Conceptual and Methodological Problems with Comparative Work on Artificial Language Learning

Jeffrey Watumull, Marc D. Hauser & Robert C. Berwick

Several theoretical proposals for the evolution of language have sparked a renewed search for comparative data on human and non-human animal computational capacities. However, conceptual confusions still hinder the field, leading to experimental evidence that fails to test for comparable human competences. Here we focus on two conceptual and methodological challenges that affect the field generally: 1) properly characterizing the computational features of the faculty of language in the narrow sense; 2) defining and probing for human language-like computations via artificial language learning experiments in non-human animals. Our intent is to be critical in the service of clarity, in what we agree is an important approach to understanding how language evolved.

1.   Introduction

Within the past several decades, starting with the synthetic reviews of Lieberman (1984), Bickerton (1990), and Pinker & Bloom (1990), there has been increasing interest and empirical study of the evolution of language (e.g., Fitch 2012, Taller-man & Gibson 2012). Nevertheless, considerable confusion remains regarding the central theoretical issues and core concepts to be engaged, leading to empirical studies that are sometimes far off the mark.

Perhaps nowhere has this confusion been greater than in reaction to the issues raised by Hauser *et al*. (2002), and this is especially the case with respect to comparative studies of artificial language learning in animals (Fitch & Hauser 2004, Gentner *et al*. 2006, Murphy *et al*. 2008, Abe & Watanabe 2011, Rey *et al*. 2012). Here we focus on two problems that have hindered work in this area, especially its potential contribution to linguistics, cognitive science, neuroscience, and evolutionary biology.

First, despite broad interest in the mechanisms underlying the capacity for language, and especially what is unique to humans and to language, studies with non-human animals are often not appropriately designed to answer questions

about these mechanisms; running artificial language learning experiments is non-trivial (Reber 1967). In particular, several studies focus too narrowly on the problem of syntactic-like embedding as *the* defining feature of our uniquely human capacity. But this approach is flawed: Embedding is neither necessary nor sufficient for a full description of human language. Furthermore, but far more peripherally, many have incorrectly suggested that Hauser *et al.*'s (2002) thesis about the evolution of language places center-embedding as a core process in human linguistic competence. Since several comparative studies of animal computation focus on this work, it is important to get it right: Hauser and colleagues specifically suggested that what is unique to humans and unique to language (the Faculty of Language in the Narrow Sense, FLN) is *recursion and its mappings to the sensory-motor and conceptual-intentional systems*.

Second, standard methodology in this research area—massive training of captive animals with reward for 'correct' behavior—bears little resemblance to experimental child language research, or to child language acquisition (Wanner & Gleitman 1982, Ambridge & Lieven 2011); studies of children explore acquisition by means of spontaneous methods, using passive exposure or habituation–discrimination. Consequently, animal researchers cannot so easily draw conclusions about either the trajectory of human language development or its computational-representational properties.

Our central aim, therefore, is to clarify these conceptual and methodological issues, and then end with a few suggestions on how empirical work in this important area might progress.

2. Testing for Uniquely Human Mechanisms of the Language Faculty

Given the broad set of factors that enter into language, empirical research is only tractable by first defining a narrow subset of core linguistic properties. This was one motivation for Hauser *et al.* (2002) to define the language faculty in the narrow sense (FLN) as "the abstract linguistic computation system alone, independent of the other systems with which it interacts and interfaces" (p. 1571) in the language faculty defined broadly (FLB). FLN comprises "the core computational mechanisms of recursion as they appear in narrow syntax and the mappings to the interfaces with [conceptual-intentional and sensory-motor systems]" (p. 1573). The FLN/FLB distinction leaves open, of course, what factors are part of FLN and FLB, while establishing methodological criteria for future investigation.

The FLN/FLB distinction was also developed as a conceptual guide. FLN characterizes linguistic *competence* in the form of recursive (computable) functions that generate a discrete infinity of structured expressions, formally analogous to the procedure for the inductive generation of sets and so the natural numbers. The set of linguistic expressions and the set of natural numbers are thus *effectively computable* in that, though infinite, they are "calculable by finite means" (Turing 1936: 230). For example, a function—a finite representation—can be specified to generate the infinite, nonrandom decimal expansion of π. Because this expansion is infinite, it cannot be physically represented as such. However this is an

independent—and arbitrary—fact of the *performance* mechanisms that implement the finite function; π does not cease to be a computable number if the physical resources required to calculate it are exhausted, or even nonexistent. It is in this same sense that FLN is a competence system—a system of recursively generated discrete infinity—logically isolable from the performance systems with which it interfaces to form FLB.

