken words – onomatopoeia. These are words that seem to imitate sounds, like 'meow' and 'bang'. These words, however, are often marginalised as comprising just a trivial fraction of the vocabulary. Onomatopoeia "only applies to noisy things," psychologist Steven Pinker explains.)

This dominant theory of linguistics has presented a

challenge to scientists trying to understand how the first languages might have got started. If words are arbitrary, then their origins are impenetrable. There is just the current vocabulary of arbitrary words, preceded by older arbitrary words, and so on. But we cannot push the problem back forever – how did words originally come to be associated with their particular meanings? Did a community of ancient humans agree one day to create a random sound, assign it a meaning, and *poof!* the first word was born? The prospect of such a process – in which the first words were deliberately invented – appears dubious. But then what could be the alternative?

Over the centuries, a number of influential scholars have argued that a more plausible scenario stems from a gesture-first hypothesis of language origins. According to this idea, the first words were *gestures*, not vocalisations.

Were the first words actually gestures?

The benefit of gesture is obvious if you have ever travelled to a place where you didn't speak the local language. In these circumstances, you probably found yourself – and your interlocutor – doing a lot of gesturing to communicate with each other. This is because gestures very naturally look like what they mean. Unlike speech – at least the way Hockett describes it – gestures *do* have a lot of potential for iconicity. People readily use gestures to trace shapes and pathways (e.g. giving directions), show spatial arrangements (depicting the layout of a room), point to things ('I'll take *that* one') and pantomime actions (scribbling your signature for 'check please'). By connecting form with meaning, iconic gestures enable two people to understand each other, even when they lack a common vocabulary. This allows them to establish a meaningful convention that is shared between them. Thus, iconic gestures provide a means for people to create new words – or *signs,* as the case may be.

In fact, the idea that people can use iconic gestures to build the vocabulary of a new language is not just speculation. Although spoken languages are extremely ancient, new signed languages have occasionally sprung up in deaf communities around the globe. These cases have provided scientists with the unique opportunity to observe the birth of a language, including the creation of a vocabulary where there was none before. In these cases, researchers have observed that people do indeed make use of iconicity to build a lexicon of signs. Over time, iconic gestures become increasingly conventionalised – especially as they are used over multiple generations – and they eventually develop into the standardised signs (and grammar) of an emerging sign language.

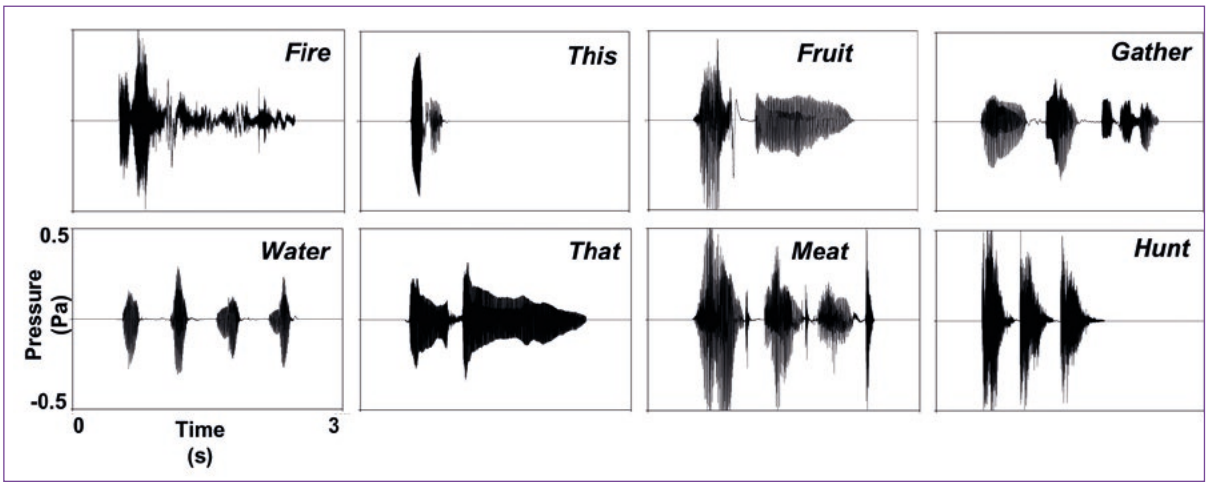
But when it comes to spoken languages, scientists have been left mainly to conjecture. One common line of reasoning is illustrated in a thought experiment proposed by psychologist Michael Tomasello. Tomasello invites the reader to imagine two groups of young children, each growing up on an isolated desert island. The children are entirely well cared for (somehow), but there is no one else on the islands – and no language. The key methodological twist is that on one island, the children are limited to communicating just with gestures, while on the other island, they must use only vocalisations. (For our purposes, we can blithely ignore the details of implementing this manipulation.) As Tomasello notes, we have a solid empirical basis to predict what would happen in the case of the gesturing children – they would, over a few generations, use iconic gestures and pointing to build a sign language.

