Journal of Neurolinguistics xxx (2016) 1–17

Contents lists available at ScienceDirect

Journal of Neurolinguistics

journal homepage: www.elsevier.com/locate/jneuroling

Review

The Universal Generative Faculty: The source of our expressive power in language, mathematics, morality, and music

Marc D. Hauser [a,*], Jeffrey Watumull [b]

[a] Risk-Eraser, PO Box 376, West Falmouth, MA, United States
[b] Oceanit, 828 Fort Street Mall, Suite 600, Honolulu, HI, 96813, United States

ARTICLE INFO

Article history:
Received 1 August 2016
Accepted 15 October 2016
Available online xxx

Keywords:
Domain-specificity
Evolution
Generative functions
Language faculty
Recursion
Turing machine
Universal generative faculty

ABSTRACT

Many have argued that the expressive power of human thought comes from language. Language plays this role, so the argument goes, because its generative computations construct hierarchically structured, abstract representations, covering virtually any content and communicated in linguistic expressions. However, language is not the only domain to implement generative computations and abstract representations, and linguistic communication is not the only medium of expression. Mathematics, morality, and music are three others. These similarities are not, we argue, accidental. Rather, we suggest they derive from a common computational system that we call the *Universal Generative Faculty* or UGF. UGF is, at its core, a suite of contentless generative procedures that interface with different domains of knowledge to create contentful expressions in thought and action. The representational signatures of different domains are organized and synthesized by UGF into a global system of thought. What was once considered the *language of thought* is, on our view, the more specific operation of UGF and its interfaces to different conceptual domains. This view of the mind changes the conversation about domain-specificity, evolution, and development. On domain-specificity, we suggest that if UGF provides the generative engine for different domains of human knowledge, then the specificity of a given domain (e.g., language, mathematics, music, morality) is restricted to its repository of primitive representations and to its interfaces with UGF. Evolutionarily, some generative computations are shared with other animals (e.g., combinatorics), both for recognition-learning and generation-production, whereas others are uniquely human (e.g., recursion); in some cases, the cross-species parallels may be restricted to recognition-learning, with no observable evidence of generation-production. Further, many of the differences observed between humans and other animals, as well as among nonhuman animals, are the result of differences in the interfaces: whereas humans promiscuously traverse (consciously and unconsciously) interface conditions so as to combine and analogize concepts across many domains, nonhuman animals are far more limited, often restricted to a specific domain as well as a specific sensory modality within the domain. Developmentally, the UGF perspective may help explain why the generative powers of different domains appear at different stages of development. In particular, because UGF must interface with domain-specific representations, which develop on different time scales, the generative power of some domains may mature more slowly (e.g., mathematics) than others (e.g., language). This explanation may also contribute to a deeper understanding of cross-cultural differences among human populations, especially cases where the

* Corresponding author.
E-mail address: marc@risk-eraser.com (M.D. Hauser).

http://dx.doi.org/10.1016/j.jneuroling.2016.10.005

generative power of a domain appears absent (e.g., cultures with only a few count words). This essay provides an introduction to these ideas, including a discussion of implications and applications for evolutionary biology, human cognitive development, cross-cultural variation, and artificial intelligence.

Contents

1. Introduction

The ideas developed in this essay grow out of several different intellectual traditions within the formal and cognitive sciences. Broadly speaking, we are interested in what enables human minds to generate a limitless range of ideas and expressions across many different domains of knowledge. To what extent is this facility enabled by domain-general or domain-specific mechanisms? To what extent are these facilities shared with other organisms and to what extent are they uniquely human? To what extent are the generative mechanisms that operate in different domains of knowledge the same or different, and why? What accounts for the developmental timing and maturation of different domains of knowledge? And could the creative, generative power of human intelligence be realized in computing machinery? This essay provides an introductory sketch of an idea that, we believe, helps shed new light on these fundamental questions.

2. Different traditions of thought

One tradition that not only launched many of the questions noted above, but developed a significant position on the answers, is Chomsky's (1955; 1995) work in linguistics, and the nature of mind more generally. The argument, in brief, is that humans are endowed with a finite cognitive computational system that generates an infinity of meaningful expressions. This is a linguistic system or faculty, with unique — specific to our species and the domain of language — recursive procedures that interface with both the conceptual-intentional (semantics/pragmatics) and sensory-motor (phonology/phonetics) systems to generate hierarchically structured representations. This intensional system — I-language — is internal to an individual, and is often described as forming a language of thought. The sets of expressions this system enumerates have been described (not by Chomsky) as E-languages (e.g., English, French, Japanese, etc.).

Based on Chomsky's linguistic framework, some have argued that language enables the expressive power of all other domains, and in many cases, provides the cognitive glue across domains. Thus, for example, Spelke (2016) has argued that what enables us to integrate different domains or modules of thought, including aspects of space and number, is language. In a classic set of experiments (Hermer & Spelke, 1994; 1996) on spatial reorientation following disorientation, young children appear incapable of integrating information about landmarks with information about the geometry of the space, a result that parallels those originally carried out on rats (Cheng, 1986). Such integration only occurs when children acquire spatially-relevant words (e.g., *right of, in front of*), the linguistic glue that integrates information from the landmark and geometry systems. Moreover, the flawless performance of adults was reduced to that of young children and rats when they were required to carry out a verbal shadowing task, one that effectively blocks access to the language faculty. This perspective sets up language as both the generative machinery of thought and the system that enables interfaces across domains. Similar ideas have inspired models of artificial intelligence in which the human-like AI understands the world by using linguistic machinery to combine commonsense knowledge with perceptual (particularly visual) representations in the form of explanatory "stories" (Winston, 2012).

Other traditions recognize the generativity of language, and in some cases, acknowledge what may be its unique computations, but argue that other domains of thought and expression are also generative: a generative system being a finite computational procedure for *explicitly* enumerating (potentially infinitely many) expressions; the most interesting expressions being in some sense "meaningful". For instance, Monti and Osherson (2012) as well as Dehaene, Meyniel, Wacongne,

Please cite this article in press as: Hauser, M. D., & Watumull, J., The Universal Generative Faculty: The source of our expressive power in language, mathematics, morality, and music, Journal of Neurolinguistics (2016), http://dx.doi.org/10.1016/j.jneuroling.2016.10.005

Wang, and Pallier (2015) have noted the parallel generativity of both language and mathematics, while also recognizing that each domain deploys both similar and different generative operations. For example, mathematics makes use of the successor function, a recursive operation that generates the integer list, and ultimately is what enables our understanding of cardinality and ordinality (Leslie, Gelman, & Gallistel, 2008). As discussed next, the development of the integer list is a beautiful example of the necessity of interfaces across domains or at least more domain-general systems of computation and representation.

Before young children know what count words mean, they rapidly develop a memorized count list. Ask any young child between the ages of two and three to count, and they will rattle off the list. But ask a two-year-old to give you *only* four cookies out of a pile of ten, and they will just grab *some* cookies, usually more than one. Understanding what each count word means develops slowly over the first few years of life. Children first understand what "one" means, then "two", then "three", and then, by recursive induction and the integration of the count list with the successor function, they understand the meaning of all other count words (Carey, 2009; see Yang, 2016 for a different explanation of the timing of this process). Thus, one interpretation of this finding is that although the successor function as a generative computation is in place, presumably from birth, it is not integrated or interfaced with the conceptual and linguistic systems until much later in life (Leslie et al., 2008). Interestingly, this integration in the number domain occurs after children have integrated other recursive procedures with their lexicon to generate linguistic expressions (e.g., Crain & Thornton, 2000; Yang, 2013). And in some cultures (e.g., the Mundurucu from the Brazilian Amazon), the integration may by limited as the lexicon only includes words for the first few integers and then describes all other, larger numbers, as "many" (Hale, 1975; Izard, Pica, Spelke, & Dehaene, 2008; Pica, Lemer, Izard, & Dehaene, 2004).

The idea that other domains of knowledge may be equipped with generative computations and abstract representations that are non-linguistic is further supported by recent imaging work in the domain of mathematics (e.g., Monti, Parsons, & Osherson, 2012; Amalric & Dehaene, 2016b). For example, when professional mathematicians evaluated statements in geometry, topology, or algebra, they activated areas unrelated to language and general knowledge semantics (Amalric & Dehaene, 2016b). Interestingly, however, there was significant activation in the Inferior Frontal Gyrus (IFG), an area that is consistently activated in imaging studies involving structural processing of language (Ding, Melloni, Zhang, Tian, & Poeppel, 2015; Pallier, Devauchelle, & Dehaene, 2011; Udden & Bahlmann, 2012), as well as music, morality, theory of mind, and tool technology (Fitch & Martins, 2014; Fitch, 2014; Friederich & Friederici, 2009; Hamzei et al., 2016; Hecht et al., 2014; Koelsch, Rohrmeier, Torrecuso, & Jentschke, 2013; Maess, Koelsch, Gunter, & Friederici, 2001). Neuropsychological and transcranial magnetic stimulation studies also suggest that when the IFG is damaged or temporarily silenced, there are significant consequences for both processing and producing richly structured sequences in these different domains (Anderson, Bechara, Damasio, & Tranel, 1999; Goldenberg, Hermsdorfer, Glindemann, Rorden, & Karnath, 2007; Monte et al., 2014; Young & Dungan, 2012).

It is because of these parallels across domains — effectively, shared resources — that we propose here the idea of a *Universal Generative Faculty* (UGF), a system of generative functions that link to all domains of human thought, and are available for both recognition-learning as well as generation-production. By fluidly combining and analogizing across different domain-specific representations, UGF creates a global system of thought — our reframing of the classic *language of thought*. In some sense, the kernels of the ideas developed here can be seen in Lashley's (1951) work on the serial ordering of action (including language, music, and motor actions), and more recently in the writings of Corballis (2014), Dehaene (Dehaene et al., 2015), Fitch (Fitch & Martins, 2014), and Osherson (Monti & Osherson, 2012). The UGF framework goes beyond these kernels by formalizing the architecture and then drawing out several interesting empirical implications in the areas of evolutionary biology, neurobiology, developmental psychology, and artificial intelligence.

3. The architecture of UGF

In formulating his theory of the (*domain-specific*) language faculty, Chomsky (1966) revived many ideas of Cartesian science. One of the most profound ideas, however, remained unexhumed: Descartes' (1637) notion of the human mind as a "universal instrument", enabling us to exercise *general* intelligence in any of an infinite number of situations. Such universality is, so Descartes claimed, unique to our species; the specialized "organs" of other animals and machines (from the automata of the 17th Century to the deep learning neural nets of today) are restricted to specific domains, enabling a highly myopic or narrow intelligence. This Cartesian notion of universality lay dormant until Turing (1936) formulated his abstract "universal machine", modeled on the mathematical competence of human "computers." Turing of course idealized human competence, abstracting away from the finitude of the mind/brain — a rational move we follow in this essay. And although Turing's work influenced the formulation of generative grammar, the (possible) universality of human cognition would not be Chomsky's explicandum, for he argued (correctly) that an arbitrary Turing machine is too unstructured to serve as a generative grammar (Chomsky, 1963). But, in our judgment, the specificity of linguistic (and mathematical, moral, and musical) structures and the generality of intelligence can be unified and explained by our theoretical formulation of UGF.