FLN qua recursive function is thus typified by three essential properties (see Watumull *et al*. 2014 for further discussion): (i) *computability*, (ii) *definition by induction*, and (iii) *mathematical induction*. Computability is reflected in a procedure—equivalent to a type of *Turing machine*, discussed below—that generates new and complex representations by combining and manipulating discrete symbols. The computable function must be defined by a sophisticated form of induction: Outputs must be carried forward and returned as inputs to generate a hierarchical structure over which can be defined complex relations (e.g., syntactic, semantic, phonological, etc.). In technical terms, the function *strongly generates* structures corresponding to *weakly generated* strings (e.g., the weakly generated string *the boy saw the man with binoculars* is one string with (at least) two syntactic structures, {{the, boy}, {saw, {the, man}}, {with, binoculars}} and {the, {boy, {saw, {the, {man, {with, binoculars}}}}}}, corresponding to (at least) two different semantic interpretations). Finally, mathematical induction is realized in the jump from finite to infinite, as in the projection from a finite set of words to an infinite set of sentences.

Given this specification of FLN, it is false to conflate recursive *functions* with center-embedded (CE) *patterns* of the form $a^n b^n$ (e.g., *the antelope* [a1] *the lion* [a2] *ate* [b2] *ran like a snail* [b1]). The most recent example of this error is by Rey *et al*. (2012) in experiments with baboons: "[T]he central claim [of Hauser *et al*. 2002 is] that the ability to process CE structures is a critical cognitive feature distinguishing human from nonhuman communication" (p. 180). Following this line of argument, 'success' by non-human animals in processing CE structures leads to the overly strong conclusion that, "[c]ontrary to the commonly accepted claim that recursion is human specific[,] CE structures produced by humans could have their origins in associative and working memory processes already present in animals" (p. 182–183).

As noted, this conclusion is problematic because it falsely equates center-embedding with recursion, and more narrowly, attributes to Hauser *et al*. (2002) the incorrect thesis that the ability to process CE patterns is what defines FLN. The correct thesis is that FLN characterizes the uniquely human character of language. To repeat, Hauser et al. proposed that FLN comprises "the core computational mechanisms of recursion as they appear in narrow syntax and the mappings to the interfaces with [conceptual-intentional and sensory-motor systems]" (p. 1573). Recursion, as noted in this hypothesis, was understood in the standard mathematical sense given above. Expressions generated by this system may be (center-)embedded or not; whether a function is recursive or not is independent of the form—or even existence—of its output. Theorems from the formal sciences (e.g., the work by Rice, Ackermann, and others) demonstrate that in general it is exceedingly difficult to infer anything truly germane as to the nature of a computational mechanism from patterns in its outputs. Consequently, test-

ing for the ability to process (center-)embedding does *not* constitute a test of the FLN claim, contrary to what is claimed by Rey *et al.* (2012) and studies on which it builds. Here we work through the Rey *et al.* study as an illustration of these problems, but note that they arise in other work as well (e.g., Gentner *et al.* 2006, Murphy *et al.* 2008, Abe & Watanabe 2011).

In the Rey *et al.* (2012) experiments, captive baboons were conditioned to associate pairs of visual shapes a_ib_i to test whether they would order selection of those shapes in a 'center-embedded' $a_ia_jb_ja_i$ pattern. Rey *et al.* summarize their results: "[B]aboons spontaneously ordered their responses in keeping with a recursive, centre-embedding structure" (p. 180). They then conclude that "the production of CE structures in baboons and humans could be the by-product of associative mechanisms and working memory constraints" (p. 183). In other words, *neither baboons nor humans are endowed with FLN*—a surprising and un-evidenced result in the case of humans. This *non sequitur* derives from the failure to distinguish associative processes from recursive computations.

Association is indeed the most parsimonious explanation of the baboon results: intensive, repetitive, conditioned associative learning that is ubiquitous in the animal kingdom, from invertebrates to vertebrates (Gallistel 1990). As Rey *et al.* observe, "the [baboon's] preference for producing CE structures requires (1) the capacity to form associations between pairs of elements (e.g., $a1b1$ or $a2b2$) and (2) the ability to segment these associations and maintain in working memory the first element of a pair ($a1$) in order to produce later its second associated element ($b1$). [T]hese two requirements are satisfied in baboons and are sufficient for producing CE structures having one-level-of-embedding" (p. 182). Two implications follow.