But what about the children who must rely on vocalisations? Would *they* be able to create a language? According to Tomasello, it is difficult to see them doing this. In particular, he suggests that the children would have trouble "inventing on their own vocalisations to refer the attention or the imagination of others to the world in meaningful ways - beyond perhaps a few vocalisations tied to emotional situations and/or a few instances of vocal mimicry." Therefore, they would not be able to create "already meaningful communicative acts" that could "serve as starting points" for new words.

It goes without saying that such an experiment could not be conducted in real life. Nevertheless, as a thought exercise, it raises a provocative claim about the way languages are created – one that warrants empirical evaluation. Is Tomasello's intuition correct? Are vocalisations really so limited in iconicity as he – and Hockett – would have us believe?

A game of 'vocal' charades

When I was a graduate student, I had the idea that it would be fun to test Tomasello's hypothesis with a game of 'vocal' charades. I wanted to see just how effectively people could communicate using only the sound of their voice, when they were prohibited from using words and gestures. In



Waveforms of eight vocalisations from the winning submission of the Vocal Iconicity Challenge. For each vocalisation, the x-axis shows the duration of the vocalisation, from 0 to 3 seconds. The y-axis shows sound pressure (loudness) in Pascals. For example, you can see that the vocalisation for *this* was shorter and louder than *that*, and that *water* was represented by repeated dripping sounds, and *fire* by a crackling, more continuous sort of noise.

particular, I wondered about their ability to communicate beyond the domains of emotion and sound. While the need to use modern English speakers as participants was clearly a limit to my investigation of language origins, the method offered, at least, an ethical semblance of Tomasello's desert-island thought experiment. If my participants turned out to be really bad at vocal charades, then probably Tomasello and Hockett were right. However, I had a hunch that people might prove to be better than these prior accounts would suggest.

Beginning with that initial experiment, I have – over the years, and with the help of my collaborators – run a series of studies asking people to use their voice to communicate a diverse array of concepts. Participants typically perform the task with a partner, taking turns vocalising and guessing meanings that span conceptual domains like speed (e.g. *fast, slow),* texture (*rough, smooth*), luminance (*bright, dark*), temperature (*cold, hot*), actions (*cook, hide*), space (*here, far*), and naturally occurring things (*water, rock*) – to name a few. Afterwards we have used Internet crowd-sourcing platforms like Amazon Mechanical Turk to test whether recordings of the vocalisations could be understood by new listeners, who had not been part of the game. These additional experiments ensure that guessers cannot take advantage of visual information like facial expressions and inadvertent gestures, or other cues arising from the shared context of playing the game together.

So how do people do? Are they any good at vocal charades? After a *lot* of time spent analysing the many delightful sounds people make – as well as testing the ability of other people to make any sense of them – I think there are two principal findings to draw from this work. First, more often than not, people share similar ideas about how to use their voice to represent a particular meaning. Ask different people to communicate *bright*, and most will produce a melodic, high-pitched sound. For *rough*, they will make a noisy sound that is lower in pitch. *Near* is a Doppler effect coming towards you; *far* is the effect going away. *This* is a shorter, louder sound than *that*. As these examples reflect, participants utilise the full expressive potential of their voice when they create iconic vocalisations, manipulating properties like their duration, pitch, loudness, and timbre to convey their intended meaning. Thus – far from being a single dimension – participants exhibit how their voice can serve as a rich, multi-dimensional source for iconic representation.