UGF, we argue, must comprise the computational machinery necessary for natural language. This is the machinery for a particular, tripartite formalization of *recursion* (Watumull, Hauser, Roberts, & Hornstein, 2014b): (1) *computability*, enabling *explicit* enumeration of nontrivial patterns in recognition-learning and generation-production, as opposed to mere retrievability of data (memorized or genetically installed) from a look-up table; (2) *definition by induction*, from which hierarchical structure (the "architecture of complexity" (Simon, 1969) and information) is generated; and (3) *mathematical induction*, which allows for unbounded generation (and creativity). This machinery endows humans with a *strong generative capacity*: in

Please cite this article in press as: Hauser, M. D., & Watumull, J., The Universal Generative Faculty: The source of our expressive power in language, mathematics, morality, and music, Journal of Neurolinguistics (2016), http://dx.doi.org/10.1016/j.jneuroling.2016.10.005

the terminology of mathematical linguistics, a grammar *strongly generates* a set of structures and *weakly generates* a corresponding set of strings. And it is over strongly generated structures (interpreted at the interface with the conceptual-intentional system) that meanings are constructed, analogies made, and inferences drawn. On this view, structures are the stuff of intelligence, regardless of domain, and the core of UGF's functionality. Strings are the residue of mapping the grammar-generated structures to the sensory-motor system (or the artifacts of applying associativity to the structures).

Mathematical linguistics defines a *grammar* as a schema of rules — codified in an equivalent *automaton* — for enumerating and recognizing a set — called a *language* — of well-formed formulas. The standard taxonomy of grammars is arranged as a hierarchy — the Chomsky Hierarchy (Fig. 1) — of increasing generative capacity, from simple regular grammars (equivalent to

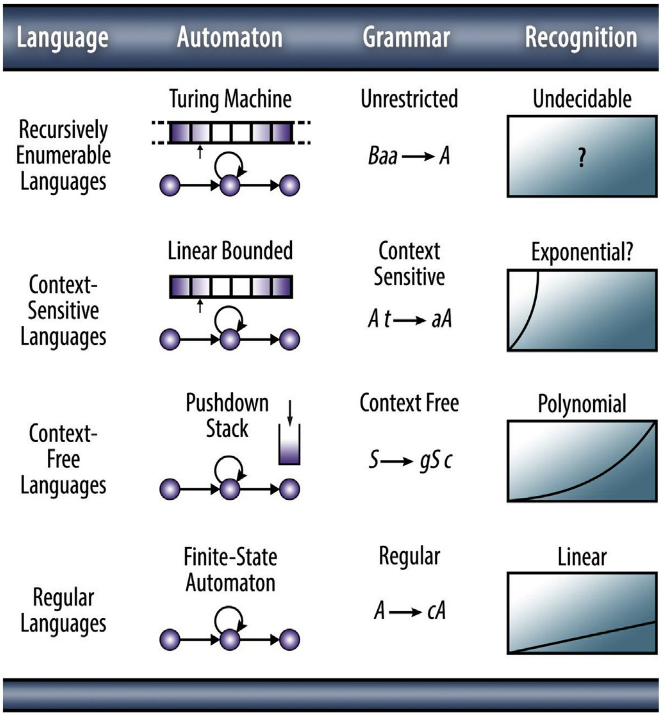

Fig. 1. The Chomsky Hiearachy (adapted from Searls, 2012). Mathematical linguistics defines a *Language* (first column on the left) as a set of strings of symbols assembled from a given *alphabet*. An *Automaton* (second column) is a machine that performs a function according to a predetermined set of coded instructions. A *Grammar* (third column) is a schema of rules for specifying a language. And the *Recognition* graphs (fourth column) show the computational complexity, in the general case, of recognizing whether a given string is a member of a given language, charting how the time required grows as a function of the length of the input string.
The grammars consist of a set of rewrite rules that assume forms such as A → xB. Here, upper-case letters denote temporary or *nonterminal* symbols, which are not members of the alphabet; lower-case letters are permanent or *terminal* symbols that are members of the alphabet. The example rule specifies that any occurrence of the nonterminal A may be rewritten as an x followed by a B. Starting with a nonterminal S, a grammatical derivation consists of a series of rewrite steps that halts when the final nonterminal is eliminated. For instance, consider a simple grammar with alphabet x and y, and rules S → xS and S → y. The grammar generates all strings beginning with an arbitrary number of x's and ending with one y. It yields derivations such as S ⇒ xS ⇒ xxS ⇒ xxxS ⇒ xxxy, where a double arrow signifies the application of a single-arrow rule.
A grammar with rules that rewrite a nonterminal as a terminal followed by at most one nonterminal is a *regular grammar*, generative of a *regular language*. An equivalent method for generating such languages is a *finite-state automaton* (FSA), a mathematical machine for representing particular styles of computation, comprising states (circles in figure) which are interconnected by transitions (arrows) that output symbols from the alphabet as they are traversed.
Grammars that generate any arrangement of terminals and nonterminals on the right-hand sides of rules have greater expressive power. These are the *context-free grammars*, which can generate all regular languages, and, in addition, non-regular languages such as strings of x's followed by an equal number of y's (e.g., xxxxyyyy). Such languages cannot be specified by a regular grammar (or FSA) because these devices have no read/write memory for representing the number of x's generated—a number necessary when it is time to derive the y's. But with context-free rules such as S → xSy, which necessarily generate an x and a y simultaneously, there is no memory problem. Alternatively, an automaton augmented with a *push-down memory stack*, which pushes/pops symbols to/from a stack during transitions, provides the necessary counting capability. In either case, context-free languages contain strings that encode dependencies between terminals, such as the relations linking x's and y's in the example, assuming those dependencies can be nested, either strictly within or independent of each other, but never crossing.
Even context-free grammars are insufficient for some languages such as, say, strings of consecutive x's, y's, and z's in equal number. This necessarily entails “cross-serial” dependencies, which require additional symbols on their left-hand side (and never more than one their right-hand side). Such *context-sensitive grammars* correspond to *linear bounded automata*, which are machines equipped with bounded read/write memory tape. Context-sensitive languages include all context-free languages (and many more), yet theoretically there exist languages beyond this set—the recursively enumerable languages, generated by unrestricted grammars or universal Turing machines.
In this figure, the language classes in the left column contain precisely those languages that can be generated by the automata and grammars listed in the next two columns. Each level contains all levels below it.

finite-state machines) to the arbitrarily complex grammars of a Turing machine. The Chomsky Hierarchy has proven profoundly important in mathematics and computer science, and most recently in the cognitive sciences (Fitch & Friederici, 2012), where it is defined in terms of weakly generated strings, for mathematical convenience. But the price to pay for such convenience — the price that must be paid to classify other organisms' capacities on the Chomsky Hierarchy (see Section 4) —is considerable: *irrelevance for natural language* (see Chomsky, 2007). The UGF proposition revamps the Chomsky Hierarchy in terms of strong generation, thereby opening the door to new phylogenetic and ontogenetic research questions.

The revamp we have in mind is actually a redesign to align with cognitive reality: the human mind is a universal Turing machine (universal *in principle* if given access to unlimited memory) that interfaces with — is constrained by — different domains (effectively different programs) so as to generate different representations. The claim that the human mind is a universal Turing machine, but not "beyond the Turing limit" (Siegelmann, 1995), can be corroborated theoretically and empirically (Deutsh, 2011; Watumull, 2012) and proved mathematically (Watumull, 2015a, 2015b); indeed neuroscientists have begun to investigate how the particular components of Turing's architecture can be realized in the brain (Zylberberg, Dehaene, Roelfsema, & Sigman, 2011). The Universal Generative Faculty is thus the suite of generative functions as they interface with different domains of knowledge. On this view, it is the specific constraints within each domain that delimit — indeed define the essence of — the flexible sets of varyingly "well-formed" representations. More specifically, it is the interface between UGF and each domain that determines why only particular patterns of words are considered grammatical (more or less), only particular patterns of notes are perceived as melodic (more or less), and only particular patterns of actions are deemed ethical (more or less). Indeed, the notion that constraints are essential to the creative generation of patterns (representations), appreciated in the philosophy of esthetics and some artificial intelligence research (Hofstader, 1985), should be further developed within the cognitive sciences.

Building from the ideas presented thus far, we hypothesize that Turing universality is ubiquitous in the natural world (Davis, 2011; Deutsh, 2011; Wolfram, 2013). The mental faculties of many nonhuman animals — and thus, their neural design — are most likely based on universal Turing machines, a thesis that is eminently testable and to some extent has, as we review in the next section. This comparative proposition could be characterized as the virtual conceptual necessity of Turing computation that Gallistel and King (2009) have emphasized, and that Berwick and Chomsky (2016) pointed out in their writings on human uniqueness: "It seems that insect navigation, like that of dead-reckoning ants […], requires the ability to 'read from' and 'write to' simple tape-like memory cells. But if so, that's all one needs for a Turing machine. So if all this is true, then ants have already climbed all the way up Nature's ladder. The puzzle[…] is that apparently ants don't build arbitrarily complex hierarchical expressions the way people do" (pp. 131–132); and, we would add, ants don't exapt the generative navigational operations for other domains. What's true for dead-reckoning is probably also true of birdsong and a number of other cognitive systems in nonhuman animals, a topic we pick up in the next section. But the puzzle is solved, or partially so, if one views UGF as a universal Turing machine, with components of the human UGF, including its universality, homologous or analogous with some nonhuman animals. From this perspective, the uniqueness of human cognition (Hauser, 2009) derives from the uniquely empowering constraints on universality (Watumull, 2015a, 2015b). In its design of the mind, Evolution's ode could well have been:

> *Every task involves constraint,*
> *Solve the thing without complaint;*
> *There are magic links and chains*
> *Forged to loose our rigid brains.*
> *Structures, strictures, though they bind,*
> *Strangely liberate the mind.*
> (Falen in Hostadter (1997): p. 272).

Constraints on universality would define levels of the Chomsky Hierarchy, reformulated in terms of strong generation. This is a task for the UGF research program, one that we begin to sketch next.

Consider the claim, repeated frequently in attempts to model language processing, that linguistic phenomena can be explained by "sequential" rather than hierarchical structure (Frank, Bod, & Christiansen, 2012). A reformulation of the Chomsky Hierarchy would refute the claim, showing that even the simplest regular language RL → $\{a^n\}$ is not the sequential {a, aa, aaa, aaaa, … }, but rather the set of strongly generated hierarchical structures {a, {a, a}, {{a, a}, a}, {{{a, a}, a}, a}, … }. Only associativity — an added, unnecessary, and empirically unmotivated complication — can flatten it into a sequence of strings. The parsimony of strong generation makes mathematical and evolutionary sense. With such a reformulation in hand, research would focus on identifying the constraints that define the levels of the Chomsky Hierarchy, as these determine how different cognitive systems can be classified.