For Rey *et al.*, the 'language' to be recognized is strictly finite, in the form $a_ia_jb_ja_i$ for $i, j = 1, \ldots, 6$ (with i, j distinct). As such, it is unnecessary to posit any embedded structure—let alone any underlying grammar—to correctly recognize this language. Furthermore, such a result runs precisely counter to the original aim of the study: Instead of showing that baboons are endowed with a capacity that parallels the characteristic unboundedness of human language, it shows that baboons display a finite, bounded processing ability. Second, if association suffices for 'one-level-of-embedding', this in turn implies that the extension of such an ability to process two levels of embedding would demand extensive additional training (i.e. listing additional associations), a result that has been amply demonstrated as fatal in connectionist networks (Berwick 1982, Elman 1991), and is fundamentally different from human language acquisition.

Another way in which Rey *et al.* err can be seen in the fact that the linguistic patterns that are 'easy' and 'difficult' for people to process do not align well with center-embedded word sequences and their possible foils—such patterns are both too strong and too weak. As noted by Rogers & Hauser (2010), while people find language patterns in the form a^nb^n difficult to process (e.g., *people^n left^n* (e.g., *people people people left left left*)), their corresponding paraphrased forms (*people who were left* (*by people who were left*)^n *left* (e.g., *people who were left by people who were left left*)) seem easier for people to analyze; several authors, including Rey *et al.*, assume that these latter patterns are within the reach of non-human animal abilities. Notably, the processing of center-embedded structures in humans is

known to be limited by working memory, a point acknowledged by Rey *et al.*, as well as from the classic studies by Miller & Isard (1964). But memory by itself is not an ability or competence. As Rey *et al.* acknowledge, it is simply the workspace within which particular procedures are executed. Human performance in such cases can be extended indefinitely without any change to the internal 'program' (competence) if time and access to external memory are increased (Miller & Chomsky 1963); and far from being unfalsifiable (see Gentner *et al.* 2006 for such a claim), the independent existence of a particular linguistic competence can be demonstrated by varying performance as a function of computational complexity. In contrast, this effect has not been demonstrated in baboons, nor is it obvious how one would run the relevant tests.

Rey *et al.* conclude that "increasing the levels-of-embedding could be too demanding for baboons" (p. 182), and then speculate that "[a]lthough the present results indicate that baboons are not qualitatively limited in processing CE structures, their performance could be limited quantitatively to the processing of one or two embeddings" (p. 182). But this is misleading. Rey *et al.* provide no evidence to indicate that the qualitative limits do not simply reduce to quantitative limits, that is, that an unlimited competence underlies the baboons' limited performance. Finally, as Rogers & Hauser (2010) observe, center-embedded $a^n b^n$ patterns correspond to the 'simplest' possible kind of embedding structure. For example, they allow for Sentences embedded within other Sentences (e.g., *John knows that the baboon learned language)*, but not Sentences embedded within Noun Phrases, as in relative clauses (e.g., *the baboon who learned language*), let alone many other constructions in human language. In short, $a^n b^n$ patterns—the proxy for center-embedded structure—are simply not what is essential to FLN; they are not good 'human language detectors', being both too simple and too complex. This critique holds independently of the method used to demonstrate how individuals acquire such patterns, a point we explore below.

To think that human linguistic competence can be reduced to association and working memory reveals a misunderstanding of the critical difference between a look-up table—a finite list of associations—and a Turing machine—a mathematical model of computation represented by a control unit of stored instructions and an unbounded read/write memory tape enabling unbounded computation. If one takes the computational theory of the mind/brain seriously, it is the Turing machine (or one of its formal equivalents) that serves as the natural model for human cognition, including language; the look-up table is a nonstarter (see Gallistel & King 2009).

The distinction between finite and infinite memory, more specifically the independence of assumptions about working memory from those about syntactic competence, has proved fruitful for the bulk of research in human syntax during the past sixty or so years. While it is true that the human brain is finite, and so *could* be represented as a (large) finite-state machine or look-up table, this isn't relevant. The set of outputs a human can generate is in principle unlimited and, importantly, *non-arbitrary*, (i.e. the set of outputs is nonrandom, inclusion in the set being determined by the generative function). It is infinite models of these finite systems that yield scientific insight (see Turing 1954 on the generate/look-up distinction).