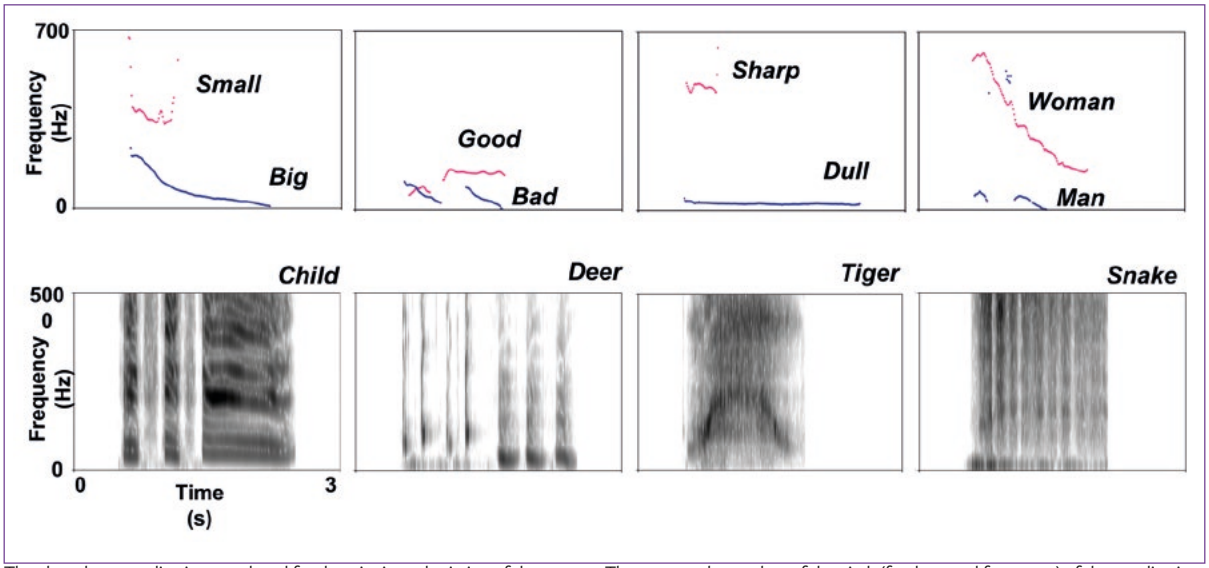
The second finding of these studies is that people are, for the most part, pretty good at guessing the meanings of the vocalisations. In one experiment, for example, participants listened to vocalisations representing eighteen different meanings and selected their responses from ten choices. Compared to the chance guessing rate of ten per cent, participants on average selected the correct answer more than a third of the time. Indeed, for the great majority of meanings we have tested, people are able to select the correct response at a rate that is significantly higher than chance. Not surprisingly, accuracy varies a lot between different meanings: *sleep* and *attractive* are both fairly easy to guess, whereas *this* and *near* are much more difficult. Notably, however, when people do make mistakes, we find that they tend to confuse similar meanings – *bad* is confused with *rough* and *ugly*, and *short* with *fast* and *small*, for instance.

Most recently, we completed what may stand as the culmination of the vocal charades experiments. Working with psychologist Gary Lupyan, we asked people to put their vocal skills to the test in a high-stakes contest we called The Vocal Iconicity Challenge. Contestants were challenged to record a set of vocalisations to communicate thirty different meanings that were selected for their possible relevance to the lives of our ancestors – for example, *fire, hide, tiger, sharp, big* and *pound*. We then played the vocalisations from each submission to ten different listeners who attempted to guess their meanings. The contestant whose vocalisations were guessed most accurately was crowned the Vocal Iconicity Champion and awarded the $1,000 Saussure Prize. We were pleased that our contest drew several submissions from researchers associated with top linguistics, psychology, and language evolution programmes from across the USA, as well as from the UK and Europe. With such a prestigious award at stake, we found that contestants were up to the challenge. For the best submissions, average guessing accuracy across all thirty meanings was higher than 50%, compared to the chance rate of 10%. Accuracy was 57% for the championship submission.

Can vocal iconicity bridge disparate languages and cultures?

Our results suggest that – as I had suspected – people are actually pretty good at playing vocal charades. Participants in these studies certainly did not appear so severely constrained – limited to just emotional expression and vocal mimicry – as Tomasello argued in his thought experiment. The vocalisations that participants produce are rich and multi-dimensional, and listeners – based only on the sound – are able to guess their meanings with considerable accuracy. But before leaping to any conclusions about language origins, there is an important limitation to these studies. Namely, all of the participants were English speakers, and the vast majority were Americans. Of course, they were always instructed not to use words, including interjections like 'boo' or 'mm-hmm'. But still, it is likely they were able to draw on other pieces of shared cultural knowledge related to sound and meaning. As one example, consider the vocalisations that participants tend to make for *up* and *down*, producing sounds with rising and falling pitch, respectively. This mapping between space and pitch is second nature to English speakers, but it is far from universal. In Farsi, for instance, low pitches are 'thick' and high pitches 'thin'.