At a higher level of the Chomsky Hierarchy, but also part of UGF's suite of generative procedures, consider *Merge*, first defined with respect to language by Chomsky (1995). In the simplest terms, Merge forms a set out of two objects. More formally, S is the set of possible syntactic objects (SOs) —"syntactic" in the most general, algebraic sense of a formal symbol — and its binary function f —*Merge* — constructs complex SOs by forming 2-sets (i.e., sets containing two SOs (either or both of which can be complex)): $f_{\mathrm{MERGE}}(\mathrm{SO}_1, \mathrm{SO}_2) \rightarrow \mathrm{SO}_3$ (with $\mathrm{SO}_3 = \{\mathrm{SO}_1, \mathrm{SO}_2\}$); indeed the magma axiom states that for any two members of S, α and β, application of f to α and β generates a member of S: $\forall\ (\alpha, \beta) \in S$: $f(\alpha, \beta) \in S$. For instance, in English, given that the elements *a* and *book* are members of S, then so is the set {*a*, *book*} formed by f_{MERGE}; informally, call simple syntactic objects (e.g., *a*, *book*, etc.) *lexical items* and complex syntactic objects (e.g., {*a*, *book*}, etc.) *phrases*.

Please cite this article in press as: Hauser, M. D., & Watumull, J., The Universal Generative Faculty: The source of our expressive power in language, mathematics, morality, and music, Journal of Neurolinguistics (2016), http://dx.doi.org/10.1016/j.jneuroling.2016.10.005

As a member of the suite of generative procedures within UGF, Merge is accessed by the language faculty to construct a discrete infinity of hierarchically structured expressions legible at interfaces with the conceptual-intentional and sensory-motor systems. Other systems, such as the music faculty, also access Merge to generate hierarchically structured expressions, but the simple syntactic objects (the "atoms of computation") used in those expressions differ. Roberts (2011) and Katz and Pesetsky (2011) have shown that many — perhaps all — of the formal properties of language (Merge, most importantly) have analogues in music (Fig. 2), the domains being defined and differentiated by the values of variables in Merge (e.g., lexical items for language, notes for music, etc.) and the systems with which they interface. For example, language interfaces with systems of knowledge, belief, and reasoning that enable formal semantic notions of truth and reference, whereas music interfaces with, among other things, our emotions, pitch perception, and rhythm in unique, ineffable ways to articulate notions of harmony and beauty.

Merge may also play a role in moral judgment, combining social concepts into complex representations in a computable process, defined by induction (thus arranging the representations hierarchically), over a limitless range via mathematical induction. One of the most important functions of such representations is in enabling sophisticated social cognition. For instance, Thomas, DeScioli, De Freitas, and Pinker (2016) provide experimental evidence for the power of "recursive mentalizing" — a recursive theory of mind in which one infers what others know about what one knows — in both moral and non-moral situations. The recursive theory of mind functions as a game theoretic constraint, possibly explaining the morally relevant bystander effect: the more potential helpers there are, the less likely any individual is to help. Individuals will therefore "strategically shirk" when they think others feel obligated to help. This is but one example of how recursion and structured representations may enter into moral reasoning more generally, and recent neuroimaging work by Frankland and Greene (2015) suggests that a region of the superior temporal cortex may play a role as it supports agent-action encoding in the form of *who did what to whom*.

Here we sketch how UGF could interface with domain-specific moral representations. Note that this sketch has some parallels to the important work initiated by Mikhail (2011) on moral action trees. A central difference is that on our view, the generative machinery comes from UGF, not a moral faculty (see also Hauser, 2012); the moral faculty or domain supplies the specific conceptual representations and constraints that interface with UGF. More specifically, UGF makes available its suite of generative procedures to the moral domain, as it does for all other domains. Which procedure is selected will depend in part on the interface conditions, as well as the domain's function in thought and action-expression. We assume that Merge is one such option for morality, especially given the parallels between language and morality that have been drawn by several authors (Dwyer, 1999; Harman, 1999; Hauser, 2006; Mikhail, 2011), dating back to John Rawls (1971).

To begin, an *event* is to a moral situation as a *verb* is to a sentence: it is the dynamic chassis of the mental representation (Pinker, 2007). For morality, the inputs to recursive Merge are actions, causes, consequences, and so on — as contrasted with lexical items in language, notes in music, and numbers and variables in mathematics. The output is an analogue of a sentential phrase, but here it is a moral principle, judgment, or statement, to be adjudged for its ethical or social permissibility. As with complex patterns in language, mathematics, or music, this structure is "algebraic" or "formal" in its representation of the shape, arrangement, and relations of variables (symbols). Thus we can define a syntax of moral events, analogous to "event structure" in theoretical linguistics (Ramchandran 2010).

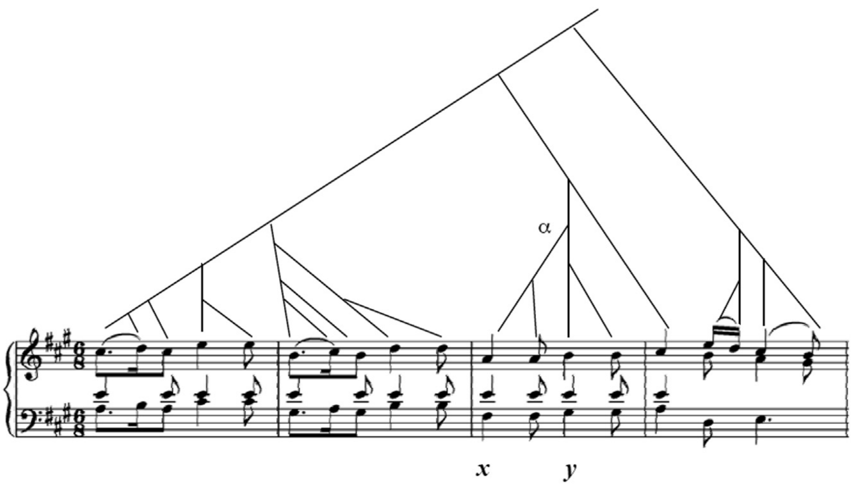

Fig. 2. The syntactic structure for Mozart piano sonata K. 331 (Katz & Pesetsky, 2011). The tree (the "prolongation reduction structure") is the harmonic interpretation of the piece (represented in conventional musical notation). The gist of this is that prolongation reduction, like natural language, forms binary-branching, endocentric, acyclic directed graphs; in addition, the constituents of music and language encode structural relations between elements that are not necessarily string adjacent. The fact that relationships between elements need not be adjacent, and can be located at arbitrarily long distances, rules out simple statistical or probabilistic rules or procedures. In short, there is an isomorphism in the formal structures of language and music, and importantly, an identity in the generative procedure that creates them.

Please cite this article in press as: Hauser, M. D., & Watumull, J., The Universal Generative Faculty: The source of our expressive power in language, mathematics, morality, and music, Journal of Neurolinguistics (2016), http://dx.doi.org/10.1016/j.jneuroling.2016.10.005

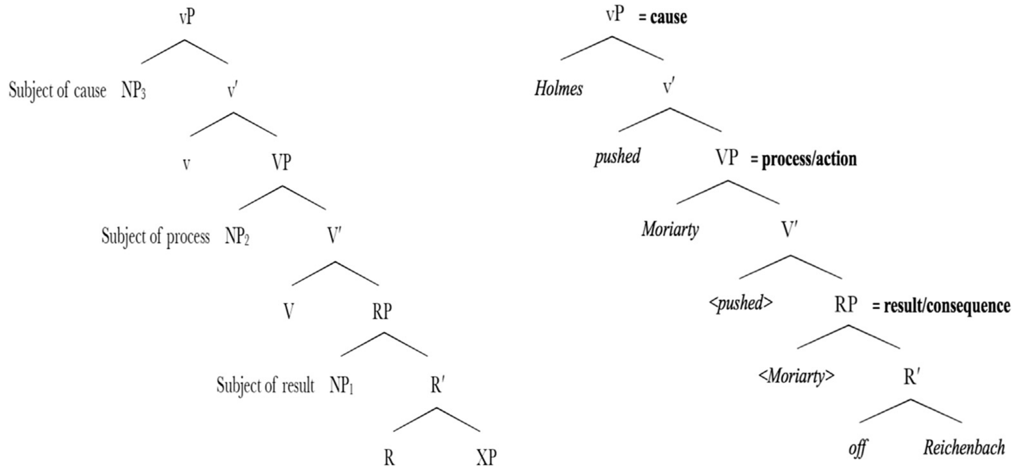

Fig. 3. LEFT: Canonical event structure syntax. Here, "syntax" is used in a general sense of a purely symbolic (or "algebraic"), formal representation. **vP** is to moral syntax what a transitive verb phrase is to linguistic syntax, representing the cause of the event. **NP** is the moral analogue of a noun phrase, representing the subject (e.g., the agent) causing the event. **v′** (and **V′** and **R′**) is an "intermediate projection"—a set theoretic outcome that emerges in any syntactic structure (linguistic, moral, musical, mathematical), with empirically detectable effects. **VP** is analogous to a verb phrase, and in the moral syntax expresses the process of the event. **v** and **V** characterize subtly different aspects of the cause and process dynamics. **RP** represents the result ph(r)ase of the event for **R** (whatever/whomever that may be) and **XP** (the moral analogue of a variable (X) phrase, representing additional properties of the event). RIGHT: The moral analogue to the canonical event structure.

The vP, VP, and RP of Fig. 3 (left side) are projections of the subcomponents of the event: the vP (formally equivalent to a transitive verb phrase) is a projection of the cause of the event; the VP (formally equivalent to a verb phrase) is a projection of the process of the event; and RP is a projection of the telos or result of the event. The VP is necessary for all dynamic verbs/actions, but the vP and RP are not. *Here the language/morality distinction is blurry or arguably nonexistent*; the following linguistic "constructions" may be — and we presume are in fact — effectively equivalent to the moral "constructions" they express — there is an isomorphism of *conceptual patterns*. The *result-transitive* construction "q saved r" contains an INITIATOR vP (q), an UNDERGOER VP (r), and a RESULTEE RP (r). We observe that the majority of morally charged thoughts (e.g., those involving harm to others) and expressions are of this form. The *process-transitive* construction "x pushed y" contains an INTIATOR vP (x) and an UNDERGOER VP (y). A substantial number of morally charged thoughts and expressions are of this form. The *process-result* construction "z died" contains an UNDERGOER VP (z) and a RESULETEE RP (z). Some morally charged thoughts and expressions are of this form. The *process-intransitive* construction "w ran" contains an UNDERGOER VP (w). Some morally charged expressions are of this form. All of these expressions represent mental structures encoding moral content. The thesis here is that UGF establishes the general form for such structures and that specific domains supply the content — in this case, the moral domain. The contentful structures are generated recursively and can be extended without bound by virtue of mathematical induction.