Consider human arithmetical competence. Here, the finite/infinite distinction seems so clearly necessary that the cognitive science literature assumes without question that this competence is somehow internalized (perhaps not transparently) in the form of some finite set of rules; it further assumes that these rules, unmodified for any particular arithmetic task, determine an infinite—and non-arbitrary—range of outputs. Here, performance may be 'truncated' by working memory, among many other factors, in recognizable ways (e.g., Hitch 1978, Dehaene 1999, Trotzke *et al*. 2013). Indeed, multiplication cannot even be carried out by a finite-state machine. What is required for multiplication is something similar to a Turing machine with a potentially unbounded input/ output tape, so that intermediate results can be written to an external tape and carried forward ('recursed') to later stages of the computation. Any purely association-based method must fail at some point. Yet no one doubts that people have internalized the rules for multiplication (operating on an internal tape). Nor is there any confusion that the same holds for any physically realizable computer, like a laptop. Unsurprisingly, in all cases, the infinite model yields the proper theory for the physically realized device.

Arithmetical competence corresponds in many important respects with linguistic competence. As observed above, both arithmetic and language are systems of digital infinity, each enumerating inductively a potentially infinite and non-arbitrary set of discretely structured objects via computable functions. As Chomsky (1959) noted, the grammar for generating a set of linguistic expressions can be characterized as a function mapping the integers onto this set. As hypothesized for FLN (Watumull 2012), the discrete elements of a syntactic expression (e.g., words) are read as input and, as instructed by internalized linguistic rules (principles and parameters, etc.), combined into sets (e.g., phrases) and written onto the memory 'tape' to be carried forward as 'intermediate results', serving as inputs to subsequent computations. This enables the unbounded combination of words into phrases, and phrases into sentences, and sentences into discourses.

The generative process just described is essentially the "iterative conception of a set", with sets of discrete objects, linguistic or arithmetic, "recursively generated at each stage" such that "the way sets are inductively generated" is formally equivalent to "the way the natural numbers […] are inductively generated" (Boolos 1971: 223). Thus both language and arithmetic draw on similar generative procedures, a point reiterated in Hauser *et al*. (2002). Though non-human animals appear to be able to carry out some arithmetical operations using analog quantity representations, or perhaps subitizing for small integers, there seems to be no evidence for anything resembling the computable rule systems sketched above or the inductive generalization to an unbounded domain of structured arithmetical expressions. Even when animals are taught the Arabic integers through reinforcement, they never acquire anything remotely like the successor function, generalizing beyond the trained input (Kawai & Matsuzawa 2000). Moreover, and of direct relevance to the methodology of most animal studies in this area including the artificial language studies discussed here, the research on animal integer processing also demonstrates that this capacity is entirely different from children's development of arithmetical competence: Animals never exhibit the kind of inductive leap (the best evidence for discrete infinity) that all children

take once they have acquired knowledge about the first few integers (Leslie *et al.* 2008, Carey 2011). What is required is some way to carry forward arbitrarily large, inductively generated intermediate results, say by means of an arbitrarily long input/output tape (mentally represented), as in the multiplication and syntactic examples described earlier.

3.   Methodology for Experiments with Non-Human Animals

Understanding how behavior is acquired is essential to comparative inquiry. It is particularly important in work on artificial language learning because children do not acquire language by means of massive, long-term training. Further, a hallmark of virtually all aspects of language acquisition is the inductive aspect of recursion: Once a particular component of linguistic competence develops, it rapidly generalizes to a virtually limitless range of possible expressions. In the case of most work on artificial language learning, whether on birds, rodents, or primates, the method entails massive training with little evidence of anything remotely resembling unbounded generalizations. The animals seem merely to be compiling a list—a look-up table—rather than internalizing rules. Thus, even if one were to grant that animals exhibit certain linguistic-like behaviors, their mode of acquisition is nothing like that evidenced by human children, and whatever has been acquired appears extremely bounded in its expressive power.

A counter-example to this approach is the original study of finite-state and phrase-structure grammars by Fitch & Hauser (2004) with cotton-top tamarins, and pursued in a slightly different way by Abe & Watanabe (2011) in Bengalese finches. Here, the method paralleled those used by researchers working on artificial language learning in human infants, and in particular, a familiarization-discrimination technique. In brief, this technique exposes subjects in a passive listening context to the relevant input, and then follows with presentations of exemplars that match the input as well as exemplars that are different in some fundamental way. If subjects have picked up on the pattern inherent in the familiarization phase, they should respond more intensely to the exemplars in the discrimination phase that are different than to those that are the same.

