CrossMark

Did language evolve in multilingual settings?

Nicholas Evans[1]

Abstract   Accounts of language evolution have largely suffered from a monolingual bias, assuming that language evolved in a single isolated community sharing most speech conventions. Rather, evidence from the small-scale societies who form the best simulacra available for ancestral human communities suggests that the combination of small societal scale and out-marriage pushed ancestral human communities to make use of multiple linguistic systems. Evolutionary innovations would have occurred in a number of separate communities, distributing the labor of structural invention between populations, and would then have been pooled gradually through multilingually mediated horizontal transfer to produce the technological package we now regard as a natural ensemble.

> *instead of seeing bilingualism as a peripheral ability to be*
> *studied after monolingualism is well understood,*
> *bilingualism can be a central part of the story of language evolution*
> *Roberts 2013:192*

✉ Nicholas Evans
  nicholas.evans@anu.edu.au

[1]   CoEDL (ARC Centre of Excellence for the Dynamics of Language), Australian National University, Canberra, Australia

Introduction[1]

In this article I argue that there are many benefits to conceiving the evolution of language as having taken place in a multilingual setting. (To avoid the Catch-22 that this implies, from the outset, I should make clear that at each relevant stage 'multilingual' is to be taken as qualified by 'to an appropriate level of linguistic complexity', since at the early stage in particular the communicative varieties which our ancestors were using would have been been primitive.)

The computational (social) energy that went into creating pieces of linguistic technology was substantial—far more than we can appreciate, now we take the existence of language for granted. Multilingual conduits, by linking populations together, forced structural re-organisation and generalisation of structures towards the full suite of features that we now consider a human language. No single, isolated population had the resources to develop these, in the small-group demographics that characterised our species at the time language emerged. The model thus solves two problems at once. First, it predicts that higher-order structures in language result from individuals whose multilingual repertoires positioned them to induce generalisations about language that are less evident or even unnecessary to monolinguals (such as the arbitrariness of the sign—obvious to bilinguals, but famously not so to many monolinguals). Second, it provides a mechanism to distribute the enormous population-level innovation cost that must have gone into building the earliest languages across a multilingual web of intercommunicating groups.

More specifically, my argument will build on three assumptions:

(a)   The gradualist assumption that the evolution of language, to modern levels of complexity, required the assembling together of a number of innovations, which were at least partly independent. Language, as a communicative technology (Dor 2015), is a technological package, and just like other packages (e.g. the modern 'farming package', or the internet) its elements could in principle have been innovated by different groups, at different times and places, before gradually being brought together in a more powerful combination.

(b) The assumption, based on induction from those human populations most similar to our forebears, that even at the narrowest point of our human evolutionary bottleneck the human population would have been way too large for a single language to have been used, or maintained as a single code, across the population, within the types of social grouping then available.

(c) The further assumption, again based on induction from contemporary human populations that are the best analogues of our early forebears, that exogamy in marriage (marrying out, commonly if not universally) was parallelled by exogamy in language (learning the language of parents from two groups, of intended spouses, etc.)

Taken together, these assumptions set up a scenario in which different parts of the modern linguistic package would have been innovated among different populations, then spread across the mosaic of the early linguistic landscape by multilingual individuals. Useful traits developed in other groups would readily have been transmitted and adopted by these means, and the juxtaposition of differently structured systems would have promoted complexification—introducing more finely graded sets of structural tools and semiotic choices.

For a long time, research on language evolution has been dominated by 'the idea that monolingualism is the default, most basic state and so needs to be explained before considering bilingualism' (Roberts 2013: i; see this work for a survey of the monolingual bias in work on language evolution). But recent simulations by Roberts (2013) and Roberts et al. (2014) have shown that bilingualism can evolve from the outset, in situations where linguistic elements have a social signalling function: agents will select for more than one sign candidate if sign occurrence is sensitive to social context. They do not, however, make the case I will be arguing for here: that not only is primal multilingualism a natural evolutionary outcome from early in our speaking history, but that it was a necessary mechanism for the emergence of the suite of abilities we call language. Whereas those works are based on agent-based modelling, the present paper builds the case for primal multilingualism on the scenarios suggested by actual human societies, particularly those that form the best contemporary analogs to the small hunter-gatherer populations in which language evolved.

My paper proceeds as follows. In "Gradualism and package assembly" I make explicit the advantages of taking a gradualist position in understanding language evolution. In "The ethnographic evidence for ancient multilingualism" I survey the ethnographic evidence for regarding proto-multilingualism as plausible, and in "Multilingualism, innovation transfer, and complexification" I illustrate how it would have underpinned both the transfer of useful innovations and the complexification of subsystems when innovations originating in two distinct languages were co-present. In "Coevolution and diversity: trait evolution versus trait adoption" I relate this to some broader coevolutionary questions about the embeddedness of linguistic innovations in both biological and cultural diversification, before concluding in "Conclusions" with some final observations and proposals for future explorations that follow from the current proposal.

Gradualism and package assembly

Languages are packages of many elements at various levels, and so are the cognitive abilities underlying their use.

To conceptualise how technological packages emerge, it is helpful to think of the functioning ensemble (as an organisational, productive, economic and social unit) constituted by an eighteenth-century farm in northern Europe. The farmer grows a range of cereals (wheat domesticated in Anatolia 12 kya, sorghum domesticated in Ethiopia 6 kya, maize domesticated in central America 9 kya), raises a variety of animals for food (cattle domesticated in the Fertile Crescent 10.5 kya, pigs from Eastern Anatolia 9 kya, chickens from southeast Asia 8 kya). The land is prepared using a heavy plough developed during the Middle Ages to deal with the heavy soils of northern Europe, from an earlier and lighter prototype developed in Egypt and the Indus Valley ca. 4 kya, hitched to horses domesticated in the Eurasian steppes ca. 3.5 kya. Numerous cross-connections make this an integrated unit: some of the cereals are fed to the animals being raised, whose dung is in turn used to make the soil more fertile for crop growth. The point is that what appears, at a given snapshot in time, as a single organisational system is in fact the product of a number of quite distinct adaptations—technological breakthroughs, some requiring millennia to perfect—by distinct populations, in different times and places, afforded by different local conditions, to solve different problems. For example, the heavy plough developed in response to the clayey soils of northern Europe, which unlike the sandy soils of the Mediterranean could not be well prepared with the preceding 'scratch plough', and obviously the domestication of an animal like the chicken depends on its local availability in the wild—Southeast Asia, not Northern Europe.

It is helpful to take apart the many innovations which humans needed to make before anything like a modern language would come into existence.[2] This applies to all of the following elements, some more fundamental than others but all or most needing to be put in place before we can speak of language at modern levels of sophistication.

Adoption of an interactional engine, in the sense of dyadically coupled, closely timed conversational interaction (Levinson 2006), as a type of socially coordinated action, embedded in *cultural transmission* which allows the ratchetting up of cultural complexity (Tomasello 2008) as tried-and-tested solutions to communicative problems are streamlined, conventionalised, and transmitted.

Major architectural principles, such as compositionality, dual patterning, recursion, and arbitrariness.

---

[2] Within certain approaches to language evolution, particularly the saltationist view associated with Chomsky and his collaborators (e.g. Hauser et al. 2002), just one of these innovations, namely the development of recursion (or its intellectual descendant *Merge*), gets privileged as THE crucial step in the development of language (or, more precisely, Faculty of Language in the Narrow Sense, in their terminology). Obviously such approaches are not automatically compatible with the gradualist, multi-adaptation view adopted here, though even in that intellectual tradition a 'Faculty of Language in the Broad Sense' would include most or all of the above. To that extent, the arguments made in this paper would be limited to evolution of 'language in the broad sense', but are applicable nonetheless.

Distribution over channels, most importantly mouth/ear for speech versus hand/eye for sign and gesture.

Distribution of semiotic labour, between lexicon, morphology, syntax, prosody, gesture, and between pragmatics (inference in context) versus semantics (encoding of meaning in a context-independent way).

The evolution of shifters (deictic words), which transfer the task of contextualising communication to the here-and-now from the context, or from attention-directing gestures, into the grammatical and lexical system.

The evolution of combinatorially defined classes of signs (word classes), allowing grammatical rules to generalise over productive numbers of signs.

Evolution of grammatical categories and structures—tense, aspect, mood, evidentiality, negation. Evolving these was vital in one of Hockett's (1963) fundamental design features, 'displacement',[3] enabling the discussion of non-existent or hypothetical scenarios, locating events in time, coordinating reference with one's interlocutor's mental models, and so forth. Another type of grammatical category, concerned with *audience design and information flow*, deals with such problems as setting up question–answer pairs to seek and give information, with using *definiteness devices* (*the man* vs. *a man*) to indicate whether the referent is conceived as part of established common ground, or with using *evidentiality* to indicate the grounds for an assertion (direct perception, hearsay, etc.).

Social signalling systems for signalling individual identity, group membership, relative social roles and the like.

Developing semantic properties in the lexicon, e.g. abstracting properties like 'round' or 'green' from entity-words like 'grindstone' or 'leaf', solving the problem of developing generative numeral systems, kinship terms, systems for pulling out the dimensions of event structure (Aktionsart, volitionality, thematic roles), and metalinguistic terminology for talking about language itself.

Four consequences of the above listing are crucial to the argument I will advance below.