As noted earlier, mathematical induction is the mechanism for unbounded generativity. This aspect of recursion applies only to reifications of the natural numbers (or well-ordered sets), and these are intrinsically unboundedly hierarchical. Consequently, non-hierarchical mathematically inductive systems cannot exist. We equate "unbounded" with mathematically inductive, and thus birdsong, path integration, etc. are not unbounded, as we discuss next; these systems can be iterated indefinitely (as can, for instance, the locomotion of any organism), but infinite iteration is qualitatively different from the infinite recursion of Merge. Iteration does not carry information forward in time (i.e., it has no memory), and therefore cannot construct cumulative (increasingly complex) representations.

The UGF framework, as formulated here, allows for and in fact predicts differences between the language and action domains, given differences in the interfaces — predictions that are testable. Consider, for example, a claim made by Rizzi (2009) that because a recursive function as applied in the language faculty is "one that can indefinitely apply to its own output" (p. 65), it is homologous to a motor routine. But this is an invalid conclusion, predicated on a misunderstanding of recursion. A simple concatenative function such as a motor routine can apply indefinitely to its output, but it is not recursively generable if recursion is properly and rigorously defined: the property obtaining of a function that defines its output value in terms of the output values of *all* of its antecedent applications, ultimately grounded in an initial (base) application. This property of *self-reference* is an essential property of recursion. And as a result of self-reference, Merge generates the *self-similar* structures of fractal geometry. There is nothing homologous or analogous in the motor routines of humans, animals, or AI robotics (Husbands, 2014).

From our comments thus far, it does not follow that nonhuman animals are incapable of *recognizing-learning* patterns generated by Merge. In other words, it is perfectly possible for a computational system to recognize-learn patterns that it could not itself generate-produce. This is an elementary truth of computability theory, related to the idea of computable and

Please cite this article in press as: Hauser, M. D., & Watumull, J., The Universal Generative Faculty: The source of our expressive power in language, mathematics, morality, and music, Journal of Neurolinguistics (2016), http://dx.doi.org/10.1016/j.jneuroling.2016.10.005

computably enumerable sets (Martin, 1958). It is similar to the distinction in computational complexity theory between problems that can be *solved* in polynomial time versus solutions that can be *recognized* in polynomial time (the P versus NP problem, for which it is assumed that P does not equal NP). Similar distinctions are quite familiar from mathematical games such as Rubik's Cube in which it can be easy to generate a particular pattern, but hard to explain how precisely it emerged (and vice versa). And thus we would not be surprised to discover that, for instance, some avian species could distinguish patterns that, while never appearing in its native song, are distinctive of human language. Indeed we see such a phenomenon in humans: in the first stages of language acquisition, infants recognize-learn many more linguistic patterns than they can generate-produce. This developmental dissociation may have an analogue in some nonhuman animals, the only difference being that recognition and generation of Merged structures eventually overlap in human competence, but not in (some or all) nonhuman animals. Ultimately, the constraints on the universal Turing machines with which many animals are endowed limit their generative capacities to relatively simple structures (see Section 4). Human creativity, by contrast, is *recursively enumerable* or even *productive*[1]: it is not possible to enumerate all and only its “well-formed” representations.

The complexity of human creativity — bound up with what has often been called *general intelligence* — is a function of specific domains of representation accessing a universal Turing machine (UGF) which, in turn, provides a suite of generative procedures (e.g., Merge for linguistic and musical syntax, iteration for motor routines, the successor function for the number sense and mathematics, etc.). UGF is thus the Mycroft Holmes (Doyle, 1887) of the mind/brain: “The conclusions of every department are passed to him, and he is the central exchange, the clearinghouse, which makes out the balance. All other men are specialists, but his specialism is omniscience” (p. 400).

4. The evolution of UGF

If UGF is part of the human brain's design, with an architecture as proposed in the previous section, then it will be necessary to shift the theoretical focus of many phylogenetic or comparative problems. We think this shift opens the door to new questions that are more empirically tractable, while also making more transparent some of the essential causes of species differences in both cognitive capacity and behavioral expression. Instead of asking “Do animals have language, culture, mathematics, music, or morality?”, we ask “Do animals have the relevant generative computations that interface with domain-specific representations to recognize and generate the structures and/or contents of linguistic, cultural, mathematical, musical, or moral expressions?” But also, and importantly, we ask: “What generative computations underlie the specific specializations observed in animals during foraging, predator avoidance, communication, navigation, and so on?”

We begin by discussing this theoretical shift and then summarize a small subset of the work to illustrate its significance and implications for future research; more comprehensive reviews of the existing research can be found elsewhere (Dehaene et al., 2015; Fitch, 2014; ten Cate, 2016).

4.1. A theoretical shift

A common thesis in the study of comparative cognition, dating back at least as far as Darwin (1888), is that of mental continuity: though there are clearly quantitative differences between humans and other animals, and among nonhuman animals, there are no qualitative differences (de Waal, 2016). Thus, for example, though humans have language, an ever-growing technology, rich moral and legal rules, gastronomically and esthetically over-the-top food, and both tonal and atonal symphonies performed by large orchestras, nonhuman animals have “smaller” or proto-versions of these — bee language, New Caledonian crow tools, chimpanzee cooperation and punishment, ant fungal farms, and the dawn choruses of song birds and gibbons. If anything is considered qualitatively different, at least by some, it is language. For some, language is what allows us to think complex thoughts, invent technologies, teach and communicate ideas, and create cultures, legal doctrines, and culinary feasts. Language, in this sense, not only permeates all of human thought, but is the catalyst for the limitless creativity of these other domains; indeed, natural language essentially constitutes *the* language of thought. On this view, the comparative questions are how, when, and why did language evolve, and in what ways has our linguistic faculty transformed, perhaps quantitatively and qualitatively, other faculties? A second view, one that nonetheless acknowledges the infiltration of language into other domains of thought, unpacks the language faculty into distinctive mechanisms with the goal of exploring how these evolved.

Hauser, Chomsky, and Fitch (2002) provided a framework for exploring this second perspective on language evolution. Within this framework, there are processes or mechanisms that are involved in language, but are neither unique to language nor unique to humans. These constitute the Faculty of Language in the Broad sense (FLB), and include mechanisms such as working memory, attention, breathing, articulation, and certain aspects of both auditory and visual perception. In contrast, so the proposal goes, there are mechanisms that are both unique to language and unique to humans. These constitute the Faculty

[1] Tarski and Gödel demonstrated that truth is a *productive* set (neither recursive—meaning that all members/nonmembers of the set can be generated—nor recursively enumerable—meaning that all members of the set can be generated, but not the nonmembers): the productive set is that which cannot be produced by any finite rule schema but which, most importantly, can be approximated to higher and higher degrees of completeness by increasingly sophisticated rule schemata. And given that truths are mentally representable and linguistically expressible, it follows that human creativity is a productive set.

Please cite this article in press as: Hauser, M. D., & Watumull, J., The Universal Generative Faculty: The source of our expressive power in language, mathematics, morality, and music, Journal of Neurolinguistics (2016), http://dx.doi.org/10.1016/j.jneuroling.2016.10.005

of Language in the Narrow sense (FLN), and, as originally hypothesized, may include only the generative procedures underlying syntax and the mappings to the interfaces with the conceptual-intentional and sensory-motor systems.

The FLB-FLN perspective, along with others on language evolution, have led to considerable comparative work, including studies exploring natural communicative repertoires, the capacity to acquire elements of human language (spoken and signed), and the ability to perceive or extract patterned structures that map to different mathematical formalisms or expressions (Cheney & Seyfarth, 2008; Hauser, 1996; ten Cate, 2016). The latter approach is of particular interest to us here as it is in part what inspired our proposed shift to thinking about UGF (Fitch & Hauser, 2004; Fitch & Martins, 2014; Hauser, Barner, & O'Donnell, 2007). From the FLB-FLN perspective, the issue is quite narrowly focused on language, and the conclusions drawn by some (Berwick & Chomsky, 2016), including us (Hauser et al., 2014), is that nonhuman animals have nothing like the recursive syntax of language — no precursors or proto-languages. For example, Yang's (2013) comparative analysis shows that very young children generate structures consistent with a productive grammar whereas Nim Chimpsky, who was trained in American sign language, only shows evidence of memorized sequences. Similarly, Truswell's (2012) analyses of the bonobo Kanzi's ability to respond appropriately to spoken English reveals that there is no capacity to comprehend hierarchical structure, a core aspect of all human languages. Conclusions such as these effectively terminate the comparative research project. From the UGF perspective, in contrast, the central issues are about generative computations and interfaces to different domains of knowledge and systems for expression or action. Here, there are large, mainly unexplored vistas for a comparative research program.

Comparative work in this area (see reviews by Dehaene et al., 2015; Fitch & Martins, 2014; ten Cate, 2016) has sought to explore whether different species are capable of extracting the sequential patterns in artificially created grammars. The material developed for this work falls into three broad classes: 1) statistical or probabilistic sequences, 2) algebraic rules, and 3) generative algorithms within the Chomsky Hierarchy (see Section 3, Fig. 1) ranging from finite-state to Turing-universal, the latter being the most appropriate characterization of human language (Dehaene et al., 2015; Fitch, 2014; Watumull, 2015a, 2015b). Although all of the published work within this framework has been aimed at understanding human language, including its precursors, we see it instead as directly relevant to the evolution of UGF, and thus of far greater significance in terms of mapping out the phylogeny of the underlying system. In particular, within the UGF framework, we would ask: Do certain species have the capacity to both recognize-learn and generate-produce patterned sequences that are based on a probabilistic expression or rule, one that is used in vocal communication as well as in social situations involving cooperation and fair exchanges of goods or services? And to what extent does their access to these generative procedures extend up the Chomsky Hierarchy (in terms of strong generation), including the more powerful recursive procedures?

4.2. Comparative studies of UGF

As a starting point into relevant comparative research of UGF, consider the three types of sequential patterns noted in Section 4.1: statistical-probabilistic relationships between constituents or elements, algebraic rules, and generative algorithms within the Chomsky Hierarchy. These three types comprise a space of possible computations. Notably, each type is associated with a suite of computational procedures that vary in their abstractness and expressivity. We can imagine that for each type, a given species may only be capable of recognizing-learning structured patterns that fit this type, or may be capable of only generating-producing such patterns, or may be capable of both (analogous to the recursive/recursively-enumerable/productive distinction discussed in Section 3). Further, we imagine that either or both recognition-learning and generation-production are restricted to a particular domain, or deployed in more than one domain. This set of hypotheticals sets up a comparative approach, sketched in Fig. 4, and filled in to some extent with what is currently known about some species.

