The Dual Neurobiological Systems (DNS) framework places the neurobiological and evolutionary origins of language center-stage, and views the communicative and combinatorial capacities of the modern human as a dynamic coalition of two intersecting but evolutionarily and functionally distinguishable sets of systems. Strong evolutionary continuity between humans and their primate relatives is provided by a distributed, bi-hemispheric set of capacities that support the dynamic interpretation of multi-modal sensory inputs, in the context of social communication between members of the same species. Here we use this set of capacities to derive a neurobiologically constrained approach to the evolution of speech-based communication in the modern human lineage. A key challenge for such an approach is to identify the neurocognitive ancestral state from which the modern dual systems framework emerged.

Addresses
1 Centre for Speech, Language, and Brain, Department of Psychology, University of Cambridge, UK
2 Department of Psychology, University of Cambridge, UK

Corresponding author: Marslen-Wilson, William D (wdm10@cam.ac.uk)

A neurobiological framework
It is a foundational assumption across the neurosciences that there is a deep and explanatory continuity between the structural and functional architecture of the major cortical and subcortical systems seen in nonhuman primates and those that underpin the corresponding systems in the modern human. A critical question for human language and communication is to determine how far the neurobiological systems supporting these capacities can also be traced back to systems shared with our primate ancestors. Historically, in the context of linguistically dominated approaches to language that emphasize the grammatical and combinatorial aspects of language function, these neurobiological linkages have not been systematically pursued. Indeed, in the biolinguistic framework (e.g., [1]) the causal evolutionary event driving the emergence of human language is claimed to be the ‘saltatory’ appearance in the human lineage, perhaps 100,000 years ago, of the neurocomputational capacity for ‘recursion’. This key mutation is seen as bestowing unique combinatorial capacities on the human species, and to be fundamentally and formally distinct from the multiple roles of language as a means of communication between species members.

Here we advocate a very different approach to human language and communication, based on a Dual Neurobiological Systems (DNS) framework [2,3]. This places the neurobiological and evolutionary origins of language center-stage, and sees the communicative and combinatorial capacities of the modern human as reflecting a dynamic coalition of two intersecting but evolutionarily and functionally distinguishable sets of systems. A strong evolutionary continuity between humans and their primate relatives is provided by a distributed, bi-hemispheric set of capacities that support the dynamic interpretation of multi-modal sensory inputs, most relevantly in the context of social communication between members of the same species. These systems are the object of intensive study in monkeys and apes, and the DNS approach claims that this basic architecture underpins comparable communicative functions in the human.

At the same time, the bi-hemispheric, whole-brain systems inherited from our primate ancestors are by themselves not sufficient to support modern human language. Primate communication systems—even those of our closest relatives among the great apes – are not remotely comparable to human language in their expressive capacities [4]. While there are many factors feeding into this – capacity for symbolic representation, neural substrates for speech motor control, cooperative communicative intent, and so forth – one important area of divergence indeed lies in the domain of combinatorial grammatical function. Human language is much more than a set of signs that stand for things. It constitutes a powerful and flexible set of grammatical devices for organizing the flow of linguistic information and its interpretation, allowing us to represent and combine abstract linguistic elements (words and morphemes) which not only convey meaning but also provide the structural cues that indicate how these elements are linked together to drive sentence-level interpretation.
Neurobiological Systems approach. This left hemisphere system, furthermore, depends on neuroanatomical developments that are significantly more extensive in the human than in even our closest primate relatives. There are substantial increases in the size and complexity of the frontal and temporal areas critical for language function, coupled with the emergence of major white matter tracts connecting these areas [5]. While the functional consequences of these changes are not yet fully understood, explaining how they provide the neural substrate on which certain grammatical functions depend, it is clear that their presence is essential. When the left hemisphere system is damaged, the parallel right hemisphere regions cannot fully take over (e.g., [5,7]).

A critical claim of the DNS approach, however, is that the emerging left hemisphere system neither replaces nor displaces the bi-hemispheric system for social communication and action, so that we are dealing in the modern human with two evolutionarily distinct biological systems, however well integrated they may be in ordinary language use. Direct evidence for this comes from the functional separability of the two systems. Using a combination of behavioural and neuroimaging techniques, it is possible to demonstrate this separability both in unimpaired young adults (e.g., [8]) and in patients with left hemisphere brain damage [6–10]. These and other studies provide strong evidence for the global separability of a morphosyntactically-focused left-lateralized system and a bilateral lexically and semantically-oriented system that links into broader capacities for social cognition and pragmatic interpretation [3].