Firstly, it sits naturally with a gradualist view of language evolution. Some of the elements above—particularly the coupling of the interactional engine with cultural memory—are necessary to drive along the elaboration of the whole edifice. But many other elements could be evolved independently of the others—there is no logical reason why the evolution of *I* and *you*, as crucial conversational shifters, should be coupled with, precede, or follow, the evolution of tense, for example. Further, steps are scalar rather than all-or-nothing: in evolving question words, for example, we might evolve *who?* and *where?* before *why?* or *how?*, or in evolving the dual patterning of phonemes some sounds may be freed up for promiscuous recombination while others—perhaps sounds in the onomatopoeic words of some birds—would

---

[3] As Daniel Dor (p.c.) points out, humans have also invented other technologies for experiential displacement, such as drawings or maps. Language, crucially, allow for displacement of material that cannot readily be experientially displaced by these means, such as possible worlds (modalities), differences in common ground between interlocutors (e.g. definiteness), or chains of evidence ('evidential' inflections in e.g. Quechua).

remain bound to particular combinations, and not yet liberated from particular referential ties. Once this view is adopted, the evolution of language is removed from the need to be an all-or-nothing process, can respond to adaptive selection pressure (some variants, or innovations, being obviously advantageous), and, most importantly for our argument here, can in principle follow many locally different trajectories as different elements are developed in different orders.

Secondly, this view allows readily for *differential affordances*. Structures for some functions may evolve more readily in one modality than another—deictic reference through the eye-hand modality, by pointing, rather than the ear-mouth one; interactional signalling (agreement, disagreement, repair, curiosity, clamouring for attention) through the coordinated ear-mouth modality. If, as seems plausible, early language was more of a cross-channel hybrid, our model allows intermediate steps where one group runs with the affordance of the eye-hand modality in using pointing for deictic reference (and perhaps for self-reference, pointing to one's chest or nose), while another group takes the less-favoured route of encoding this function within the ear-mouth modality (developing, by some means, ways of using sound to direct attention in words like *this* or *I*).

Thirdly, it sits easily within a view of language as *coevolving* against a background of both biological and cultural difference across human populations. Different genetic or cultural biases can make the emergence of particular structures or functions more likely in some groups than others—at the phonetic level, for example, it looks increasingly likely that the emergence of tone in some populations, and clicks in others, are linked to genetic differences across populations, respectively relating to pitch perception and the shape of the mouth (see Dediu 2011; Dediu and Levinson 2013a for a survey).[4] This uneven biological baseline would make it easier for some groups to get over the innovation hump than others, but once one group has developed a particular linguistic tool—say tone, or clicks—it can be readily adopted by others, since diffusion is easier than invention.[5]

Fourthly, language evolution, like other forms of cultural evolution, exhibits cumulativity effects. The larger the vocabulary in a domain, the more precise we can be by choosing a word from the set. Consider the accreted precision of musical nomenclature in English, as alongside its own word *song* it has borrowed such

---

[4] A referee asks whether we can project this genetic variability back into the small, early populations who were evolving speech. At present we do not have an empirically based answer to this question—which depends, among other parameters, on knowing what time in the past we are talking about and exactly how the population of proto-speakers is to be delimited (Sapiens only? Neanderthals and Denisovans as well?) However, induction from current human populations of comparable size makes it seem unlikely that it would be genetically homogeneous.

[5] Ideas linking 'race' to language features were firmly rejected, within linguistics, by Franz Boas' argument that a child of any racial background can acquire total mastery of any language provided they are exposed to it from early life. But this is not incompatible with the findings of Dediu, Ladd and others that certain linguistic features correlate with genetic ones: small genetic differences between populations, iterated over many generations, can differentially favour the evolution of particular linguistic traits: 'mathematical and computational models suggest that genetic biasing of language, even if small at the individual level, can act as a forcing factor on the trajectory of language change' (Dediu 2011: 286), but this does not render them unlearnable by those from other populations once they have been evolved.

words as *chanson* or *Lied* from other languages and their associated musical traditions. In their source languages, *chanson* and *Lied* simply mean 'song', but once borrowed into English and arrayed in this choice set they take on more precise denotations drawing on the musical associations of particular periods and composers from the French and German musical traditions. Travelling in the other direction, the English word *song*, welcomed into German, denotes a particular type of song associated with Anglo-American, twentieth or twenty-first century music.[6] Cumulativity effects allow the expressiveness and precision of language to be frequently augmented through the addition of vocabulary, but also of new grammatical devices, as we will see below.

Drawing these points together, we propose two crucial characteristics of language evolution.

Firstly, it was a *cumulative, multi-sourced, socially distributed cultural invention*. Since elements of the total package are partly independent of each other, it is plausible to see them as having been 'invented' separately in different groups speaking different nascent languages, and gradually integrated into the powerful overall package we know today, just like elements in the north European farming package.

Secondly, it diffuses the hoary old opposition between monogenesis and polygenesis: instead, it is more plausible to assume something we might call *polysemigenesis*. If the emergence of the complete package was gradual and, through a long period, different groups had different partial assemblages, it follows that the distinction between monogenesis and polygenesis of language is an artificial one. How many of the above features needed to be present before 'language' had come into existence? What makes one, or some, more criterial than others? What seems more likely is that there was a process of multiple but partial emergence of the suite of features we now regard as language.

And, crucially, one can conceive of a situation where different groups had solved different sub-problems in the development of the whole package, but no one group had brought them all together. At this point in our model, multilingualism enters the picture, forming a natural conduit for the flow of adaptive linguistic innovations between groups and their assembling into an integrated system.

---

[6] Cf. this definition in the German version of Wikipedia, where the German word *Song* is defined as a particular kind of *Lied*: *Ein Song (englisch für Lied) ist ein Lied des 20. oder 21. Jahrhunderts, das sich an anglo-amerikanischen Vorbildern orientiert. Der Begriff findet vor allem in der populären Musik Verwendung und grenzt sich ab zum Kunstlied, zum Volkslied bzw. Folksong, zum Schlager im deutschsprachigen Raum und zum französischen Chanson. Anders als im englischsprachigen Raum, wo der Begriff „Song" weitgehend synonym zur weiten Bedeutung des deutschen Wortes „Lied"verwendet wird, ist im deutschsprachigen Raum der Song eine Liedgattung.* [https://de.wikipedia.org/wiki/Song]. Translation: 'A Song (English for *Lied*) is a *Lied* of the twentieth or twenty-first century, which is oriented to Anglo-American examples. The concept is primarily used in popular music and is delimited from the *Kunstlied* ('art song'), the *Volkslied* and the *Folksong*, to the term *Schlager* ('hit') in the German-speaking area and to the French *chanson*. Unlike in the English-speaking area, where the concept 'song' is broadly used as a synonym for the broad meaning of the German word Lied, in the German-speaking area the *Song* is a type of *Lied*.'

The ethnographic evidence for ancient multilingualism

As in other questions concerning language evolution, direct evidence from ancient populations is hard to find. On the other hand, when it comes to extrapolation from observable populations we are arguably in a better position when it comes to sociolinguistics and demography than we are with regard to language structure. Apart from pidgin/creoles and emergent sign languages, linguists generally hold the view that no modern languages are primitive—all have developed to equally high levels of structural sophistication (with some additional add-ons due to literacy perhaps being an exception). On the other hand, we observe very different patterns of language use as we move along the double continuum of economic organisation and scale of social unit. As we move into the realm of hunter-gatherers, and certain other types of small-scale society such as shifting cultivators, we observe certain characteristics wherever these groups are found in the world. Many have pointed out that hunter-gatherer societies provide the best analogues to the social and demographic conditions that shaped us through the longue durée of most of our shared human past: 95–99.999% of our history, depending who we are. The key features of interest here are:

(a) *small demographic size for languages*—bands or, later, clans—that sometimes produce stable language⟺group numbers of as low as seventy-five in the Australian language Gurrgone (Green 2004),[7] and only rarely exceed a few thousand. In Australia—often characterised as the only 'continent of hunter-gatherers'[8]— the average number of speakers per language at the time of European contact was probably somewhere between 650 and 3000.[9] On the island of New Guinea, which had a predominantly agricultural population but no significant larger state formation, the number of speakers per language was probably 3300–5000 before colonial contact.[10] In fact, without the formation of some form of complex state, we can take speaker-populations at this level or below to be the norm for human groups—the centrifugal political mechanisms for diffusing and integrating linguistic norms, and the value of using linguistic difference to signal group membership is high enough to promote an almost incessant dynamic of language

---

[7] This is not too far below Dunbar's (1992) 'comfortable' human group size of 148, though I hasten to point out that, as argued here, multilingualism ensures that the social group is substantially larger than the language group.

[8] Though this may be exaggerated—see Pascoe (2014) for an important recent critique arguing for forms of agriculture and other types of sedentary food production (fish traps, eel-channels etc.) over much of the continent.

[9] Further, the likely skewing of speaker-population sizes along a log-normal distribution means that the average speaker population is likely to have been even smaller than this average suggests. I am grateful to an anonymous referee for pointing out this consequence.

[10] Before European contact, there were probably around 250 languages in Australia, though some recent estimates push this up to 407 (Bowern 2016)—and given that estimates of precontact population range from 250,000 to 750,000, this is roughly 650–3000 speakers per language. For New Guinea, a rough estimate of total number of languages is around 1200, for a precontact population of perhaps four to six million, giving the number of speakers per language as somewhere between 3300 and 5000.

diversification (cf. François 2012). The relevant human population size, at the period during which language evolved, is difficult to investigate—it depends on whether we see language origins as happening 200,000 years ago, or earlier before the Sapiens-Neanderthal split, or perhaps gradually over many hundreds of years. If we take a figure of 10,000–50,000 people as a low ballpark figure, and a population size per language of 50–1000, that gives us a numerical range of 10–1000 languages during the key period of language emergence. If we go with the much larger population size of 120,000–325,000 individuals proposed by Sjödin et al. (2012) for Sub-Saharan Africa some 130 kya, this would give us a numerical range of 120–6500 languages, and the thin spread of humans over a vast area would favour exactly the sort of variation and multilingualism being argued for.