To date, comparative work on this problem has largely focused on birds, non-primate mammals, and nonhuman primates, with the majority of studies exploring recognition-learning of patterned auditory sequences. As summarized by Dehaene et al. (2015), Fitch (2014), and ten Cate (2016), there is evidence that species within each of these animal groups show abilities to extract statistical-probabilistic patterns and algebraic rules, with extremely limited evidence within the space of generative algorithms, mostly restricted to the lowest level of the Chomsky Hierarchy —that is, regular languages (Rogers & Hauser, 2010). Of the studies showing successful recognition-learning of patterned sequences, however, the majority entail artificially created patterns within one modality, based on extensive training procedures — and in some cases it is arguable whether the animals did in fact perform the claimed computations (discussed in Watumull, Hauser, & Berwick, 2014a); a far smaller set of studies have created patterned material from the species-specific repertoire (Comins & Gentner, 2013), explored visual stimuli (Ravignani, Westphal-Fitch, Aust, Schlumpp, & Fitch, 2015), or used non-training spontaneous methods (Abe & Watanabe, 2011). No study has looked at pattern recognition across multiple domains, and to our knowledge, only one study has explored and showed successful transfer across modalities (visual to auditory (Murphy, Mondragon, & Murphy, 2008);). This is significant because "transfer learning" is essential to the general intelligence of humans and one of the most difficult to replicate in artificial intelligence (Hinrichs & Forbus, 2011).

As Fig. 4 reveals, most of the species capable of recognizing-learning the statistical-probabilistic or algebraic rules do not use these procedures in generating-producing behaviors. This dissociation arises in so many different contexts and in so many different species that it represents a significant contrast with our own species, which can analogize across domains and from which emerges a more general intelligence. For example, despite the fact that rats and several nonhuman primates can recognize-learn adjacent and non-adjacent statistical relationships in patterned acoustic strings (Endress, Carden, Versace, & Hauser, 2009; Newport, Hauser, Spaepen, & Aslin, 2004; Sonnweber, Ravignani, & Fitch, 2015), as well as sequences that fit

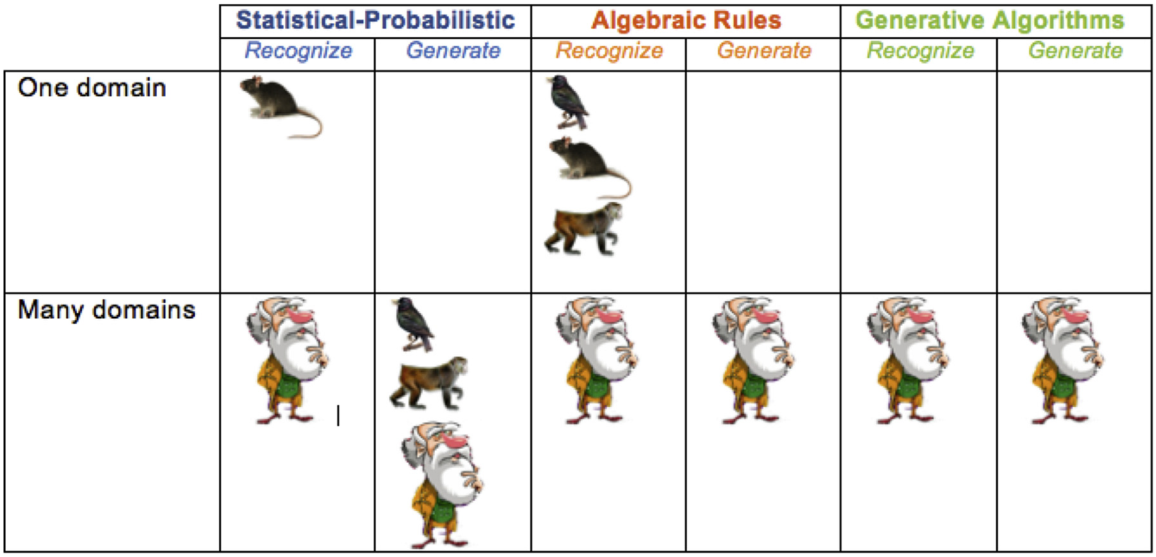

Fig. 4. A space of possible computations available for both recognition-learning and generation-production, in one or many domains. In cases where there is empirical evidence, a cell is filled with a representative taxonomic group (see KEY).
KEY:

: Birds—the majority of birds tested are Oscine songbirds (e.g., finches, starlings).

: Non-primate mammals; domesticated rats are the primary species tested.

: Nonhuman primates; different New World (e.g., tamarins, squirrel monkeys) and Old World monkeys (baboons, macaques), and Apes (chimpanzees).

: Humans; Western college students, primarily.

certain algebraic rules involving, for instance, repetitions (Murphy et al., 2008), none of their vocalizations or communicative gestures is patterned on the basis of such statistical or algebraic rules or regularities. The communicative systems of these animals are primarily based on single, discrete signals; when the same signal is repeated, or linked to a different signal, the sequence reflects changes in the signaler's motivational or emotional state. It is of course possible that despite the lack of evidence in the communicative domain, some of these animals generate actions in other domains that are mediated by such statistical or algebraic rules; this is a challenge for future research.

Song-learning birds represent a partial exception to the dissociation between recognition-learning and generation-production noted above for mammals. Among the Oscine songbirds, some are classified as close-ended learners whereas others are open-ended (Petkov & Jarvis, 2012). The close-ended learners typically acquire their species-specific song during a sensitive period of development, within the first year of life. Once acquired, the notes, as well as their specific sequence, remain fixed throughout life. Thus, any rule-based system for generating the song is only operative during early ontogeny, and then closes down, effectively reduced to a look-up table (a non-recursive mechanism). In contrast, the open-ended learners are not only endowed with a generative system that remains open throughout life, thereby enabling different combinations of notes (perhaps based on something like a productive set), but for some species (e.g., parrots, starlings, lyre birds), the system allows them to integrate sounds that fall outside their species-specific inventory, including artificial sounds (e.g., telephone rings, chain saws). As Berwick, Okanoya, Beckers, and Bolhuis (2011) note, analyses of different species of songbirds reveal that they can be classified according to different levels of the Chomsky Hierarchy. For example, starlings generate songs that last from 1 to 2 min, consisting of different motifs or patterns of notes that appear in a fixed order. The transitions between these elements are best captured by a first-order Markov model. Thus, for starlings, if you know one motif, you have a high probability of predicting the next motif; bigrams suffice to capture the statistics of this song system. In contrast, a Markov model is insufficient to capture the patterning of Bengalese finch song. The song of this finch is generated from a repertoire of

Please cite this article in press as: Hauser, M. D., & Watumull, J., The Universal Generative Faculty: The source of our expressive power in language, mathematics, morality, and music, Journal of Neurolinguistics (2016), http://dx.doi.org/10.1016/j.jneuroling.2016.10.005

approximately eight different notes, with notes clustered into 2–5 higher-order units or constituents. Bigrams fail to explain transitions because the constituents can appear in one place early in a song and then reappear in many, arbitrarily distant locations later in the song. To explain the patterning of constituents in a Bengalese finch song, and most likely other open-ended learners, requires the generative procedures or rules of the regular languages within the Chomsky Hierarchy, a level in which recognition-learning and generation-production is still restricted to local relationships among constituents. Surprisingly, and as discussed next, the generative computations that we see among songbirds doesn't extend beyond their song system into any other domain or sensory modality. We believe this says something profound about the limitations of such systems in nonhuman animals, and the evolution of a fundamentally different system in the human animal.

Before concluding this section, we mention a different, but highly relevant literature in terms of sequences — action sequences. As noted earlier, several authors, starting with Lashley, have noted parallels between the generativity of language and the motor or action domain (Pulvermüller & Fadiga, 2010). Though both domains are clearly generative, it is important to recognize key differences (Moro, 2013): linguistic syntax is mind-internal, whereas actions are externalized; a word is quite unlike an action as a unit; and constraints on the generative operation (Merge) in language relate to locality concerns, whereas they relate to the physical or motor systems in the action domain.

Of interest to us here is the comparative evidence in the action domain of computations beyond concatenation, as it reveals that animals recognize-learn distinctive action sequences, but are far more limited with respect to both generation-production and translation across different domains and modalities. This is an area that will require further comparative research.

At a behavioral level, several studies of nonhuman primates reveal that they perceive the difference between accidental and intentional gestures, including actions that result in danger or the indication of food — patterns that go beyond those of a simple concatenative function. Thus, for example, though rhesus monkeys are incapable of throwing, they recognize that in humans, certain kinematic rotations of the arm, together with the grasp of a rock and the alternation of eye gaze toward a target, combine to create a threatening situation (Wood, Glynn, & Hauser, 2007a); using the same sequence of actions without eye gaze, or including eye gaze but with a slow rotation of the arm, are perceived as functionally different (i.e., non-threatening). Similarly, rhesus will selectively approach one of two boxes concealing food when a human experimenter moves his arm and hand and then grasps the box, but not when this same motion is used but the hand flops onto the box (Wood, Glynn, Phillips, & Hauser, 2007b). These data, and others, reveal that rhesus monkeys, as well as other animals, are capable of recognizing-learning the significance of different action sequences. What has yet to be explored is whether rhesus, as well as other animals, are capable of recognizing-learning action sequences that are generated on the basis of other statistics, algebraic rules, or different levels of expressions within the Chomsky Hierarchy.

4.3. *Comparative studies of interfaces between UGF and other domains*

In our own species, the different generative procedures are available in different domains, and thus, take as input different types of perceptual and conceptual representations, enabling fluid concepts to enter into creative analogies. What is striking about the evidence to date for nonhuman animals is how seemingly un-available or inaccessible such procedures are with respect to implementation across domains. This comparative claim says something profound, we believe, about the interfaces from UGF to other mind-internal processes, pointing the way to questions about neural implementation and evolution.

To illustrate the comparative claim made here about interfaces, consider the Oscine song birds, a relatively large avian group in which all species learn their species-specific song through a process of imitation or copying — a process that, in many ways, resembles how children acquire language (Bolhuis & Everaert, 2013; Lennenberg, 1967; Marler, 1970). Not only is the ontogeny of this system similar in important ways to the acquisition of language, but so too is the recognition-learning and generation-production of song structure. In particular, Oscine song consists of a sequence of notes, arranged in a particular order to convey information about the individual, the local population, and species identity. Depending on the species, the notes are (i) either restricted to those that are part of the species-specific repertoire or may also include other sounds, or (ii) may be acquired once and remain fixed throughout life or change year to year, combining notes in different ways. What is striking about the system that underlies the acquisition, recognition-learning, and generation-production of song is that it is entirely restricted to song, with no evidence whatsoever of interfaces to either other domains or functions or modalities. In particular, despite the exquisite evidence of the songbird's capacity for imitation — a capacity that entails copying or duplication — this capacity is entirely restricted to song, with no evidence that this sub-order of birds can use this ability in the visual modality or with other acoustic signals. For example, no male songbird imitates the song of his competitive neighbor (though lyre birds can imitate other species' songs, giving the impression of a saturated habitat), a move that might be useful in recruiting extra sexual favors. Similarly, despite the songbird's prowess in combinatorially generating novel varieties of note sequences, they never deploy this capacity in any other domain, or even within the same domain of communication, extending to their calls or visual signals. The entire system, including its ontogeny, recognition-learning, and generation-production, is restricted to one signal and one function or domain. There are no interfaces outside of this system, and no transfer learning. Thus this is a classic case of Fodor (1983) “encapsulation”.