**Implications for human language and its evolutionary trajectory**

The two systems approach, and the separability of these two systems, brings into sharp focus the need to build a neuroscientifically interpretable account of what is ‘ancestral’ and what is ‘human-specific’ in modern language. The *ancestral state* becomes critical here. It is essential to understand the neurobiological context from which modern human language emerged if we are to understand how it emerged, and in what respects it is qualitatively different from preceding modes of communication. This applies both to the ancestral state as it is reflected in the modern human and to the historical states from which it evolved. Evidence for the neurobiological properties of these ancestral states is, however, severely limited where conventional sources are concerned. The fossil record and archaeological traces of material culture are not only sparse but intrinsically limited in their ability to support inferences about the cognitive capacities they imply, let alone the neural systems that support these capacities.

The DNS framework, by virtue of its emphasis on the continuity of the neurobiological systems underpinning the evolutionary trajectory from shared primate ancestors to the modern human, can provide a qualitatively richer and more constrained framework for interpreting the evidence provided by the historical paleoanthropological record – and for the new kinds of evidence streaming in from paleogenetics. Inferences about the neurocognitive ancestral state at a given point in human evolution can be constrained both by its evolutionary past and by its evolutionary future. Working backwards and forwards from living neurobiological systems whose properties we can directly investigate – on the one hand the modern human, and on the other a range of living nonhuman primate species – we can develop a better defined and more constrained toolkit of explanatory constructs in terms of which to develop a neuroscientific account of human evolution.

Where human language is concerned, we will focus on the bi-hemispheric system hypothesized for the modern human, and on the implications of this for more distant evolutionary ancestral states. In doing so, we take into account evidence from both the modern human and from our nonhuman primate (NHP) relatives, ranging from our closest relatives among the great apes – where the Last Common Ancestor (LCA) is estimated at 7–8 million years ago (mya) – to more distant but much more intensively studied species such as the macaque (LCA estimated at 25–30 mya). As noted above, a broad range of neurobiological processes, already well established in the macaque and other primates to support their ecological niche as intelligent social primates, form the primary neural substrate for parallel functions in the modern human. In the context of language comprehension, these bi-hemispheric systems support not only the ability to identify words in the speech input – integrating visual and auditory cues in face-to-face interaction – but also provide the basis for our ability to make sense of these word-meanings in the general context of the listener’s understanding of the world and of the specific context of speaking, allowing rapid and appropriate action in complex social and physical environments.

One example of this is the extension into the auditory perceptual domain, by Rauschecker and colleagues [12,13], of the dorsal/ventral distinction already established for visual perception. This distinction, developed in the macaque in terms of ‘What’ and ‘Where’ auditory pathways, strongly informs current thinking about human processing of complex auditory inputs, and of speech and language in particular [14,15,16,17]. Another important domain, directly relevant for conspecific social communication, relates to the close parallels in the human brain to the neural systems that support the integration of auditory and visual signals in the perception (and production) of NHP communicative signals (e.g., [18,19]). The presence and the critical role of these ancestral systems in modern human speech comprehension is illustrated by the patterns of preserved function following damage to the LH.
perisylvian system and by the selective patterns of activation seen in neuroimaging studies with intact individuals. The basic process of relating incoming speech sounds to representations of lexical form and content is primarily supported by bilateral temporal lobe systems. In a language like English, it is only forms that are grammatically complex – such as regular past tense forms like *jumped or waited* – which obligatorily engage left fronto-temporal systems in the intact listener and whose processing is disrupted in LH patients [8**,10,20].

A second (and related) domain of substantial overlap between neural systems and cognitive functions in humans and NHPs concerns the social context in which – necessarily – conspecific communications will be interpreted. It has long been known that the incremental interpretation of spoken utterances involves an immediate embedding in the pragmatic context of speaking (e.g., [21**22]). While the neural basis for pragmatic inference in humans is not fully resolved (for review see [23]), there is consistent evidence for activation of brain areas involved in mentalizing and Theory of Mind (ToM), such as medial prefrontal cortex (mPFC) and right temporoparietal junction (TPJ), which fall well outside the classic LH regions supporting morphosyntax. At the same time there has been increasing recognition of the complexity of social communication in large NHP social communities, and the dynamic role of pragmatic inference in the interpretation of vocal calls in these complex contexts (e.g., [24]). Coupled with this is an on-going revision of the conventional view that NHP vocal calls are essentially fixed affective reflexes, claiming instead that both monkey and ape species show sophisticated modulation of their vocalisations as a function of the pragmatic and informational context of speaking, with particular sensitivity to the dominance status of speakers and addressees (see [25,26]).