(b) *exogamy (out-marriage) and open social networks*—in small groups out-marriage is common, and is seen as bringing many advantages, such as a clear means for avoiding incest, far-flung allies, and access to territorial resources of in-laws or through alternative lines of descent (e.g. through one's mother's as well as one's father's line). This is particularly important in fragile environments where unpredictable rainfall patterns mean it is useful to have a range of potential allies, or distant family, in times of local resource scarcity. Exogamous patterns may range from direct sister-exchange through more complex systems of circulation of marriage-partners between social units such as clans, up to exchange between larger social units like subsections in Australia. And a good proportion of these individuals come from other language groups. Looking upwards in the family tree, this results in situations were parents may speak different languages, and grandparents may introduce even more. It may also produce alliance units which are stably bi- or multilingual by their very constitution. In the Morehead district of southern New Guinea, for example, marriage involves direct sister exchange and the ideal family structure is binuclear—a pair of brother-sister pairs, each residing in their own location (e.g. different villages, with each sibling pair belonging to a different clan). Special kinship terms, in languages of the region like Nen, designate such relationships as *miti* for 'double cross-cousins resulting from direct sister exchange' (i.e. if my mother is your father's sister, and my father is your mother's brother) or *mitadma* for 'aunt/uncle who is the sibling of one of my parents and the spouse of the sibling of my other parent'—see Fig. 1.

The two halves of this binuclear unit visit each other frequently. If, as commonly happens, the exchanged siblings come from different languages, this guarantees an intense lifelong exposure to both languages for the children of such unions. Moreover, since the other members of one's parent's sib-sets normally contract exchange relationships in different directions (e.g. my father may take his wife from the Gecko clan, but my father's brother takes his wife from the Crocodile clan), the outward links emanating from any particular lineage look rather like a spiral staircase where each marriage goes out in a different direction, bringing further languages into the mix, giving further language groups to whom one is closely related. In consequence, one commonly finds individuals with impressive multilingual portfolios. Jimmy Nébni is a typical Bimadbn village resident: he speaks his 'own' language Nen (also

that of his father and his wife's mother), his mother's language Idi (also that of his father's mother), his wife's language Nambu, several other local languages to varing degrees, English, Hiri Motu (the local lingua franca), and Tok Pisin (the national lingua franca). This portfolio spans four quite unrelated language families (Germanic/Indo-European, Yam, Pahutori River and Austronesian). His situation is far from atypical in the regions I have been discussing—and note that, since many of the languages are acquired early, one does not find the kind of reduction in complexity which typically accompanies later-life learning of a second language (Lupyan and Dale 2010).

Looking outwards in the mate-search, learning the language of one's future spouse and parents-in-law is a good strategy whose value is recognised in many parts of the traditional world (see White 1997 for northeast Arnhem Land and Leenhardt 1946 for New Caledonia).[11] In some regions, such as the Vaupes region of the upper Amazon, this tendency even gets formalised to the point of 'linguistic exogamy', a stipulation that one's spouse should come from another language group. As one Barasano speaker from the Vaupés region of the upper Amazon told anthropologist Jean Jackson (1983: 70): 'If we were all Tukano speakers, where would we get our women?'. Moore (2004), who worked in the Mandara mountains of northwestern Cameroon, offers a fine ethnographic study of how young men bone up on the clan languages of girls they are courting—even if they have one or two languages in common already.[12]

Of course it is likely that early humans did not yet have the instititional or conceptual superstructure to formalise more complex and juridicalised arrangements expressible in language, but a general principle of out-marriage, at least for a

---

[11] The usefulness, in the quest for a future spouse, of learning the language of one's mother in addition to that of one's father, comes out particularly clearly in this characterisation by Leenhardt (1946) of the traditional situation in New Caledonia: Les femmes enseignent aux enfants leur langue maternelle; de quelques pays qu'elles viennent, elles préparent leurs filles à aller un jour au pays de l'oncle utérin, et la connaissance de la langue du kaña leur paraîtra toujours indispensable dans ce but. De même, leur fille devra comprendre la langue du "frère" boru ña, qui sera un jour son mari. … nombre de jeunes gens continuaient leur séjour, et ne revenaient qu'après avoir épousé la femme qu'ils allaient ramener chez eux. Durant ce temps, ils avaient appris à fond la langue. Leur femme parlera sa langue en même temps qu'elle apprendra celle de son mari. Ainsi tout indigène était pour le moins bilingue. ['Women teach their children the maternal language; from wherever they come, they prepare their daughters to go one day to the country of their maternal uncle, and the knowledge of the language of the *kaña* is indispensable for this. In the same way, their daughter must understand the language of the 'brother' *boru ña*, who one day will be her husband… [After certain feasts in the country of the maternal clan] a number of young people will stay on, not returning until they have married a woman who they will bring back to their country. During this time, they will have master the language. Their wife will talk her language at the same time as she is learning that of her husband. Thus every indigenous person is at least bilingual.] (Leenhardt 1946: xvi; my translation).

[12] Moore gives, as an example, the courting by a young man called Jonas of a girl called Gogo in her mother's compound in the Mandara Mountains, Cameroon. To enhance his chances, Jonas courts her in Mada, her father's language, even though they already had two other languages in common: Wandala, the local lingua franca, and Wuzla, the first language of Jonas' father and Gogo's mother. (In addition Jonas speaks five other languages). Prior to visiting Gogo's mother's compound, Jonas had jotted down a list of topics of conversation and relevant vocabulary on a piece of paper, but did not need to use them during the conversation.

proportion of individuals, is likely to have obtained.[13] Indeed, as an anonymous referee points out, some form of exogamy is found in the majority of human societies: going by information from the Ethnographic Atlas compiled in the D-PLACE database (https://d-place.org), of the 1102 societies with data, only 344 have no form of exogamy, while 758 have some type of exogamy.

(c) *egalitarian multilingualism*—In modern societies in which two or more languages are deployed, these tend to be functionally specialised, e.g. the language of the home versus the language of schooling, the language of the local group versus the language of the state. But in many small-scale societies, multilingualism is 'egalitarian', in the sense that each group sees their own language as appropriate and emblematic for their own social unit, while conceding the equivalent role to other languages in the broader social universe.

As François (2012: 93) puts it, in his discussion of the highly variegated linguistic mosaic of northern Vanuatu:

the two phenomena—socially emblematic differentiation vs. widespread contact—should really be viewed as two sides of the same coin. The reason why Melanesian communities could afford such linguistic diversity is precisely their constant willingness to learn the tongues of their neighbors. Within such a unified social network as the Torres and Banks archipelago, the indulgence towards language fragmentation is only sustainable as long as the social norm is to preserve egalitarian multilingualism. While linguistic diversity is arguably triggered by the desire for social emblematicity, it needs egalitarian multilingualism to be maintained over generations. (François 2012: 93)

In many regions where this is the norm, language choice is a powerful group-signalling mechanism, of relevance not just to showing group membership but also for validating one's relationship to 'country', through a host of cultural connections. Among such regions one may count indigenous Australia, many parts of New Guinea, Vanuatu, many parts of South America (e.g. the Upper Vaupes, the Gran Chaco), and the Mandara mountains of Cameroon.[14]

For example, in northern Australia the existence of multiple languages is cosmologically legitimated and an essential part of ensuring the complementarity of

---

[13] A 'stretched frontier', for example the linear peopling of a coastline, may have somewhat limited options, as compared to a more densely populated area where one has neighbouring groups on all sides, but all but the 'tip group' would have at least two options, one in front and one behind, from whom to draw mates.

[14] I am ignoring many differences of detail here across the different regions mentioned here. For example, multilingual capacity may, in everyday practice, be played out in 'asymmetric bilingual conversations', where each party speaks their own language but understands the others, or participants may shift into whichever language is appropriate for their location. In some regions, this may result in people regularly exhibiting an active command of several languages, while in others they only speak one, but 'hear' others. Either way, however, the knowledge enabling them to interact must extend to the structures of two or more languages.
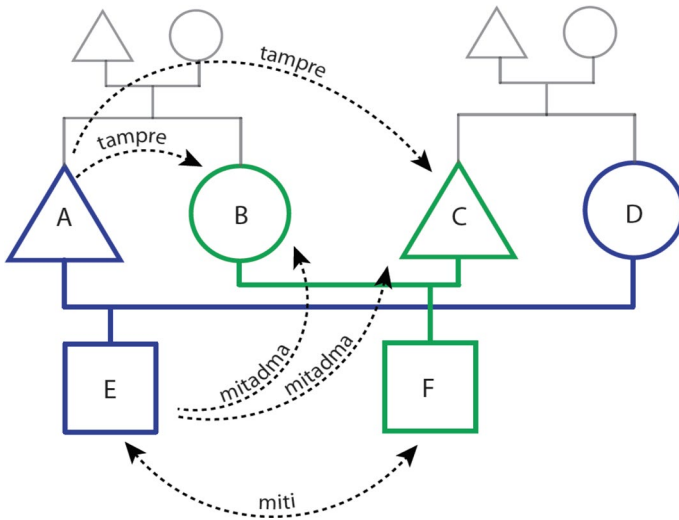
Fig. 1 Terms used between kindred in sister-exchange relations (Nen language, Morehead district, Southern New Guinea)

groups in terms of both territorial and intellectual/cultural assets (Evans 2003a, 2010, 2011; Merlan 1981; Rumsey 1993). Sutton (1997) captures this well in his seven principles of multilingualism and language difference in Aboriginal Australia, including that languages (1) are owned, (2) belong to specific places, (3) imply, through a particular linguistic choice, knowledge of, and connectedness to a certain set of people in a certain part of the country, (4) are relational symbols, connecting those who are different in a wider set of those who are the same, (5) are internal to society, not markers of the edges of different societies.