A second example of the interface problem comes from the action system, and in particular, the discovery of mirror neurons in macaque monkeys (Rizzolatti, Fadiga, Gallese, & Fogassi, 1996). Mirror neurons are cells in premotor cortex that fire in response to *both* recognizing-learning and generating-producing an action. This discovery raised the idea that like our linguistic lexicon, we are also endowed with a motor lexicon (Caggiano et al., 2011). And like language, the motor lexicon interfaces with a syntactic or generative procedure that allows elements within the lexicon to be combined to create novel sequences. Regardless of whether the parallel or analogy between language and action is strong or weak, the comparative evidence suggests that the capacity available for recognition-learning is either non-existent or weakly available for generation-production in the motor domain. In particular, though the mirror neuron system recognizes different action sequences as evidenced by different neural responses, this system does not support the generation-production of either visual or vocal gestures with comparable complexity. To date, there is no evidence that macaques are capable of either visual or vocal imitation, and the lack of evidence can't be explained by either working memory or motor limitations (Iacoboni, 2009). Thus, despite the presence of a system that recognizes-learns motor sequences, one based on a motor lexicon of gestures, this system plays no role in the highly functional or adaptive motor outputs that are required for imitation. As with the songbird system, the work on mirror neurons further pushes us to consider limitations on the interfaces between UGF and both different domains and sensory modalities. It also adds a kink in those theories that see mirror neurons as “the DNA of the mind” and the source of all human brilliance, including language (Ramachandran, 2010, pp. 1–7).

In sum, the tripartite system that makes up our human UGF, in its entirety, appears unique to our species. There are components of UGF that, however, have analogues or homologues elsewhere in the animal kingdom. *Computability* is common to all generative systems, and *definition by induction* is ubiquitous in the generation of hierarchical structure. Thus it is possible for a system to be computable but not generative of hierarchy (e.g., Lashley's motor routines), as well as computable and generative of hierarchy but not unbounded (e.g., birdsong or human phonology). The UGF framework sets up an empirically tractable research space to address questions of evolutionary significance.

5. Conclusions

If our thesis regarding UGF is on the right track, it has several significant implications for our understanding of human thought, including how it evolved, develops, breaks down, varies cross-culturally, and could be modeled or replicated in artificial systems. Here, we briefly summarize our main points, and then raise a few implications, especially for human development, cross-cultural variation, and artificial intelligence.

The Universal Generative Faculty is a universal Turing machine comprising a suite of generative functions accessible by specific cognitive domains. These functions vary in their strong generative capacity (e.g., human language is defined by Turing universality, birdsong by a finite-state machine). The uniqueness of human cognition derives from the general intelligence that emerges from UGF's tripartite recursive system—computability, definition by induction, and mathematical induction. More specifically, UGF and its interfaces to different domains of knowledge enables the unbounded generation of hierarchically structured (indeed fractal) representations exapted for language, mathematics, music, and morality. The different domains differ only by the values (e.g., words, numbers, notes, and events) they substitute for the variables in the generative procedures. Critically, the Turing universality of human cognition is not so unrestricted as to be vacuous — its coherent creativity is a function of constraints imposed on UGF by the specific domains (analogous to the way constraints of rhyme and scansion engender the beauty of poetry).

As noted in Section 4, the UGF thesis and the framework that it entails, changes the conversation regarding phylogenetic or evolutionary issues, embracing several empirical studies carried out under a different theoretical focus, and opening the door to new studies that can help support or refute our ideas. In particular, instead of looking at nonhuman animal studies using artificial grammars (patterns that adhere to particular statistical, algebraic, or generative rules) as providing evidence for language-like competences or precursors to language, we look at these studies as directly relevant to UGF, exploring which generative procedures, formulated in terms of strong generative capacity, are shared in common with humans and which are not, which procedures underly either or both recognition-learning and generation-production, and which are deployed in one or more domains. Moreover, by pursuing the UGF framework, we also ask which generative procedures mediate the capacities of animals during foraging, navigation, communication, and social interaction, irrespective of parallels with humans. What is striking about the data thus far is that for much of the animal world, even when there is evidence for generative procedures at the level of regular languages (the lowest level of the Chomsky Hierarchy), there is virtually no evidence that such procedures mediate both recognition-learning and generation-production across different domains. Part of what it means to be a normally developing human is that such procedures are available for both recognition-learning and generation-production across multiple domains, though in cases of expertise (composer, professional mathematician, poet), the generation component may far exceed what mere mortals can produce. What this tells us is that the generative procedures that are the essence of UGF are promiscuously available to different domains of knowledge in humans but perhaps no other species; alternatively, it is possible that the UGF of other species lacks some of the procedures contained in the UGF of humans. The comparative evidence also suggests that the issue of availability arises because of legibility at the interfaces. These observations, in turn, raise fascinating questions, and some solutions, for studies of human development, cross-cultural variation, neurological deficits, and AI.

Across human development, different domains of knowledge appear to come on-line at different ages, with different time tables for recognition-learning and generation-production. The latter distinction has often been referred to as the competence-performance distinction. In other words, a child's knowledge of, say, language, morality, or mathematics appears very early in development as a matter of competence, but due to factors that enable generation-production (e.g., motor development and working memory), appear as actions or expressions only much later in development. Based on hundreds of studies, infants show an early capacity to recognize-learn the meaning of words, predict the properties of moving and hidden

Please cite this article in press as: Hauser, M. D., & Watumull, J., The Universal Generative Faculty: The source of our expressive power in language, mathematics, morality, and music, Journal of Neurolinguistics (2016), http://dx.doi.org/10.1016/j.jneuroling.2016.10.005

objects, calculate the outcomes of elementary arithmetical operations, recognize false beliefs in others, and determine who is morally good and bad, years before they can produce meaningful words, reach for the correct location of hidden objects, correctly use count words and compute nontrivial arithmetical tasks, state who has false beliefs, and help morally good agents while punishing bad ones (Banaji & Gelman, 2013; Barrett et al., 2013; Carey, 2009; Crain & Thornton, 2000; Hamlin, 2013; Spelke, 1994; Wellman, 2014; Wynn, 2002).

When one effectively controls for potential performance effects by looking at children at the same age, one nonetheless sees differences in capacity across domains. For example, Yang's (2013) elegant analyses of 2-year old children's expressions reveal that they combine syntactic categories by means of Merge. That is, by the second year of life the generate-produce system is capable of engaging Merge to construct a virtually limitless range of, at least, two word expressions. Soon thereafter, by at least 3 years, children can embed one phrase inside another, shifting from "Give kitty ball" and "Nice kitty" to "Give nice kitty ball" (Crain & Thornton, 2000; Roeper, 2007). In contrast, if one compares the ontogeny of theory of mind capacities, fully recursive comprehension and expression (e.g., Sally has a false belief) does not appear until 5–6 years of age (Wimmer & Perner, 2002), followed a couple of years later by the ability to engage in Gricean embedding that involves an inductively recursive operation (e.g., "I know that you know that I am lying"; Thomas et al., 2016; Wellman, 2014). And in mathematics, order of operations — a form of embedding, ((3 + 4) x 7) — is not understood and used appropriately until the age of 9–10 years. Given that each of these domains — language, theory of mind, mathematics — tap the same form of recursive operation — embedding — why would each domain show different developmental timing trajectories?

To address the timing issue across different domains, consider the core architecture of UGF: a suite of generative procedures that interfaces with different domains of knowledge, including their conceptual representations and their means of expression or externalization. Logically, we must assume that the generative procedures are part of the brain's evolved design. That is, neither young nonhuman nor human animals learn how to use these procedures. In this sense, the generative procedures are like vision: animals don't learn how to see, but they do have to learn about the relevance of what they are seeing. Given this perspective, there are at least two possible ways in which the generative procedures could be laid out in the brain's circuitry. Either UGF is a centralized system of procedures that any domain can access, or UGF is more distributed, with both different procedures enabled in different domains (e.g., Merge in language, successor function in mathematics) as well as the same procedures duplicated for different domains (e.g, Merge for language and Merge for music). Though we will return to these two possibilities in a moment, in either case the procedures are available to the developing child as they are part of its endowment. What must develop is the interface between the relevant generative procedure and the particular conceptual representations. For reasons that remain unclear, it appears that the interfaces to language mature far earlier than those for mathematics, theory of mind, or morality; at present we are unaware of any relevant data for music. It could be that the relevant conceptual representations are slower to develop in different domains, which would delay the interface, or it could be that interface itself is slower to develop in different domains, which could be the result of different constraints on the generative procedure (e.g., working memory puts greater limits on embedding for theory of mind than for language); it is also possible that the differences arise due to different methodologies, with some providing an earlier lens on the child's generative capacity. This is a fertile area of research that, ultimately, can decide between different competing hypotheses.

The differences in development may also help explain certain cross-cultural patterns, including one that we discussed earlier: cultures that have words for all numbers and those that have only a few count words. Cultures that show a limited range of count words typically express "one", "two", "three", and sometimes "four" and "five". Numbers that exceed these values are dumped into the more generic quantifier "many". On the UGF thesis, these individuals certainly have access to the successor function — as do all humans — but simply do not implement this generative procedure. In these cultures, other generative procedures of course appear in language, theory of mind and morality. Our explanation for these truncated count systems is that they represent a limitation at the interfaces. For example, the Mundurucu have words for 1, 2, 3, and 4, and then use a word for "many". The Mundurucu also show, as do all other humans, a facility with the approximate number system, one that is not limited in magnitude but is limited with respect to ratios. Thus, the generative procedure maps neither to the lexicon nor to the analogue magnitude representations, leaving the Mundurucu with a non-generative, implicit system of mathematics. This interpretation differs from at least two common alternatives: one, that the exact number system develops along with the recursive properties of language (Hauser et al., 2002; Leslie et al., 2008) or two, that the exact number system is culturally constructed by mapping the child's developing understanding of the words "one, two, three, four, …" to an object tracking system that precisely quantifies small numbers of objects up to about four (Carey, 2009; Hurford, 1987). The first proposal sees number as piggy-backing on language, the second as being derived from experience, of literally learning the list of integers. The UGF perspective provides a more accurate explanation of the data, and in particular the fact that young children show an inductive leap once they have acquired the meaning of the words "one, two, three" (see Yang, 2016 for a different account of when the inductive leap arises). This is easily explained by two features of the UGF framework: 1) all that is needed to generate the count words is the interface between the successor function and the linguistically labeled concepts — once formed, the entire integer list emerges at once – and 2) language and number do not share completely overlapping neural structures (Monti et al., 2012; Amalric & Dehaene, 2016b).