The evolutionary implications of these parallels between the interpretative contexts for conspecific communications in humans and NHPs are greatly strengthened by striking new findings [27**], which fall well outside the classic LH regions supporting morphosyntax. This is a whole-brain fMRI study where awake macaques viewed videos of interaction between conspecifics, activating a network centered in medial and prefrontal cortex that is exclusively engaged in social interaction analysis, and that significantly overlaps with human ToM networks. While this does not require us to assume that macaques ‘mentalize’ the actions of their conspecifics in the same way as humans, it suggests a deep evolutionary continuity in the neural systems that underpin these human capacities, thereby considerably enriching the properties attributable to more recent ancestral states.

A final and more controversial domain is the capacity of the modern human bihemispheric system – and, by extension, potential evolutionary ancestral states – for the sequential interpretation of spoken inputs. The clearest evidence here comes from neuropsychological and neuroimaging research involving patients with LH fronto-temporal damage (e.g., [6,11**]). This damage profoundly disrupts patients’ syntactic processing capacities, with no evidence that recruitment of right hemisphere activity can compensate [7]. Critically, however, the amount of left hemisphere damage, and the extent to which it involved key LH fronto-temporal circuits, does not affect these patients’ ability to identify the words being spoken or to understand the messages being communicated – so long as syntactic cues were not required to do so. Disruption of LH systems supporting complex syntax does not eradicate the capacity to interpret sequences requiring only computation of linear adjacency relations, as in canonically ordered sentences like “the woman pushed the girl”. Detailed case-studies of the processing strategies of such patients reveals, furthermore, an enhanced dependence on real-time semantic and pragmatic interpretation strategies based on local relations between successive words in the spoken input [28,29]. Further studies of unimpaired populations, combining fMRI with multivariate RSA analyses [30**,31] in a natural listening paradigm, shows that simple phrases (‘I play’, ‘the smile’), engage bilateral temporal regions with no evidence of LIFG involvement, confirming that some aspects of local sequential analysis are indeed supported outside the LH fronto-temporal system.

In the evolutionary context, these basic sequence interpretation capacities have surfaced from a number of different perspectives. For instance, Dehaene et al. [32] describe separate but coordinated systems for the encoding of simple transitions between items, their grouping into chunks, and temporal ordering, all of which are widely neurally distributed across both hemispheres and observed in both humans and non-human primates. Speculations that language in the modern human was preceded by the emergence of a protolanguage (e.g., [33,34]) presupposes elementary sequencing capacities (see also [35]), while the proposals of Klein and Perdue [36] for what they call the Basic Variety – the spoken communication systems developed in untutored second language acquisition – reflect extensive cross-linguistic evidence for robust capacities for the interpretation and production of linguistic sequences without recourse to either hierarchical syntax or inflectional morphology.

Potentially direct evidence for a shared neurobiological substrate for sequence analysis in humans and nonhuman primates comes from recent cross-species studies using fMRI to examine the brain regions involved in processing the ordering relationships between auditory nonsense words in rule-based sequences [37**,38]. These studies find that key regions in human bilateral ventral frontal and opercular cortex have functional counterparts in the
monkey brain. These regions are also claimed to be associated with initial stages of human syntactic processing (c.f., [39]). This raises the possibility that certain bilateral neural systems, which play a significant role in language function in modern humans, originally evolved to support domain-general abilities involved in sequence processing, and that these sit outside the core LH perisylvian system.

In summary, as illustrated above, it is possible to characterize the bihemispheric system in the modern human in terms of salient current properties that can be related to ancestral systems with deep evolutionary roots. The first two sets of claims, concerning the infrastructure for the perceptual analysis of audio-visual conspecific signals and the powerful role of social cognition in the interpretation of such signals, seem relatively straightforward. The status of the observed capacities for sequential interpretation is less clear cut, in the sense that the bihemispheric system in the modern human cannot be said to be a pure reflection of ancestral capacities in this regard, and may incorporate co-evolutionary changes related to the emergence of the left fronto-temporal system for complex syntax.