The Warramurrungunji ancestress in the creation story of the Cobourg and Western Arnhem Regions[15] travelled through the landscape, sowing each ecozone with its own food type (yams here, waterlilies there) and putting her children in different places, telling them what language they should speak there. Map 1 shows a 200-km transect of part of Warramurrungunji's route, passing through the territories of nine clans and seven languages from four language families, at least as different from each other as Germanic, Slavic, Indo-Aryan and Romance.

Unlike the Abrahamic Babel myth which inflicts multiple languages on the world as a curse for human presumption, in regions such as those mentioned multilingualism is a magnificent boon, assuring complementarity of groups and tying them to

---

[15] Needless to say, this is presented here not as an actual account of early human lingualism, but rather to show in a particularly vivid way how deeply ingrained the social-signalling function of language can be in many cultures, and also how taken for granted it can be that cultural blocs can span multiple languages. The Warramurrungunji myth has been recorded in a number of languages—Iwaidja, Kunwinjku, Gun-djeihmi—from people belonging to quite different clans. See Evans (2010: 5–8 for more details of this myth).

their own clear territories. As the linguist Don Laycock (1982) was told by a man from the Sepik region of Papua New Guinea: "it wouldn't be any good if we all talked the same; we like to know where people come from."

Particular tracts of land will be associated with particular languages and it is hazardous to speak other tongues there. A custodian will introduce visitors to those places by calling out to the spirits in the local language as a guarantee of safety and recognition, and mythical narratives will index the movement of their characters from one place to another by shifting languages during the performance—in the full expectation that listeners, suitably polyglot, will appreciate their art. Religious ceremonies typically have several 'legs', for each of which a different clan/language group is responsible, which depict events in a different country and may be sung in a different language and musical idiom—as if the *Odyssey* were not confined to Greek but instead shifted through various languages of the ancient Mediterranean as it traces the steps on Odysseus' journey. In many areas, languages names themselves reflect widespread metalinguistic knowledge of multiple tongues. Thus in Southern New Guinea (Evans 2012a) most language names—Nen, Len, Nambo, Idi and so forth—are simply the word for 'what' in the respective language, as if English were called *whattish*, French *quoiais*, German *wassisch*, Welsh *bethaeg*, and Russian *shtoskiy* based on the respective words *quoi*, *was*, *beth* and *shto* for 'what'.

But even at this fine level of grain, the variation does not stop—it keeps going all the way down, as shown for example by Meyerhoff's (2017) work on vowel realisations and subject agreement in the Vanuatu language Nkep, or ongoing work by the Wellsprings of Linguistic Diversity project targeting such variables as initial velar nasals in Bininj Kunwok (Marley 2018), final nasals in Idi (Schokkin 2018), or emerging prominence markers in Nen and Nmbo (Evans et al. 2018). Even in small communities, therefore, variation is constantly being generated and harnessed to semiotic use.

For the purposes of our larger argument, the ethnographic detail we have been examining is not intended to serve as an exact model of how early humans would have been. But it is intended to remind us how readily people acquire impressive levels of polyglot proficiency, without any need for formal training. More importantly, it is a reminder of how closely multilingualism is tied to a cluster of factors that include small group size, out-marriage, and the harnessing of linguistic difference to the signalling of group membership (and ties to land) and, for an individual, of the 'ropes' of alliance, contacts, knowledge and credibility with other groups that they have managed to build up through their lives. For an individual, polyglot mastery suggests an unusual breadth of ceremonial contacts and far-flung social capital, eliciting expressions of admiration, in indigenous Australia, like 'he travellin man himself' (Evans 2011; Sutton 1997). For a group, positioning themselves as bi- or multilingual emphasises their connections as brokers between other groups—such as clans like the Barabba in Central Arnhem Land that define themselves as 'Kune-Dangbon' (two different languages) (Evans 2003b). It is reasonable to assume that all of these factors were present, albeit in less sophisticated ways, among our early ancestors. This sets up a plausible scenario of widespread early multilingualism—at both individual and group levels—whose consequences for the development of language we now examine.

Multilingualism, innovation transfer, and complexification

Multilinguals are the natural agents of horizontal transfer across languages. Their mental representations contain the distinct structures and units of two or more languages, and their communicative practices potentially draw on this whole pool as they seek to solve expressive problems (including positioning themselves socially). Not infrequently, this process produces new, elaborated linguistic systems drawing on elements from two or more languages.

A celebrated example is Michif (Bakker 1997), a mixed language combining Cree and French elements that emerged from the children of Quebec French trappers and Cree women in a bilingual setting where this community had to move between these two worlds, from a bilingual matrix of French-speaking trappers and their Cree wives.[16] (Michif derives from the Québec French pronunciation of *méti* 'mixed, person of mixed descent'.)

Not only does Michif put together sounds from both contributing languages (its phoneme inventory is close to the union of the French and Michif inventories), but elements of its grammar combine separate sets of grammatical distinctions made in the two languages. French gender and number—manifested in article choice and adjective agreement—opposes masculine to feminine singulars against plural. Cree gender—manifested in demonstratives—opposes animate to inanimate. The basic French noun phrase combines an article with a noun; the basic Cree noun phrase combines a demonstrative with a noun. Michif can put all of this together, lining up a French style article, then a Cree-style demonstrative, then the noun. Crucially, agreement needs to take both semantic contrasts into account (Fig. 2): an expression like *this girl* picks animate for its demonstrative and feminine for its article, while *yon fields* picks inanimate for its demonstrative and plural for its article:

'Mixed languages' like Michif are relatively rare, frequently short-lived, and linguists have only become interested in them recently. (For interesting discussions of two nascent mixed languages in indigenous Australia, recorded as they emerge, see Meakins 2011; O'Shannessy 2012, 2016).

But we can illustrate the same principles of elaboration by bilingual contact with many less dramatic examples representing more 'normal' processes of language contact. For example, the Dravidian language Kannada descends from an ancestral language that resembled Tamil in lacking a voicing contrast (e.g. p vs. b), or aspiration (e.g. p vs. pʰ), but new sounds adopted through contact with Indo-Aryan languages (from Sanskrit onwards) have introduced these phonetic contrasts, which are reflected in the writing system (even if not all speakers maintain these in-grafted

---

[16] The most likely sociohistorical scenario is that, in the first generations, descendants of these first mixed marriages were bilingual, and served as go-betweens between Cree hunters of fur and Quebec French fur-traders. Subsequent legal changes in Canada, which formalised ethnic group membership, left the Métis in a position where they belonged neither to the recognised indigenous tribes nor to the mainstream white population. At that point fluency in one of both of the formant languages would have atrophied, and a new mixed language appears to have emerged, though a process of what Bakker (1997) calls 'language intertwining', as a group marker. In subsequent generations this left Michif speakers whose language repertoire did not include the source languages.

Map 1 The pathway taken by founding ancestress Warramurrungunji, showing the clans and languages she established

distinctions in casual speech). As a result, the Kannada phonological inventory has been substantially expanded. A well-known and comparable case from Southern Africa involves the adoption by Bantu languages such as Xhosa and Zulu of a

number of click sounds from the Khoisan languages in which they came into contact as they moved southward.

Similar processes of complexification can occur in all parts of the language system.[17] As a semantic example, we examine the semantics of noun classes in the bilingual borderlands of the Bininj Kunwok and Dangbon languages of Western Arnhem Land (Evans 2003b). Dangbon was traditionally in intensive contact with the eastern dialects of Bininj Kunwok (Kune, Kuninjku), and some clans (such as Barabba) even defined themselves bilingually as 'Kune Dangbon'. Other varieties of Bininj Kunwok, such as Kunwinjku, were further away from Dangbon and knowledge of it was much less prevalent. Though fairly closely related, the subclassification of nouns in Bininj Kunwok and Dangbon follows quite different principles.

The original Bininj Kunwok system, as exemplified by the conservative Kunwinjku dialect (Fig. 3) has a five-class system shown by prefixes and which assigns classes to semantically-based ontologies (masculine, feminine, vegetable, neuter, with a residual fifth class unprefixed).

The Dangbon system (Fig. 4) makes an opposition between part nouns, which are obligatorily possessed, and absolute nouns, which need not be. Part nouns predominantly include parts of the body ('his nose'), of plants ('its seed'), and of the landscape ('its billabong').

The eastern Bininj Kunwok dialects, from the clans which identified as traditionally bilingual, illustrate what happens when these two different semantic systems are combined (Kune Dulerayek, Fig. 5).

Here summative complexification has integrated the full set of distinctions made in the two neighbouring systems, maintaining the gender and vegetable features found in Bininj Kunwok on the one hand, and the part versus absolute distinction from Dangbon on the other. In doing so, it splits each of classes III and IV from Bininj Kun-wok into part (alternating structures) versus absolute (fixed), but at the same time retains the vegetable versus neuter contrast in part nouns.