Turning briefly to neurobiology, several recent studies have invoked the idea that different domains of thought make use of either similar or different generative procedures. Many of these papers emerged in response to the ideas proposed by Hauser et al. (2002) concerning the evolution of language, and in particular, what is uniquely linguistic and human — that is, what is in FLN. We quote the key passage here as it has often been overlooked or misinterpreted and is also directly relevant to our discussion of language and mathematics:

> "FLN is a computational system (narrow syntax) that generates internal representations and maps them into the sensory-motor interface by the phonological system, and into the conceptual-intentional interface by the (formal) semantic system[…]. All approaches agree that a core property of FLN is recursion, attributed to narrow syntax in the conception just outlined. FLN takes a finite set of elements and yields a potentially infinite array of discrete expressions. This capacity of FLN yields discrete infinity (a property that also characterizes the natural numbers)." [p. 1571]

Importantly, recursion is seen as the core generative procedure, along with the interfaces. Equally important is the parallel drawn between language and mathematics. Of relevance are a series of theoretical and empirical papers making the claim that recursive operations are fundamental in language, mathematics, music, and morality, and that they rely on different neural resources. For example, Klessinger, Szczerbinski, and Varley (2007) report on a neuropsychological case of a man with severe aphasia who showed no deficits in solving algebra problems, including ones with only numbers. Additional support comes from Monti et al. (2012) neuroimaging study showing a neural dissociation between the processing of combinatorial operations in language and algebra. In another study, Bánréti, Hoffmann and Vincze (2016) explored two patient groups, one with Broca's aphasia and one with Alzheimer's disease. Results showed that the aphasics had significant difficulty understanding and producing sentences with embedded phrases, but had no difficulty understanding second order false beliefs, where mental states are effectively embedded. In contrast, the Alzheimer's patients showed the opposite effect: intact sentence embedding but significant deficiencies in mental state embedding. These studies show that different domains of thought access similar generative procedures. They also suggest that if UGF is the repository for such procedures, the procedures must be duplicated in order to explain the kind of double dissociation documented by Bánréti and colleagues. That is, UGF must consist of multiple copies of different generative procedures, which makes perfect sense if the architecture of UGF is based on a universal Turing machine: such a system can replicate arbitrarily many arbitrary programs (generative procedures). On this view, if the capacity for embedding was damaged in both language and mathematics, this would be due to each domain's generative procedure being damaged.

Finally, the architecture of UGF has implications for the creation of human-level artificial intelligence. UGF, consisting of domain-general generative procedures that interface with domain-specific concepts to create a limitless number of hierarchical representations, differs dramatically in kind from the machine learning neural networks currently voguish in AI research (LeCun, Bengio, & Hinton, 2015). The latter are not even *generative*: they do not explicitly and recursively enumerate patterned representations of discrete symbols; as Turing's original work showed, their outputs, being finite, are not computable. Thus, for instance, neural nets cannot even approach human-level, human-style linguistic syntax (as evidenced in the garbled grammar of Google Translate, a deep learning neural network). And without syntax (form), there can be no semantics (content). This is why neural nets cannot understand even simple expressions (e.g., Apple Siri produces nonsense (if anything) once the topic shifts even minimally from its small, scripted routines). Any "understanding" at all is, like in nonhuman animals, myopic and non-transferable across domains. For example, IBM's Watson can play a mean game of *Jeopardy!*, but cannot comprehend one iota of Go, the ancient Japanese board game recently "conquered" by Google's AlphaGo program. Conversely, AlphaGo would play *Jeopardy!* with the proficiency of a paperweight, emphasizing again the myopia of a seemingly intelligent system. These opaque, non-explicit, black boxes were laboriously trained—and could only be trained—for their specific, narrow tasks. And even here, they have no more genuine *understanding* of the games than an abacus does of arithmetic. Neither program—indeed no existing AI—could engage in even the most elementary of free-flowing conversations, for that would require the UGF-enabled domain-general commonsense knowledge typical of any three-year-old child. In short, deep learning AI systems are not equipped with the conceptual fluidity and generativity of human cognition, which we argue can emerge only from UGF.

A human-level AI would thus need to be endowed with UGF: a universal Turing machine constrained by interfaces with specific domains of competence whose structures can be transferred across domains to construct complex linguistic, mathematical, musical, and moral representations. The key to transferring and combining knowledge in fluid and creative ways is analogy: "The ceaseless activity of making mappings between freshly minted mental structures (new percepts) and older mental structures (old concepts)—the activity of pinpointing highly relevant concepts in novel situations—constitutes the analogical fabric of thought, and the unceasing flurry of analogies that we come up with is a mirror of our intelligence" (Hofstadter and Sander, 2013, pp. 126–127). So, until a program can recursively generate and analogize complex mental representations, forming a productive set of concepts, human-level AI will remain out of reach. In other words, *human-level* AI would need to have *human-style* thinking, and the human-style derives from UGF.

In sum, our proposal for thinking about human thinking in terms of a universal Turing machine, a system of generative procedures and interfaces to domain-specific representations, provides a precise specification of human creativity. But it also provides a different way of looking at evolution, development, cross-cultural variation, and artificial intelligence. This difference not only helps unify existing empirical threads, but opens the door to new ones.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Abe, K., & Watanabe, D. (2011). Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nature Neuroscience, 14*, 1067–1074.
Amalric, M., & Dehaene, S. (2016b). Origins of the brain networks for advanced mathematics in expert mathematicians. *Proceedings of the National Academy of Sciences, 113*, 4909–4917.
Anderson, S. W., Bechara, A., Damasio, H., & Tranel, D. (1999). Impairment of social and moral behavior related to early damage in human prefrontal cortex. *Nature, 11*, 1032–1037.
Banaji, M. R., & Gelman, S. A. (2013). *Navigating the social world*. New York: Oxford University Press.
Bánréti, Z., Hoffmann, I., & Vincze, V. (2016). Recursive subsystems in aphasia and Alzheimer's Disease: Case studies in syntax and theory of mind. *Frontiers in Psychology, 7*, 1069–1121.
Barrett, H. C., Broesch, T., Scott, R. M., He, Z., Baillargeon, R., Wu, D., et al. (2013). Early false-belief understanding in traditional non-Western societies. *Proceedings of the Royal Society B: Biological Sciences, 280*, 1–6.
Berwick, R. C., & Chomsky, N. (2016). *Why only us*. Cambridge: MIT Press.
Berwick, R. C., Okanoya, K., Beckers, G. J. L., & Bolhuis, J. J. (2011). Songs to syntax: The linguistics of birdsong. *Trends Cognitive Science, 15*, 113–121.
Bolhuis, J. J., & Everaert, M. B. H. (2013). *Birdsong, speech, and language*. Cambridge: MIT Press.
Caggiano, V., Fogassi, L., Rizzolatti, G., Pomper, J. K., Thier, P., Giese, M. A., et al. (2011). View-based encoding of actions in mirror neurons of area F5 in macaque premotor cortex. *Current Biology, 21*, 144–148.
Carey, S. (2009). *The origin of concepts*. New York: Oxford University Press.
ten Cate, C. (2016). Assessing the uniqueness of language: Animal grammatical abilities take center stage. *Psychonomic Bulletin Review*, 1–6.
Cheney, D. L., & Seyfarth, R. M. (2008). *Baboon metaphysics*. Chicago: University of Chicago Press.
Cheng, K. (1986). A purely geometric module in the rat's spatial representation. *Cognition, 23*, 149–178.
Chomsky, N. (1955). *The logical structure of linguistic theory*. New York: Plenum Press.
Chomsky, N. (1963). Formal properties of grammars. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1, pp. 323–418). New York: Wiley & Sons.
Chomsky, N. (1966). *Cartesian linguistics*. New York: Cambridge University Press.
Chomsky, N. (1995). *The minimalist program*. Cambridge: The MIT Press.
Chomsky, N. (2007). Review of margaret Boden's 'mind as machine: A history of cognitive'. *Artificial Intelligence, 171*, 1094–1103.
Comins, J. A., & Gentner, T. Q. (2013). Perceptual categories enable pattern generalization in songbirds. *Cognition, 128*, 113–118.
Corballis, M. C. (2014). *The recursive mind*. Princeton, NJ: Princeton University Press.
Crain, S., & Thornton, R. (2000). *Investigations in universal grammar*. Cambridge: Bradford Books.
Darwin, C. (1888). *The descent of man, and selection in relation to sex*. Princeton, NJ: Princeton University Press.
Davis, M. (2011). *The universal computer*. New York: Norton Press.
Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., & Pallier, C. (2015). The neural representation of sequences: From transition probabilities to algebraic patterns and linguistic trees. *Neuron, 88*, 2–19.
Descartes, R. (1637). *Discourse on the method*. New York: The Floating Press.
Deutsh, D. (2011). *The beginning of infinity*. New York: Viking Press.
Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2015). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience, 19*, 158–164.
Doyle, A. C. (1887). *Sherlock Holmes: The complete novels and short stories*. London: Bantam Books.
Dwyer, S. (1999). Moral competence. In K. Murasugi, & R. Stainton (Eds.), *Philosophy and linguistics* (pp. 169–190). Boulder, CO: Westview Press.
Endress, A. D., Carden, S., Versace, E., & Hauser, M. D. (2009). The apes' edge: Positional learning in chimpanzees and humans. *Animal Cognition, 13*, 483–495.
Fitch, W. T. (2014). Toward a computational framework for cognitive biology: Unifying approaches from cognitive neuroscience and comparative cognition. *Physics of Life Reviews, 11*, 329–364.
Fitch, W. T., & Friederici, A. D. (2012). Artificial grammar learning meets formal language theory: An overview. *Philosophical Transactions of the Royal Society, B, 367*, 1933–1955.
Fitch, W. T., & Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science, 303*, 377.
Fitch, W. T., & Martins, M. D. (2014). Hierarchical processing in music, language, and action: Lashley revisited. *Annual Review of the New York Academy of Sciences, 1316*, 87–104.
Fodor, J. (1983). *Modularity of mind*. Cambridge: MIT Press.
Frank, S. L., Bod, R., & Christiansen, M. H. (2012). How hierarchical is language use? *Proceedings of the Royal Society B: Biological Sciences, 279*, 4522–4531.
Frankland, S. M., & Greene, J. D. (2015). An architecture for encoding sentence meaning in left mid-superior temporal cortex. *Proceedings of the National Academy of Sciences, 112*(37), 11732–11737.
Friederich, R., & Friederici, A. D. (2009). Mathematical logic in the human brain: Syntax. *PLOS One, 4*(5), 1–7.
Gallistel, C. R., & King, A. P. (2009). *Memory and the computational brain*. New York: Wiley-Blackwell.
Goldenberg, G., Hermsdorfer, J., Glindemann, R., Rorden, C., & Karnath, H. O. (2007). Pantomime of tool use depends on integrity of left Inferior frontal cortex. *Cerebral Cortex, 17*, 2769–2776.
Hale, K. (1975). Gaps in grammar and culture. In M. D. Kinkade, K. Hale, & O. Werner (Eds.), *Linguistics and anthropology* (pp. 295–315). Lisse: The Peter de Ridder Press.
Hamlin, J. K. (2013). Moral judgment and action in preverbal infants and toddlers: Evidence for an innate moral core. *Current Directions in Psychological Science, 22*, 186–193.
Hamzei, F., Vry, M.-S., Saur, D., Glauche, V., Hoeren, M., Mader, I., et al. (2016). The dual-loop model and the human mirror neuron system: An exploratory combined fMRI and DTI study of the inferior frontal gyrus. *Cerebral Cortex, 26*, 2215–2224.
Harman, G. (1999). Moral philosophy and linguistics. In K. Brinkman (Ed.), *Proceedings of the the world congress of philosophy* (pp. 107–115). Bowling Green, OH: Philosophy Documentation Center.
Hauser, M. D. (1996). *The evolution of communication*. Cambridge: MIT Press.
Hauser, M. D. (2006). *Moral minds*. New York: Harper Collins.
Hauser, M. D. (2009). The possibility of impossible cultures. *Nature, 460*, 190–196.
Hauser, M. D. (2012). *The seeds of humanity. Tanner lectures on human values*. Salt Lake City: University of Utah Press.
Hauser, M. D., Barner, D., & O'Donnell, T. (2007). Evolutionary linguistics: A new look at an old landscape. *Language Learning and Development, 3*, 101–132.
Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science, 298*, 1569–1579.
Hauser, M. D., Yang, C., Berwick, R. C., Tattersall, I., Ryan, M. J., Watumull, J., et al. (2014). The mystery of language evolution. *Frontiers in Psychology, 5*, 1–12.
Hecht, E. E., Gutman, D. A., Khreisheh, N., Taylor, S. V., Kilner, J., Faisal, A. A., et al. (2014). Acquisition of Paleolithic toolmaking abilities involves structural remodeling to inferior frontoparietal regions. *Brain Structure and Function, 220*, 2315–2331.
Hermer, L., & Spelke, E. S. (1994). A geometric process for spatial reorientation in young children. *Nature, 370*, 57–59.
Hermer, L., & Spelke, E. S. (1996). Modularity and development: The case of spatial reorientation. *Cognition, 61*, 195–232.
Hinrichs, T. R., & Forbus, K. D. (2011). Transfer learning through analogy in games. *AI Magazine, 32*, 70–83.
Hofstader, D. R. (1985). *Metamagical themas*. New York: Basic Books.
Hofstadter, D. R. (1997). *Le ton beau de marot*. New York: Basic Books.