Whose ancestral state?
The properties of the sketch that we have proposed here, outlining some major communicative functions supported by ancestral systems in the modern human, suggests that these can best be mapped back to a neurocognitive ancestral state where speech-based communication can reasonably be inferred, but where the hierarchical combinatorial functions associated with the modern human left perisylvian system may not yet be fully established. Various considerations suggest that the target state should therefore be located before the advent of the ‘Anatomically Modern Human’ (AMH) but after the hypothesized divergence of the modern homo lineage and the so-called archaic homo lineage (including Neanderthals and Denisovans) from the large-brained (1200–1300 cc) hominins (such as Homo heidelbergensis) thought to be the common ancestor of both lineages [40*,41]. This divergence is provisionally dated at approximately 650,000 years ago (650 kya).

First, it is persuasively argued (see [42]) that Neanderthals had relatively advanced cognitive capacities – reflected in the technologies and material culture now attributed to them (e.g., [43]) – which included some form of spoken communication system. These are capacities in common with both modern and archaic lineages. Consistent with this, the heidelbergensis common ancestor is thought to exhibit the enlargements in the thoracic spinal ganglia controlling breathing that are necessary for speech-related breath control. Second, recent evidence from paleoepigenetic analyses of ancient DNA [44] suggests that there were substantial regulatory changes in gene expression, after the separation of the modern and archaic lineages, that specifically affect the configuration of the face and larynx [45*]. These changes – the flattening of the face and the creation of the modern human right-angled structure of the larynx – optimize these structures for the creation of complexly varying sound sequences, likely to be necessary for the combinatorial phonology that is characteristic of modern human speech communication.

Since one of the diagnostic criteria for the classification of fossil skulls as anatomically modern human (AMH) is the flattening of the face and re-shaping of the jaw (relative to the more prognathous and robust Neanderthal characteristics), these regulatory changes would have to be in place by the time the AMH appears in the fossil record. This had previously been dated to around 200 kya, but new findings from Jebel Irhoud in Morocco indicate a much earlier emergence at ~300 kya [46,47]. Interestingly, the Jebel Irhoud skulls seem to have the fully flattened face of the AMH but not the globular shape of the cranium seen in more recent AMH fossils [48], suggesting that the re-shaping of face and larynx was an earlier point in this process. The implication of this is that there was substantial development in speech-based communication in the modern human lineage, subsequent to the separation from archaic homo, and preceding the emergence of the AMH at around 300 Kya.

Provisionally, therefore, we aim to reconstruct an ancestral state falling towards the end of the period 650–300 Kya, at a relatively late stage in the evolutionary trajectory to the anatomically and behaviorally modern human. The major challenge for this reconstruction will be to consider how far these early modern homo individuals had in place the multiple elements underpinning language function – symbolic representation, hierarchical compositionality, cooperative communicative intent, and fine-grained speech motor control. If the paleoepigenetic inferences drawn by Gokhman and colleagues are correct, then the neural substrates for speech control will have emerged in tandem with the shifts in laryngeal and facial structure – though surely only in the context of across-the-board advances in communicative complexity. This in turn would indicate long-term processes of differentiation from the ancient homo (Neanderthal) lineage, with the emergence of a suite of capacities closer to the behaviorally modern human state – and bearing in mind that speech-based symbolic communication offers benefits not only for communication with others but also for the internal representation of thought.

Conflict of interest statement
Nothing declared.

Acknowledgements
The writing of this review was supported in part by a European Research Council Advanced Investigator grant to WMW (AdG 230570 NEUROLEX).
References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as

- of special interest
- of outstanding interest


9. First direct evidence from fMRI for the separability of a left fronto-temporal system supporting morphosyntax from a more general-purpose bilateral system supporting processes of perceptual mapping from sound to stored meaning.


Influenal re-statement of these authors’ multi-stream account of speech comprehension, bringing the functional architecture of the macaque brain into the mainstream of the cognitive neuroscience of human language.


28. A study of cortical networks that underpin the processing of social interactions in macaques, suggesting deep evolutionary parallels with comparable systems in humans.


32. fMRI data from healthy adults, showing that simple linear adjacencies engage bilateral temporal regions only, contrasting with selective left fronto-temporal engagement by morphosyntactically complex inflected forms.


A novel fMRI-based cross-species comparison between the systems for processing auditory rule-based sequences, showing that key regions in the human ventral frontal and opercular cortex bilaterally have functional counterparts in the monkey brain.


A recent review, by some of the main protagonists, of the revolution in our understanding of modern human origins brought about by the study of ancient (and modern) genomes.


Striking evidence from paleoepigenetic analyses of ancient DNA suggesting major regulatory changes in gene expression, after the separation of the modern and archaic lineages, that specifically affect properties of the face and larynx relevant to complex vocal output.