The intersection of these two systems has created subclasses not found in either neighbouring variety: vegetable parts like 'seed' (which allow either the *man-* prefixed 'vegetable' structure of the *-no* suffixed possessive structure), and non-vegetable parts like 'eye' (which allow either the *kun-* prefixed 'neuter' structure or the *-no* suffixed structure).

There are good cognitive reasons for the type of category elaboration under multilingual contact that this example illustrates. Bilinguals must attend to, remember, and formulate information in ways that are sufficiently precise for the purposes of both speech communities they participate in. To do this, their best cognitive strategy is to use an elaborated conceptual grid of the type exemplified here, which makes

---

[17] For an interesting line of research at the intersection of semantic and social structure, see the series of studies by McConvell (1985, 2018) which trace the genesis of Australia's unique eight-class 'subsection' system—which assigns every individual to one of eight classes, effectively providing a schema for their kin relation to every other member of the social universe. Most proximally this appears to have originated through the interaction and integration of two isomorphic but differently named four-class systems across a language boundary among bilingual speakers; more distally, four-class systems in their turn may have originated through comparable interactions of two-class 'moiety' systems.

Cree *awa* 'this (animate)'   *ne:ma* 'yon (inanimate)'

awa la fij 'this girl'   *ne:ma lı šã* 'yon fields'

French *la*   *fij (fille)*   *le* (<*les*)   *šã* (*champs*)
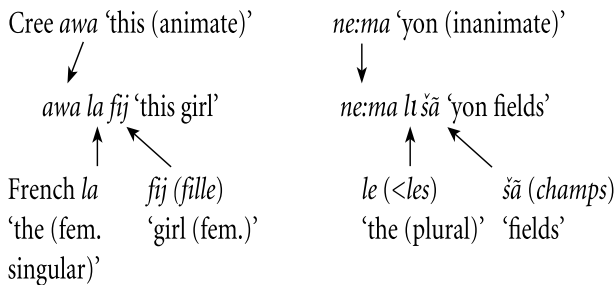'the (fem.   'girl (fem.)'   'the (plural)'   'fields'
singular)'

Fig. 2 Cree and French determiners and agreement in the Michif noun phrase

all distinctions needed and permits ready intertranslatability. From the elaborated ontology in Fig. 5, which effectively integrates two dimensions of semantic distinction, they can readily map to the semantic ontology of either of the languages they speak: to speak Dangbon, they simply retain the contrast between part (IIIp and IVp) versus others, and to speak other Bininj Kunwok dialects such as Kunwinjku, they simply drop out the part versus absolute contrast in classes IIIp versus IIIa and IVp versus IVa, and retain the categories given here by the Roman numerals.

Multilingually mediated contact between languages does not just accumulate new categories—entirely new structures can also arise through exaptation mediated by the swirl of contact. Consider the development of the Mediterranean alphabet or Japanese syllabic script. In each case these major technological breakthroughs occurred because a notational system well-adapted for one sound system did not work well for another. From the adaptations that needed to be made, qualitatively new structures emerged.

Consider the emergence of the alphabet in Greek. In Semitic languages, the linguistic structure meant that vowels could generally be worked out from context and didn't need to be shown, whereas in Greek, vowels were much more important, a notational need that was met by taking over some unneeded Semitic letters. Thus ∢ aleph, the glottal stop ʔ and originally a rebus[18] based on the first syllable of western Semitic *ʔalif* 'ox', was used for the vowel/a/, and ○, the pharyngeal fricative ʕ and originally a rebus of an eye based on the proto-Semitic *ʕayn* 'eye', was used for the vowel/o/. The transition from representing consonants to representing vowels was aided by the fact that pharyngeals colour the following vowel, so that whereas a 'clear' *a* sound would come out in the Phoenician pronunciation of ∢ as/ʔa/, the pronunciation of the vowel after ○ sounded more like an /o/ to the Greek ear. This innovation allowed humans, for the first time, to represent speech through a string of distinct consonant and vowel symbols, spawning the huge number of alphabets now used, in one form or another, to write languages on every continent. (And

---

[18] A rebus is a symbol that borrows the sounds of an easily drawn word to represent a homophonous word that is more difficult to depict visually, e.g. using the symbol 🐝 to write the English verb 'be'.

Class	Prefix	Sample word	Meaning
I (Masc)	*na-*	*na-koʔpaɲ*	'old man'
II (Fem)	*ŋal-*	*ŋal-koʔpaɲ*	'old woman'
III (Veg)	*man-*	*man-tacek*	'Grevillea pterydifolia'
		man-mim	'seed'
IV (Neut)	*kun-*	*kun-wok*	'language'
		kun-mim	'eye'
Unprefixed	*ø-*	*taluk*	'woman'

Fig. 3 The noun class system, Kunwinjku dialect of Bininj Kunwok

its technical derivative, the international phonetic alphabet, is able to write all the sounds of all human language in an extended, standardised alphabetic notation.)

Japanese, likewise, adopted a writing system that had been evolved for writing quite another sort of language, which did not fit the structures of its own language particularly well. The thousands of characters in Chinese dovetailed beautifully with Chinese phonological structure—syllabic in structure, without inflection, and with tones that significantly multiply out the number of the possible syllables. But for Japanese, with its small number of simple syllable types, its many multisyllabic words, its lack of tones, and its many inflectional suffixes, Chinese characters were not an efficient system. A first step in adapting Chinese script to writing Japanese (man'yōgana) was to fix a set of kanji (characters) by phonetic value, and use these to write grammatical elements such as suffixes. Subsequently these were simplified in form by the Buddhist priest Kūkai, whose visits to India had exposed him to the Indian Siddham script,[19] yielding the syllabary that is now the primary means of writing Japanese.[20]

With these examples I have shown some of the ways that multilingual speakers act as vectors of horizontal transmission of features between languages. While system elaboration, the outcome I have focussed on here, is by no means the only possible outcome—there can also be convergence,[21] divergence or simplification—it

---

[19] The Siddham script, from which the Bengali, Tibetan and some other scripts evolved, is neither an *alphabet*, which has a distinct letter for each sound, nor a *syllabary*, which has a distinct letter for each syllable, e.g. な *na* versus ぬ *nu* in Japanese hiragana. Like the Semitic scripts, the Siddham script is an *abugida*: letters have an 'inherent' a-vowel which will be pronounced in the default case (e.g. Devanagari न *na*), but can be deleted and replaced by modifying the main letter (e.g. the underswirl in Devanagari नु *nu*).The fact that Kūkai's exposure to the Siddham script led him to produce a script which, as a syllabary, differs typologically both from Chinese and from Indic abugidas is further testimony to the way that multilingualism (effectively between three languages in this case) can lead to the emergence of quite new structures.

[20] *Hiragana* is primary in the sense that it is learned first, and can be used to write any word; characters are still used alongside it. (There are in fact two syllabaries, *hiragana* and *katakana*, specialised for different purposes, one developed from the old regular script and one from the cursive script.).

[21] A referee raises the question of whether cross-borrowing would lead to homogenisation. The study of 'linguistic areas' in historical linguistics, whereby unrelated or only distantly related languages converge through time, has certainly identified several putative convergence zones or *Sprachbünde* (e.g. the

Class	Suffix	Sample word	Meaning
Part	*-Poss*	**ce,*	**nose
		ce-no,	'his/her nose',
		ce-ŋan	'my nose'
		etc.	
Absolute	*(-Poss)*	*ʈolu,*	'dog'
		ʈolu-no,	'his/her dog'
		ʈolu-ŋan	'my dog'

Fig. 4 Noun classes in Dangbon

Class	Affix	E.g.	Meaning
I (masc)	*na-*	*nakoʔpaɲ*	'old man'
II (fem)	*ŋal-*	*ŋalkoʔpaɲ*	'old woman'
IIIa (veg abs)	*man-*	*mantacek*	'grevillea sp.'
IIIp (veg part)	*man- ~ -no*	*manmim ~ mimno*	'seed'
IVa (neut abs)	*kun-*	*kunwok*	'language'
IVp (neut part)	*kun- ~ -no*	*kunmim ~ mimno*	'eye'
Unpref	*ø*	*taluk*	'woman'

Fig. 5 Noun classes in the Kune Dulerayek variety of Bininj Kunwok

is the one that is most relevant to our argument, since it shows how multilingually mediated complexification can lead to the accretion of linguemes[22] from multiple sources. However, the cases I have been focussing on have all involved very specific subunits rather than fundamental building blocks—for the good reason that we are exemplifying with modern languages in which these building blocks were all already in place. In the next section we ask how a comparable process might have applied to the evolution of language at earlier stages.

Footnote 21 (continued)

Balkans, South Asia, Mainland Southeast Asia) in which languages come to possess certain structural features in common (e.g. not using infinitives in the Balkans, developing retroflexes and verb-final syntax in South Asia, tone, monosyllabicity and serial verbs in Mainland Southeast Asia). But in no case is the convergence perfect, and if anything the direction of current findings for most of the classic convergence zones is to show that they are lot less well-defined than has been classically believed. As the distinguished contact linguist Sarah Thomason (2000) puts it: 'Even in the strongest Sprachbünde, the often-cited "tendency toward isomorphism" rarely if ever leads to massive overall convergence.'

[22] This is a handy term (Croft 2000) for discussing units of linguistic structure in a way that is non-committal with regard to whether they concern sound (phonemes), word-structure (morphemes), etc.