Though this technique captures the spontaneity of processing that is characteristic of language processing, it suffers from at least two problems. First, unlike the training techniques that involve highly objective and robust behavioral measures (e.g., touching a button), the familiarization-discrimination techniques involve a more subjective and ambiguous response: looking time or looking orientation. Despite methods designed to provide relatively high inter-observer reliabilities, these remain relatively fragile techniques, due in part to the often small differences in response measures across conditions (often a matter of a couple of seconds). Second, and more importantly, in studies of non-human animals, where the test population is extremely limited and small, it is necessary to run different conditions with the same population. This is not the case in studies of human infants where different conditions are tested on different populations. Given the limited test population, animals often habituate to the general test environment, and further, are exposed to many different conditions, thereby changing their experience over multiple conditions.

We are thus left with a spontaneous method that cannot deliver the requisite information about processing capacities that are like child language acquisition, or a training method that can potentially identify an ability, but one that may well be fundamentally different from what is in play for human children during acquisition. In other words, even if a training study shows that an animal can 'compute' center-embedded patterns, the underlying representations are likely to be entirely different because of the procedures used to demonstrate this capacity. In any event, such methods have, thus far, failed to demonstrate the unboundedness that is required of human linguistic computation.

4.   Conclusion

What results with non-human animals might challenge the claim that the language faculty is uniquely human? And more narrowly, what evidence might refute the hypothesis proposed by Hauser *et al.* (2002) regarding the composition of FLN? With respect to the generative component of their thesis, and in particular, its focus on recursive mechanisms, it would be necessary to show that animals *spontaneously* respond to stimuli that are characterized by (i) *computability*, (ii) *definition by induction*, and (iii) *mathematical induction*—the three properties typical of linguistic recursion that we briefly noted above. Computability requires proof of a procedure that generates new and complex representations by combining and manipulating symbols, as in human language; this *productive* process is to be contrasted with the *retrieval* of representations from a look-up table (finite and innately specified or memorized), as in non-human primate calls. The computable function must be defined by a sophisticated form of induction: Outputs must be carried forward and returned as inputs to *strongly generate* a hierarchical structure over which can be defined complex relations (e.g., syntactic, semantic, phonological, etc.); this also implies the discreteness of representations. Lastly, mathematical induction is seen in the jump from finite to infinite. This can be demonstrated by significant generalization beyond the exposure material (e.g., counting indefinitely beyond the training set) and by revealing an unbounded competence underlying bounded performance.

In conclusion, to advance this important field, greater conceptual and methodological clarity is necessary (for recent discussions, see Fitch & Friederici 2013, Zuidema 2013). Conceptually, it is necessary to understand the formal aspects of recursive functions in order to capture the fundamental generative and unbounded properties of all natural languages (where embedding is an interesting but incidental phenomenon). Experiments should focus on all aspects of the Turing-like architecture of the faculty of language in its narrow sense: aspects of the enumeration by finitary procedures and read/write memory of a non-arbitrary digital infinity of hierarchically structured expressions and relations. Devising such tests may prove difficult, but this is the critical challenge for a theoretically rigorous and empirically grounded approach to the evolution of language.

References

Abe, Kentaro & Dai Watanabe. 2011. Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nature Neuroscience* 14, 1067–1074.

Ambridge, Ben & Elena V. M. Lieven. 2011. *Child Language Acquisition: Contrasting Theoretical Approaches*. Cambridge: Cambridge University Press.

Berwick, Robert C. 1982. Locality principles and the acquisition of syntactic knowledge. Cambridge, MA: MIT dissertation.

Berwick, Robert C., Kazuo Okanoya, Gabriel J. L. Beckers & Johan Bolhuis. 2011. Songs to syntax: The linguistics of birdsong. *Trends in Cognitive Science* 15, 113–121.

Boolos, George. 1971. The iterative conception of set. *The Journal of Philosophy* 68, 215–231.

Carey, Susan. 2011. *The Origin of Concepts*. Oxford: Oxford University Press.

Chomsky, Noam. 1959. On certain formal properties of grammars. *Information and Control* 2, 137–167.

Dehaene, Stanislas. 1999. *The Number Sense: How the Mind Creates Mathematics*. Oxford: Oxford University Press.