Thus, a stronger test of whether our ancestors could have used iconic vocalisations to create the first words is to see how good people are at vocal charades when they have to communicate across disparate linguistic and cultural backgrounds. Towards this goal – in work with Gary Lupyan and linguist Jing Paul – we have taken our vocal charades experiments to China, where we have also extended our study to examine the skills of children. Notably,



The plots show vocalisations produced for the winning submission of the contest. The top row shows plots of the pitch (fundamental frequency) of the vocalisation. Each of the four plots compares the pitch of two antonymic meanings. For example, the concept of *small* was represented by a shorter, higher-pitched sound than *big*, whereas *woman* was expressed by a longer, higher-pitched sound than *man*. The bottom row shows spectrograms for four of the vocalisations related to distinguishing different kinds of animals. These show, for instance, that *tiger* was represented by a roaring sound, *deer* by the imitation of clacking hooves followed by some soft animal vocalisations, and *snake* by a noisy sibilant sound.
The bottom row shows spectrograms for four of the vocalisations related to distinguishing different kinds of animals.

"These results point to quite an impressive feat of communication, when you think about it. Children living in China – even those with minimal experience hearing the world – are able to create vocalisations that can effectively communicate notions like long, small, few, *and* a lot *to American listeners in an entirely different time and culture."*

our participants have included both children with normal hearing, and – to explore the boundaries of communication – children with congenital deafness. In a preliminary study focused on the communication of magnitude, we asked the children to create vocalisations that distinguished between paired items – a *big* versus *small* ball, a *long* versus *short* string, or *several* versus *a few* marbles, for instance. We then played their vocalisations back to adult listeners – Chinese Mandarin speakers and American English speakers – and tested their ability to guess which item the child was referring to. If American listeners in particular are able to understand the children's vocalisations, especially those of the deaf children, then this would indicate that vocalisation can be a robust means to communicate across linguistic and cultural boundaries – at least when it comes to magnitude.

The results show that the hearing and the deaf children both created vocalisations that are understandable to Chinese and American listeners alike. Listeners' accuracy was far from perfect, but statistically speaking, their guessing was very reliably above chance. Intriguingly, American listeners appear to be about as accurate as Chinese listeners, suggesting that the iconicity of the vocalisations may be universally recognisable. These results point to quite an impressive feat of communication, when you think about it. Children living in China – even those with minimal experience of hearing the world – are able to create vocalisations that can effectively communicate notions like *long, small, few,* and *a lot* to American listeners in an entirely different time and culture.

Masters of charades

We will never know for certain how our ancestors created the first words, and Plato's ambivalence is likely to continue on through the millennia. The findings from a game of vocal charades are obviously just a small piece in the complex puzzle of language origins. Nevertheless, I think these studies highlight something important about human communication and how it might have evolved. They show that whatever the medium, people are masters at the game of charades. Left without words, we show ourselves to be iconicity wizards, adept at crafting signals that conjure our meaning in the mind of another – with gestures or vocalisations alike. ¶

Marcus Perlman is a Lecturer in English Language and Linguistics at the University of Birmingham.

Find out more

Book

Michael Tomasello (2008) *The Origins of Human Communication*, MIT Press.

Articles

Marcus Perlman and Gary Lupyan (2018) 'People can create iconic vocalisations to communicate various meanings to naive listeners', in *Scientific Reports*, 8.

Charles F. Hockett (1960) 'The origins of speech', in *Scientific American,* 203.

Charles F. Hockett (1978) 'In search of Jove's Brow', in *American Speech*, 53.

Online

You can listen to the vocalisations of the Vocal Iconicity Challenge and test them yourself! These are available through the Open Science Framework at osf.io/x9w2z

To keep up with current research in language evolution, check out replicatedtypo.com. And to learn more about iconicity in spoken languages – onomatopoeia and beyond – you may be interested in the blog of linguist Mark Dingemanse, at ideophone.org