Coevolution and diversity: trait evolution versus trait adoption

As argued in "Gradualism and package assembly" section, many of the fundamental design elements of language are not inherently dependent upon each other, in terms of order of introduction. For example, there is no given order in which the three distinct problems of evolving a means of expressing negation, of expressing time/tense, and of developing personal pronouns based on role in the speech act (I: speaker, you: addressee). Imaginatively, we can recast the modern (1), which can do all this through evolved grammatical mechanisms, with the 'semi-evolved' variants (2), (3) and (4a, b, c), each of which manages without a grammaticalised solution to one of these three problems:

(1)   I won't see you tomorrow.
(2)   I will see you tomorrow. [accompanied by head-shake, pushing away gesture, etc. to show negation].
(3)   I not see you. [accompanied by pointing to sun as it moves to set in west, then looping back around to where it will rise tomorrow.]
(4a) Writer won't see Reader tomorrow.
(4b) Fred won't see Kim tomorrow. [speaker is called Fred; addressee is called Kim].
(4c) [pointing to myself] won't see [pointing to you] tomorrow.

It is thus logically possible to envisage a situation where three different speech communities each solve one of these problems, which we illustrate here by means of a thought experiment involving three invented speech communities. The Gugu group develop a conventionalised form of negation, the Bogons work out a simple way of encoding tense, while the Sabas develop a method of encoding person (*I* and *you*). A bit later, when the Gugu group comes into contact with the Bogons, start intermarrying with them, and bilingual Gugu-Bogon speakers emerge, they bring both negation and tense together into an elaborated new form of language that can do both by linguistic means. Analogous processes occur in the speech of bilingual Bogon-Sabas speakers, and in a subsequent move communication between Bogon-Gugu and Bogon-Sabas bilinguals ends up accumulating all three innovations into a single code.

So far my argument has, in principle, been neutral between whether these developments took place using an oral-aural channel, a manual-visual one, or a hybrid one. We can make the argument more interesting, and probably more realistic, by assuming that at some early phase human communication was a hybrid, with communicative tasks distributed relative to the affordances of the two channels. For example, in establishing joint attention to a locatable object, some type of pointing (finger, lip, eye-gaze), being iconic, is probably easier to evolve than an arbitrary oral sign like *this* or *kore*. On the other hand, the vocal channel is well adapted for indicating the desire for something (e.g. through the sorts of intonational contour a young child makes when wanting something) or a questioning attitude (though rising contours associated with questions around the world). In depicting the natural world, the vocal channel is a natural candidate for developing bird names (onomatopoeic names based on their calls) but the manual channel is well-suited for distinguishing

species of animal or fish (e.g. macropod types, based on imitating different gaits, or fish types, based on their morphology or manner of movement).[23]

One important set of steps in language evolution, then, is the gradual oralisation of language (naturally excepting sign languages). To achieve this, numerous functional components needed to go through a transition from manual-visual to oral-aural channel. Again, there are many such transitions, and most would have been logically independent of each other.

We bring in some more imaginary ancient groups into our argument here. The inland Fidils, though predominantly users of the manual-visual channel, have developed a way of indicating negation orally by saying something like 'uh-uh' (the negative response-marker) while signing. The east-coast Movovs, with whom they intermarry, long ago started developing a range of vocal bird names based on imitations of their calls, among them *vaakvaak* 'crow' for the birds in the dry west of their territory, and *haʁ*[24] for the penguins inhabiting the sea to their east. At some point these words develop the secondary meanings 'west' and 'east' respectively, and eventually take on the meanings 'later' and 'earlier' as well, based on the sun's trajectory, before making a final leap of abstraction to 'future' and 'past'. Intermarriage and growing Fidil-Movov bilingualism leads to the swapping of both these innovations between the two languages, making them the first people in the world able to express both negatives and tense by verbal means.

The above examples involve independent functional domains of language. But we can apply our model in an even more interesting way when we look at functional couplings. Consider the question–answer system that drives much of dialogue, and the transmission and enrichment of knowledge. Modern languages can ask questions like 'where?', 'whither?' or 'what?' and answer them with demonstrative words like 'there', 'thither' or 'that'. English, like many languages, allows a range of spatially calibrated answers—*here* versus *there* (and *yon* in older English), *hither* versus *thither*, *this* versus *that*. Moreover, many languages are like English in having proportional formal relationships between question words and demonstratives. The English formula is *wh-* for questions, *h-* for proximal demonstratives, and *th-* for distal demonstratives, but we don't do this consistently (we don't express 'now' with *hen*, or answer *which?* with *hich* or *thich*). Some languages, like Japanese or Tamil, have systems that are both richer and more consistent (see Evans 2012b for more details). Japanese makes three distance distinctions (*k-* 'near me', *s-* 'near you', *a-* 'near neither of us', e.g. *kore* 'this', *sore* 'that (by you)', *are* 'that (away from

---

[23] The initiation language Demiin, taught to second-degree initiates of the Lardil group in northern Queensland, presents an interesting example of speech-gesture hybrids (Hale 1973; McKnight 1999). Its set of spoken signs compresses the whole of Lardil vocabulary down to less than 200 words, but these are then disambiguated through gesture. Thus the word ⁀ɬi can refer to any (gilled) fish (⁀ɬ is an ingressive lateral fricative sound unique to Demiin), but the different types of fish are disambiguated by making an appropriate gesture while saying the word, e.g. for 'parrotfish' (*ngerrawurn* in Lardil) the hand is held with the thumb out and up but inclined slightly: the thumb represents the dorsal fin and the inclination the fact that these fish generally tilt while eating coral.

[24] Note for the non-linguist: in the international phonetic alphabet the upside-down capital R, ʁ is an Edith-Piaf style 'uvular trill', most often accomplished by English speakers when gargling.

us both)' and combines these plus the interrogative marker *d-* with a much wider range of ontological markers (e.g. *-oko* 'location', *-ou* 'manner'). What we needed to evolve, then, for a question-and-answer system that could function in the immediate context, was a system that opposed questions to deictic answers, and ranged them over ontological categories like space, direction, identity, time, manner and so forth.

The easy bit of evolving such a system is the distance deictics—something readily solved by pointing (with an affordance preference for the gestural, as mentioned above). Many of the ontological categories can also be expressed rather easily by gestural means, e.g. a dynamic rather than a static point for 'hither' as opposed to 'here', or pointing to sun-position (or perhaps bringing in our east:west::past:future proportion) for different types of 'then'. However, getting ontologically specified question words by gesture is much harder—where does one point to show ignorance? Here let us imagine a communicative breakthrough by one group, in the oral-aural channel, using a plaintively questioning 'want-to-know' intonation as a general question, and relying on pragmatic uptake by the addressee to work out what the questioner wants to know about.

Once again let us run this through three of our groups, the Fidils and Movovs, long in contact with each other, and the Bogons, who recently moved into their region and who have been absorbed into their multilingual bloc through the exchange of spouses.

The Fidils, continuing their track-record of exapting emotive changes in pitch for semiotic purposes, are the ones to develop a conventionalised 'I am ignorant and want to know' tonal contour.

The Movovs, long used to using pointing to distinguish *here*, *there* and *yonder*, combine this with their penchant for talking about time in terms of the sun's east–west trajectory and their birdcall-originated verbalisations for these, to become the first human beings to develop a clear word for 'then', which they do by combining the 'there' point with the penguin-derived *haʁ* vocalisation denoting 'penguin; coastal area; east, earlier, before'. They also combine the there-point with the crow-derived *vaakvaak* 'west, later' to express the meaning 'then (later)', and at some point a creative bardic spirit among them puts these together to come up with the new compound word *haʁ-vaakvaak*, combined with a sweeping east-to-west point, which means 'then' in the modern sense that is neutral with respect to past or future. (Eons later Chinese scribes would create signifiers for abstract meanings in similar ways when they combined 月 'moon' and 日 'sun' to give 明 'bright'.) Frequent use erodes the compound from *haʁ-vaakvaak* to *hʁvaak* 'then' and by this time the point-and-sweep gesture is redundant and often omitted.

So far neither group has worked out a way to ask information-questions efficiently. The Fidils ask generalised questions with their rising pitch, without it being clear what they're asking about. The Movovs just wave their fingers around and shrug their shoulders in the hope their addressee will help them out. But at some point a Fidil-Movov bilingual superimposes the generalised-question contour from their Fidil language with the *hʁvaak* 'then' word from their Movov language. The word *hʁvaák* (with rising tone indicated by the acute accent on the second vowel) is born, with the meaning 'when', giving the world its first true information-interrogative.

This useful word passes into general use in both speech communities, though only those speaking good Movov can pronounce it properly.

Monolingual Fidil speakers manage the tone alright but can't get their mouth around the initial cluster and simplify it from *hʙvaák* to *faák*. Later on a Fidil speaker creatively combines *faák* 'when' with a point, suppressing the 'when' meaning to give a general information-interrogative meaning to English *wh-*. The point supplies the ontology of spatial location, and this hybrid speech-plus-sound lexeme comes to mean 'who' or 'which'. One night, at an intertribal wedding feast, the bilingual Fidil-Movov bride makes the customary rudimentary speech asking who will marry her brother, and when. She uses both *hʙvaák* and *faák* in her address. In the dark, her Movov family fail to see the pointing gesture she makes, and some of them don't know Fidil anyway, but from the context they guess that *fáak* must be the Fidil word for 'who' and, just like English speakers dropping tone off borrowed Chinese words like *chop-suey* or *fengshui*, they adopt the sounds without their gestural counterparts.