Elman, Jeffrey L. 1991. Distributed representation, simple recurrent networks, and grammatical structure. *Machine Learning* 7, 195–225.

Fitch, W. Tecumseh & Angela D. Friederici. 2012. Artificial grammar learning meets formal language theory: An overview. *Philosophical Transactions of the Royal Society B*, 1933–1955.

Fitch, W. Tecumseh & Marc D. Hauser. 2004. Computational constraints on syntactic processing in a nonhuman primate. *Science* 303, 377–380.

Gallistel, C. R. 1990. *The Organization of Learning*. Cambridge, MA: MIT Press.

Gallistel, C. R. & Adam Philip King. 2009. *Memory and the Computational Brain: Why Cognitive Science Will Revolutionize Neuroscience*. New York: Wiley-Blackwell.

Gentner, Timothy Q., Kimberly M. Fenn, Daniel Margoliash & Howard C. Nusbaum. 2006. Recursive syntactic pattern learning by songbirds. *Nature* 440, 1204–1207.

Hauser, Marc D., Noam Chomsky & W. Tecumseh Fitch. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298, 1569–1579.

Hitch, Graham J. 1978. The role of short-term working memory in mental arithmetic. *Cognitive Psychology* 10, 302–323.

Kawai, Nobuyuki & Tetsuro Matsuzawa. 2000. Numerical memory span in a chimpanzee. *Nature* 403, 39–40.

Leslie, Alan M., Rochel Gelman & C. R. Gallistel. 2008. The generative basis of natural number concepts. *Trends in Cognitive Science* 12, 213–218.

Miller, George A. & Noam Chomsky. 1963. Finitary models of language users. In R. Duncan Luce, Robert R. Bush & Eugene Galanter (eds.), *Handbook of Mathematical Psychology II*, 419–491. New York: Wiley and Sons.

Miller, George A. & Stephen Isard. 1964. Free recall of self-embedded English sentences. *Information and Control* 7, 292–303.

Murphy, Robin A., Esther Mondragón & Victoria A. Murphy. Rule learning by rats. *Science* 319, 1849–1851.

Pinker, Steven & Paul Bloom. 1990. Natural language and natural selection. *Behavioral and Brain Sciences* 13, 707–784.

Reber, Arthur S. 1967. Implicit learning of artificial grammars. *Verbal Learning and Verbal Behavior* 5, 855–863.

Rey, Arnaud, Pierre Perruchet & Joël Fagot. 2012. Centre-embedding structures are a by-product of associative learning and working memory constraints: Evidence from baboons (*Papio Papio*). *Cognition* 123, 180–184.

Rogers, James & Marc D. Hauser. 2010. The use of formal language theory in studies of artificial language learning: A proposal for distinguishing the differences between human and nonhuman animal learners. In Harry van der Hulst (ed.), *Recursion and Human Language*, 213–232. Berlin: De Gruyter Mouton.

Tallerman, Maggie & Kathleen R. K. Gibson (eds.). 2012. *The Oxford Handbook of Language Evolution*. Oxford: Oxford University Press.

Trotzke, Andreas, Markus Bader & Lyn Frazier. 2013. Third factors and the performance interface in language design. *Biolinguistics* 7, 1–34.

Turing, Alan M. 1936. On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society* 42, 230–265.

Turing, Alan M. 1954. Solvable and unsolvable problems. *Science News* 31, 7–23.

Wanner, Eric & Lila R. Gleitman (eds.). 1982. *Language Acquisition: The State of the Art*. Cambridge: Cambridge University Press.

Watumull, Jeffrey. 2012. A Turing program for linguistic theory. *Biolinguistics* 6, 222–245.

Watumull, Jeffrey, Marc D. Hauser, Ian G. Roberts & Norbert Hornstein. 2014. On recursion. *Frontiers in Psychology* 4, 1–7.

Zuidema, Willem. 2013. Context-freeness revisited. *Proceedings of CogSci 2013*, http://mindmodeling.org/cogsci2013/papers/0309/paper0309.pdf.

Jeffrey Watumull
University of Cambridge
Theoretical & Applied Linguistics
Cambridge CB3 9DA
UK
jw647@cam.ac.uk

Marc D. Hauser
Risk-Eraser
PO Box 376
West Falmouth, MA 02574
USA
mdh102559@gmail.com

Robert C. Berwick
Massachusetts Institute of Technology
Electrical Engineering & Computer Science
Cambridge, MA 02139
USA
berwick@csail.mit.edu