Please cite this article in press as: Hauser, M. D., & Watumull, J., The Universal Generative Faculty: The source of our expressive power in language, mathematics, morality, and music, Journal of Neurolinguistics (2016), http://dx.doi.org/10.1016/j.jneuroling.2016.10.005

Hofstadter, D. R., & Sander, E. (2013). *Surfaces and Essences*. New York: Basic Books.
Hurford, J. R. (1987). *Language and number: Emergence of a cognitive system*. Cambridge, MA: Blackwell.
Husbands, P. (2014). Robotics. In K. Franklish, & R. W (Eds.), *Artificial intelligence* (pp. 269–295). Cambridge: Cambridge University Press.
Iacoboni, M. (2009). Neurobiology of imitation. *Current Opinion in Neurobiology, 19*, 661–665.
Izard, V., Pica, P., Spelke, E. S., & Dehaene, S. (2008). Exact equality and successor function: Two key concepts on the path towards understanding exact numbers. *Philosophical Psychology, 21*, 491–505.
Katz, J., & Pesetsky, D. (2011). The identity thesis for language and music. *Lingbuzz*, 1–86. http://ling.auf.net/lingbuzz/000959.
Klessinger, N., Szczerbinski, M., & Varley, R. (2007). Algebra in a man with severe aphasia. *Neuropsychologia, 45*, 1642–1648.
Koelsch, S., Rohrmeier, M., Torrecuso, R., & Jentschke, S. (2013). Processing of hierarchical syntactic structure in music. *Proceedings of the National Academy of Sciences, 110*, 15443–15448.
Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–146). New York: Wiley Press.
LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*, 436–444.
Lennenberg, E. H. (1967). *Biological foundations of language*. New York: John Wiley & Sons, Inc.
Leslie, A. M., Gelman, R., & Gallistel, C. R. (2008). The generative basis of natural number concepts. *Trends Cognitive Science, 12*, 213–218.
Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in the area of Broca: An MEG study. *Nature Neuroscience, 4*(5), 540–545.
Marler, P. (1970). Birdsong and speech development: Could there be parallels? *American Scientist, 58*, 669–673.
Martin, D. (1958). *Computability and unsolvability*. New York: McGraw Hill.
Mikhail, J. (2011). *Elements of moral cognition*. New York: Cambridge University Press.
Monte, O. D., Schintu, S., Pardini, M., Berti, A., Wassermann, E. M., Grafman, J., et al. (2014). The left inferior frontal gyrus is crucial for reading the mind in the eyes: Brain lesion evidence. *Cortex, 58*, 9–17.
Monti, M. M., & Osherson, D. N. (2012). Logic, language and the brain. *Brain Research, 1428*, 33–42.
Monti, M. M., Parsons, L. M., & Osherson, D. N. (2012). Thought beyond language neural dissociation of algebra and natural language. *Psychological Science, 23*(8), 914–922.
Moro, A. (2013). On the similarity between syntax and actions. *Trends Cognitive Science, 18*, 1–2.
Murphy, R. A., Mondragon, E., & Murphy, V. A. (2008). Rule learning by rats. *Science, 319*, 1849–1851.
Newport, E. L., Hauser, M. D., Spaepen, G., & Aslin, R. N. (2004). Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology, 45*, 85–117.
Pallier, C., Devauchelle, A. D., & Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. *Proceedings of the National Academy of Sciences, 108*, 2522–2527.
Petkov, C. I., & Jarvis, E. D. (2012). Birds, primates, and spoken language origins: Behavioral phenotypes and neurobiological substrates. *Frontiers in Evolutionary Neuroscience, 4*, 1–24.
Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science, 306*, 499–503.
Pinker, S. (2007). *The stuff of thought: Language as a window into human nature*. New York: Viking Press.
Pulvermüller, F., & Fadiga, L. (2010). Active perception: Sensorimotor circuits as a cortical basis for language. *Nature Review Neuroscience, 11*, 351–360.
Ramachandran, V. S. (2010). *Mirror neurons and imitation learning as the driving force behind "the great leap forward" in human evolution*. http://www.edge.org/3rd_culture/.
Ravignani, A., Westphal-Fitch, G., Aust, U., Schlumpp, M. M., & Fitch, W. T. (2015). More than one way to see it: Individual heuristics in avian visual computation. *Cognition, 143*, 13–24.
Rawls, J. (1971). *A theory of justice*. Cambridge: Harvard Belknap Press.
Rizzi, L. (2009). Some elements of syntactic computations. In D. Bickerton, & E. Szathmáry (Eds.), *Biological foundations and origins of syntax* (pp. 63–87). Cambridge, MA: MIT Press.
Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research, 3*, 131–141.
Roberts, I. (2011). Comments and a conjecture inspired by fabb & halle. In P. Rebuschat, M. Rohrmeier, J. Hawkins, & I. Cross (Eds.), *Language and music as cognitive systems*. Oxford: Oxford University Press.
Roeper, T. (2007). *The prism of grammar*. Cambridge: MIT Press.
Rogers, J., & Hauser, M. D. (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 H. Van der Hulst (Ed.), *Recursion and human language* (pp. 213–232). Berlin: Springer-Verlag.
Searls, D. B. (2012). The language of the genes. *Nature, 420*, 211–217.
Siegelmann, H. T. (1995). Computation beyond the turing limit. *Science, 268*, 545–548.
Simon, H. A. (1969). *The sciences of the artificial*. Cambridge: MIT Press.
Sonnweber, R., Ravignani, A., & Fitch, W. T. (2015). Non-adjacent visual dependency learning in chimpanzees. *Animal Cognition, 18*, 733–745.
Spelke, E. S. (1994). Initial knowledge: Six suggestions. *Cognition, 50*, 431–445.
Spelke, E. S. (2016). Core knowledge and conceptual change: A perspective on social cognition. In D. Barner, & A. S. Baron (Eds.), *Core knowledge and conceptual change* (pp. 279–300). New York: Oxford University Press.
Thomas, K. A., DeScioli, P., De Freitas, J., & Pinker, S. (2016). Recursive mentalizing and common knowledge in the bystander effect. *Journal of Experimental Psychology: General*, 1–9.
Truswell, R. (2012). *Constituency and bonobo comprehension* (pp. 1–39). unpublished manuscript.
Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society, 42*, 230–265.
Udden, J., & Bahlmann, J. (2012). A rostro-caudal gradient of structured sequence processing in the left inferior frontal gyrus. *Philosophical Transactions of the Royal Society, B, 367*, 2023–2032.
de Waal, F. (2016). *Are we smart enough to know how smart animals are?* New York: W. W. Norton & Company.
Watumull, J. (2012). A Turing program for linguistic theory. *Biolinguistics, 6.2*, 222–242.
Watumull, J. (2015a). A Turing program for linguistic theory. *Biolinguistics, 6*(2), 222–242.
Watumull, J. (2015b). *The linguistic turing machine*. PhD thesis. Cambridge University.
Watumull, J., Hauser, M. D., & Berwick, R. C. (2014a). Conceptual and methodological problems with comparative work on artificial language learning. *Biolinguistics, 8*, 120–129.
Watumull, J., Hauser, M. D., Roberts, I. G., & Hornstein, N. (2014b). On recursion. *Frontiers in Psychology, 4*, 1–7.
Wellman, H. M. (2014). *Making minds*. Oxford: Oxford University Press.
Wimmer, H., & Perner, J. (2002). Beliefs about beliefs: Representation and constraircing function of wrong beliefs in young children's understanding of deception. *Cognition, 13*, 103–128.
Winston, P. H. (2012). The right way. *Advances in Cognitive Systems, 1*, 23–36.
Wolfram, S. (2013). The importance of universal computation. In S. B. Cooper, & J. van Leeuwen (Eds.), *Alan turing* (pp. 44–49). New York: Elsevier.
Wood, J. N., Glynn, D. D., & Hauser, M. D. (2007a). The uniquely human capacity to throw evolved from a non-throwing primate: An evolutionary dissociation between action and perception. *Biological Letters, 3*, 360–364.
Wood, J. N., Glynn, D. D., Phillips, B. C., & Hauser, M. D. (2007b). The perception of rational, goal-directed action in nonhuman primates. *Science, 317*, 1402–1405.
Wynn, K. (2002). Addition and subtraction by human infants. *Nature, 356*, 749–750.

Please cite this article in press as: Hauser, M. D., & Watumull, J., The Universal Generative Faculty: The source of our expressive power in language, mathematics, morality, and music, Journal of Neurolinguistics (2016), http://dx.doi.org/10.1016/j.jneuroling.2016.10.005

Yang, C. (2013). Ontogeny and phylogeny of language. *Proceedings of the National Academy of Sciences, 10*, 6324–6327.
Yang, C. (2016). *The linguistic origin of the next number* (Manuscript).
Young, L., & Dungan, J. (2012). Where in the brain is morality? Everywhere and maybe nowhere. *Social Neuroscience, 7*, 1–10.
Zylberberg, A., Dehaene, S., Roelfsema, P. R., & Sigman, M. (2011). The human turing machine: A neural framework for mental programs. *Trends in Cognitive Science, 15*, 293–300.