Now Movov has a system that not only links deictic words to questions, but makes ontological distinctions as well: *hʙv-* 'time' and *f-* 'place' combined with *-aák* 'wh-?' Another change has subtly come into Movov with this Fidil borrowing: until now Movovs have had a *v* sound but no *f*, while the Fidils have had an *f* sound but no *v*. But now the Movovs have both—*vaakvaak* for 'crow' and *faák* for 'what'. Thanks to this Fidil borrowing, Movov has become the first language to use contrastive voicing in its sound system. Hundreds of thousands of years later, middle English recapitulated this sound-split, when its non-phonemic alternations between *f* and *v* (*fox/vixen*, *half/halve*) began to be reanalysed as contrastive thanks to the flood of contrasting *f-* and *v*-words from French.

The Fidils, as a result of genetic mutations that are unevenly distributed through the small human population at the time, have a much higher rate of the Microcephalin allele than their neighbours. As would be discovered hundreds of thousands years later (Dediu and Ladd 2007), this produces a higher proportion of individuals in their population who have sensitive, accurate pitch perception. Fidil speakers had for some time been depicting actions using ideophones—imitative, onomatopoeic event-depictors—like *bong!* for the action of splitting a stone core. Later, they begin using the high- versus low-pitch distinctions they could manage so easily, to symbolise the difference between large, coarse actions and small, fine actions—*bòng*, with a low tone, for the first split, *bóng* for later flaking.[25] And eventually this same contrast was coerced to combining with the primitive *o* sound they sometimes used to get attention when pointing: since objects further away look smaller, they would accompany points to nearby objects with a low-toned *ò*, and to distant ones with a high-toned *ó*. Another modality-transfer breakthrough had occurred: the first time that deictic location had been encoded in the vocal medium.

Fidil-speaking mothers brought this handy practice into the families they raised with Movov husbands, and the children managed to learn the tonal contrasts just fine

---

[25] This practice is still found in languages like Ewe from Ghana: cf. *pótópótó* 'sound of a small drum' (high tone), *potopoto* 'sound of a big drum' (low tone). See Ameka (2001: 30).

even if they didn't all have the ideal genetic background for it—just as any child can learn a tone language today. Question–answer pairs of the form *faák? ò!* 'Where? Here!' or *faák? ó* 'Where? There!' began to appear commonly, and one day—in the sort of empathetic move people sometimes make as they imitate their conversation-partner while awaiting their turn, and primed by the 'place' associations of *f-* that had developed some time ago—the sequence became *faák? fò*! 'Where? Here!' with the *f-* corresponding roughly to the *-ere* in English. And now, on top of the well-established *hʁváak* 'when' versus *fáak* 'where' contrast there was a *faák* 'where' versus *fòo* 'here' contrast. Both the *f-* and the *-aak* elements are now part of contrast pairs, and compositional morphology is launched.

Meanwhile, the gradual accretion of new words has been slowly introducing 'duality of patterning' into the languages. In the earliest forms of Movov, the only time you heard *k* was in imitation-based crow-names, *vaakvaak*, and the only time you heard *ʁ* was in imitation-based penguin-names, *haʁ*. But by now *-k* is cropping up in the words for 'when' and 'where' as well, and *ʁ* in the word for 'when'. What's more, *ʁ*, originally confined to word-final position, can now occur in the opening cluster of a word: *hʁváak*. Young children sometimes simplify this to *ʁáak* and in one dialect of Movov this becomes the normal form. By these means speech sounds are prised away from their original semantic associations, at the same time acquiring greater combinatoric freedom. Two of the most fundamental design features of modern language—compositionality, and duality of patterning—have begun to emerge.

The above parable, while fanciful, is based entirely on incremental steps, each adaptive. Moreover, each step replicates a process or change for which analogues can be found in the history of modern languages. As such, it meets the main requirements for a gradualist, adaptationist account of language evolution made up of small changes which each produce a functionally superior system while being compatible with what has evolved so far. (For reasons of space I obviously could not tackle every possible design element of language, but a longer parable in the same vein could do this.)

Multilingualism has been a crucial part of our story in three main ways.

First, it distributes the task of solving a large number of distinct communicative problems across different populations. Given the tiny populations that we can assume spoke any one language in the earliest stage of language, this maximises the likelihood of individual 'inventions' across the whole human population at any one time, rather than hobbling them into a single small group.

Second, it takes advantage of the special creative dynamics that arise when two systems interact, as we illustrated with numerous examples in "Multilingualism, innovation transfer, and complexification" section. Sometimes this simply involves complexification, in the sense of new contrasts (extra consonant types in Kannada or Xhosa, extra noun classes in Kune Dulerayek) but sometimes what emerges makes a quantum leap from what was available in either system beforehand, as in the evolution of the Greek alphabet and Japanese syllabic script discussed above.

Third, it allows, very naturally, for affordances in some populations to get over the innovation hump more easily than others. Different human populations—even small ones—had different genetic distributions and different geographical settings. Evidence is beginning to accumulate that anatomical and genetic features relevant to

speech are not evenly distributed across human populations (Dediu and Ladd 2007). In populations with higher levels of Microcephalin, tone would have evolved more easily, and click phonemes would have evolved more easily in populations lacking a prominent alveolar ridge (Moisik and Dediu 2017). Now the evolution of structural features in language often involves recursive selection for emergent structures, over hundreds of generations, through the double bottlenecks of processability and learnability, selecting for certain structures over others (Christiansen and Chater 2016). Quite small differences in selection bias between different populations can be amplified in this process, shaping the likelihood of certain structures evolving in certain populations.

But evolving a cultural structure from scratch is much harder than adopting it from others—consider the rapid adoption of parliamentary democracies within a couple of decades in countries like South Korea and Samoa, as opposed to the many centuries required to evolve the institution in the places that first gave birth to it. So having a bias that makes it more likely for a structure to gradually emerge in one human population does not preclude it from being rapidly adopted, and learned by children, once it has evolved an efficient form. In this way, our multilingual-crucible model sits very naturally with coevolutionary models for the emergence of language diversity (Evans 2016).

The model takes an agnostic position about when the assemblage of linguistic tools reached a point compatible with whatever we define 'modern language' to be. Clearly it is compatible with a scenario where all the elements are assembled in Africa before the first humans talk their way out of the mother continent—there would have been enough generations, and enough distinct groups, in early human Africa to engender all the steps put together here. But it is also compatible with a scenario where key innovations are added to the suite by groups having left Africa, provided the ideational supply line does not get stretched to breaking point. There is enough evidence for ongoing human contacts across the Straits of Gibraltar, between Africa and the eastern Mediterranean, and across the Bab-el-Mandeb at the southern end of the Red Sea—not to mention more recent linguistic intercourse across the Afroasiatic family—that leaving Africa should not be conceived as a definitive break in communication. And to the extent that other hominin lineages, such as Neanderthals, are brought into our models of early linguistic evolution (Dediu and Levinson 2013b), their largely or exclusively non-African contributions to the evolution of language must be integrated.

Conclusions

The arguments I have put forward here will, I hope, demonstrate the plausibility of hitching a gradualist account of language evolution to a scenario which distributes the cumulative ratchetting up of the linguistic toolset across a number of distinct early populations. Their expressive inventions would have been regularly exchanged through the medium of multilingual individuals to form new systems whose growing sophistication directly results from the recombination of elements originating in different original systems.

The existence of these multilingual vectors of structural diffusion was a natural consequence of small group-size in early human populations, coupled with out-marriage and spurred on by the cultivation of linguistic difference for group-signalling purposes. Our model is demographically realistic, in the sense of being compatible with what we know about the linguistic portfolios of hunter-gatherer and other small human groups, and also fits with what we know about the elaboratory effects of multilingually mediated language contact. It has the added advantage of allowing different groups to bring, to the total set of communicative problems that humans needed to solve, their own specific affordances, some based on differences of biology, ecology, or social structure.

This gives us, almost for free, an account of why human languages are so diverse at the same time as exhibiting broadly comparable levels of sophistication: although advantageous adaptations travelled fast across the multilingual mesh, different and independent solutions to the same problem, in distinct populations, sometimes blocked their advance.

For other communicative inventions we can still see the effects of this deep evolutionary history: the discovery that word order could be harnessed to the task of showing who did what to whom (John kissed Mary ≠ Mary kissed John) may have been developed in different ways in different parts of the world—*John Mary kissed* is the worldwide commonest order for John-the-kisser scenario, but is rare in Europe. And it appears never to have reached the Australian continent, where speakers have developed other means for dealing with the problem: case-tagging that discriminate the agent from the patient, whatever order they appear in (e.g. Warlpiri), or complex agreement on the verb in a language like Ilgar (leaving *John he.her. kissed Mary* and *Mary he.her.kissed John* as synonyms, both differing from *John she.him.kissed Mary*).

Speculative reconstructions of the past—as all models of human language must be—are, sadly, difficult to evaluate according to the highest standards of falsification. But the considerations advanced here, I hope, at least establish the model of gradualist, multi-sited, multi-sourced language evolution as possible and, looking at the evidence from our best simulacra of early humans in terms of their demography, even as plausible in terms of their levels of multilingualism and the small size of their groups. The model also makes predictions which it should be possible to test through better analysis of existing data sets. First, if the same forces are at work today, i.e. if exogamy and bilingualism increase rates of change (and/or complexification), then regions with high levels of exogamy and bilingualism should be more diverse with more disparate languages. Informally, regions such as New Guinea and Amazonia appear to bear this out. Secondly, widespread multilingualism should increase the rates of language change, in particular the rate at which new typological features appear. These predictions need to be tested by integrating matched linguistic and ethnographic data, though at present it is not clear how data is on the incidence of multilingualism across small-scale populations.

The alternative—in which a single group develops all the elements of the human package in pure and splendid monolingual isolation—is of course conceivable, and has probably been, at least implicitly, the most widely assumed model in discussions of human evolution. In that sense the multilingual-crucible model is not forced

upon us. But if we consider the statistics of how likely innovations are to occur in populations of different sizes, then the tiny size of any early human group makes it much less likely that they would, on their own, develop all the elements that must be combined to make a modern language than if the full population of human would-be-communicators was put on the job, gradually pooling their inventions through multilingual exchange. I hope that the scenario I have assembled here, with its mixture of induction from known cases and speculative parable, can be tested in the coming years by modelling that simulates the main assumptions of the multilingualist hypothesis and its alternatives.

References

Ameka Felix (2001) Ideophones and the nature of the adjective word class in Ewe. In: Voeltz EFK, Kilian-Hatz C (eds) Ideophones. John Benjamins, Amsterdam, pp 25–48

Bakker P (1997) A language of our own: the genesis of Michif, the mixed Cree-French language of the Canadian Métis. Oxford University Press, New York

Bowern C (2016) Chirila: contemporary and historical resources for the indigenous languages of Australia. Lang Doc Conserv 10:1–44

Christiansen MH, Chater N (2016) Creating language: integrating evolution, acquisition and processing. MIT Press, Cambridge

Croft W (2000) Explaining language change: an evolutionary approach. Longman, London

Dediu D (2011) Are languages really independent from genes? If not, what would a genetic bias affecting language diversity look like? Hum Biol 83(2):279–296

Dediu D, Ladd LR (2007) Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, *ASPM* and *Microcephalin*. Proc Natl Acad Sci USA 104:10944–10949

Dediu D, Levinson SC (2013a) The interplay of genetic and cultural factors in ongoing language evolution. In: Richerson PJ, Christiansen M (eds) Cultural evolution: society, technology, language, and religion. Strüngmann forum reports, vol 12. MIT Press, Cambridge, pp 219–232

Dediu D, Levinson SC (2013b) On the antiquity of language: the reinterpretation of Neanderthal linguistic capacities and its consequences. Front Lang Sci 4:397. https://doi.org/10.3389/fpsyg.2013.00397

Dor D (2015) The instruction of imagination. Oxford University Press, Oxford

Dunbar RIM (1992) Neocortex size as a constraint on group size in primates. J Hum Evol 22(6):469–493. https://doi.org/10.1016/0047-2484(92)90081-J

Evans N (2003a) Context, culture and structuration in the languages of Australia. Annu Rev Anthropol 32:13–40

Evans N (2003b) Bininj Gun-wok: a pan-dialectal grammar of Mayali, Kunwinjku and Kune. Pacific Linguistics, Canberra

Evans N (2010) Dying words: endangered languages and what they have to tell us. Wiley-Blackwell, Maldon and Oxford

Evans N (2011) A tale of many tongues: documenting polyglot narrative in North Australian oral traditions. In: Baker B, Mushin I, Harvey M, Gardner R (eds) Indigenous language and social identity. Papers in honour of Michael Walsh. Pacific Linguistics, Canberra, pp 291–314

Evans N (2012a) Even more diverse than we thought: the multiplicity of Trans-Fly languages. In: Evans N, Klamer M (eds) Melanesian languages on the edge of Asia: challenges for the 21st century. Language documentation and conservation special publication no. 5, pp 109–149

Evans N (2012b) Nen assentives and the problem of dyadic parallelisms. In: Schalley AC (ed) Practical theories and empirical practice. Facets of a complex interaction. John Benjamins, Amsterdam, pp 159–183

Evans N (2016) Typology and coevolutionary linguistics. Linguist Typol 20(3):505–520

Evans N, Kashima E, Ellison M (2018) Grammaticalisation trajectories in a multilingual speech community: the decategorialising prominence marker in Nen and Nmbo. Paper presented at NWAV-AP5 conference, Brisbane

François A (2012) The dynamics of linguistic diversity. Egalitarian multilingualism and power imbalance among northern Vanuatu languages. Int J Sociol Lang 214:85–110

Green R (2004) Gurr-goni, a minority language in a multilingual community: surviving into the 21st century. In: Blythe J, McKenna-Brown R (eds) Proceedings of the seventh FEL conference, Broome, Western Australia, 22–24 September 2003. Foundation for Endangered Languages, Bath, pp 127–134

Hale K (1973) Deep-surface canonical disparities in relation to analysis and change: an Australian example. In: Sebeok TA (ed) Current trends in linguistics 87: linguistics in Oceania. Mouton, The Hague, pp 401–458

Hauser M, Chomsky N, Fitch T (2002) The faculty of language: what is it, who has it, and how did it evolve? Science 298:1569–1579

Hockett CF (1963) The problem of universals in language. In: Greenberg JH (ed) Universals of language. MIT Press, Cambridge, pp 1–29

Jackson J (1983) The fish people: linguistic exogamy and Tukanoan identity in northwest Amazonia. Cambridge University Press, New York

Laycock D (1982) Linguistic diversity in Melanesia: a tentative explanation. In: Carle R, Heinschke M, Pink P et al (eds) Gava': studies in Austronesian languages and cultures dedicated to Hans Kähler. Reimer, Berlin, pp 31–37

Leenhardt M (1946) Langues et dialectes de l'Austro-Mélanésie. Institut d'Ethnologie, Paris

Levinson SC (2006) On the human "interaction engine". In: Levinson SC, Enfield NJ (eds) Roots of human sociality: culture, cognition and interaction. Berg, Oxford and New York, pp 39–69

Lupyan G, Dale R (2010) Language structure is partly determined by social structure. PLoS ONE. https://doi.org/10.1371/journal.pone.0008559

Marley A (2018) A nganabbarru or an anabbarru? A sample of phonological variation in Bininj Kun-wok (Australia). Paper presented at NWAV-AP5 conference, Brisbane

McConvell P (1985) The origin of subsections in northern Australia. Oceania 56:1–33

McConvell P (2018) The birds and the bees. Origins of sections in Queensland. In: McConvell P, Kelly P, Lacrampe S (eds) Kin, skin & clan. ANU Press, Canberra, pp 219–270

McKnight D (1999) People, countries and the rainbow serpent. Oxford University Press, Oxford

Meakins F (2011) Case marking in contact: the development and function of case morphology in Gurindji Kriol. John Benjamins, Amsterdam

Merlan F (1981) Land, language and social identity in Aboriginal Australia. Mankind 13(2):133–148

Meyerhoff M (2017) Writing a linguistic symphony: analyzing variation while doing language documentation. Can J Linguist 62(4):525–549

Moisik SR, Dediu D (2017) Anatomical biasing and clicks: evidence from biomechanical modelling. J Lang Evol 2(1):37–51. https://doi.org/10.1093/jole/lzx004

Moore LC (2004) Second language acquisition and use in the Mandara Mountains. In: Echu G, Gyasi Oben S (eds) Africa meets Europe: language contact in West Africa. Nova Science, New York, pp 131–148

O'Shannessy C (2012) The role of code-switched input to children in the origin of a new mixed language. Linguistics 50(2):305–340

O'Shannessy C (2016) Entrenchment of Light Warlpiri morphology. In: Meakins F, O'Shannessy C (eds) Loss and renewal: Australian languages since colonisation. De Gruyter, Berlin, pp 217–251

Pascoe B (2014) Dark emu. Magabala Books, Broome

Roberts SG (2013) An evolutionary approach to bilingualism. Dissertation, University of Edinburgh

Roberts SG, Thompson B, Smith K (2014) Social interaction influences the evolution of cognitive biases for language. In: Cartmill ES, Roberts SG, Lyn H, Cornish H (eds) The evolution of language. Proceedings of the 10th international conference, pp 278–285 http://pubman.mpdl.mpg.de/pubman/item/escidoc:2021642/component/escidoc:2021650/EvoLangRobertsThompsonSmith.pdf

Rumsey A (1993) Language and territoriality in Aboriginal Australia. In: Walsh M, Yallop C (eds) Language and culture in Aboriginal Australia. Aboriginal Studies Press, Canberra

Schokkin D (2018) Towards a sociogrammar of Idi. Variation in word-final nasals. Paper presented at NWAV-AP5 conference, Brisbane

Sjödin P, Sjöstrand E, Jakobsson M, Blum MG (2012) Resequencing data provide no evidence for a human bottleneck in Africa during the penultimate glacial period. Mol Biol Evol 29(7):1851–1860. https://doi.org/10.1093/molbev/mss061

Sutton P (1997) Materialism, sacred myth and pluralism: competing theories of the origin of Australian languages. In: Merlan F, Morton J, Rumsey A (eds) Scholar and sceptic: Australian Aboriginal studies in honour of L. R. Hiatt. Aboriginal Studies Press, Canberra, pp 211–242, 297–309

Thomason S (2000) Linguistic areas and language history. In: Gilbers D, Nerbonne J, Schaeken J (eds) Languages in contact. Rodopi, Amsterdam, pp 311–327

Tomasello M (2008) Origins of human communication. MIT Press, Cambridge

White NG (1997) Genes, languages and landscapes in Australia. In: McConvell P, Evans N (eds) Archaeology and linguistics: Aboriginal Australia in global perspective. Oxford University Press, Melbourne, pp 45–81