Psychological Research (2005) 69: 316–329
DOI 10.1007/s00426-004-0205-6

ORIGINAL ARTICLE

Pierre Perruchet

Statistical approaches to language acquisition and the self-organizing consciousness: a reversal of perspective

Received: 18 October 2004 / Accepted: 26 November 2004 / Published online: 5 March 2005

Abstract Recent years have seen the upsurge of a new approach to language that moves away from the rule-based conventional framework. In this approach, mostly supported by the success of connectionist models, children learn language by exploiting the distributional properties of the input. It is argued in this paper that, in the same way as conforming to rules does not imply the existence of mental rules, conforming to statistical regularities does not imply that statistical computations are performed mentally. Sensitivity to statistical regularities can alternatively be conceived of as a by-product of the recurrent interplay between the properties of the current conscious content and the properties of the linguistic and extralinguistic environment. The validity of including the content of conscious experiences in an otherwise standard dynamical approach rooted in the notion of self-organization is discussed.

Introduction

If we look at the activity of an ant colony, it appears that, among other remarkable organizational features, the ants use the shortest paths to join their anthill to the food sources. Each ant displays some elementary characteristics, such as the propensity to follow a trail of pheromones and to strengthen this trail by depositing on it the same chemical substance. However, at first glance, those characteristics do not explain the selection of the shortest path by the colony. This selection is seemingly due to unrelated and more sophisticated abilities, which, could be posited, are owned only by some kind of "surveyor" ant. However, the picture changes dramatically when a dynamical perspective is endorsed. The now well-known dynamical explanation for the selection of the shortest path is as follows. The ants first randomly forage for food, and come back to the anthill when they have discovered a food source. So doing, some fortunate ants naturally find the food through a more direct path than others. The ants that have discovered the shortest path will cover the back and forth distance between the anthill and the food source fastest, and hence, more often than other ants in the same amount of time. Because the ants deposit pheromones throughout their walk, the accumulation of pheromones on the path of the lucky ants is fastest, hence progressively attracting the other ants. Thus, the basic features of an ant's momentary behavior, namely the propensity to follow and strengthen a pheromone trail, account for the selection of the shortest path by the colony, provided the dynamical interplay between those properties and a ubiquitous property of the physical world, namely that a shorter path is covered more often than a longer path in a given period, is considered. There is no place for an external planning agent, hence the term "self-organization" used to describe this form of regulation.

The position advocated in this paper is that the same reasoning is valid for the production and understanding of language. Any English-speaking adult is able to draw meaning from the sets of phonemes composing a spoken sentence such as "Look, a kangaroo." However, at first glance, the properties of any fleeting conscious experience (such as, for instance, the limited number of phonemes that can be held in a single attentional focus) are unable to even begin to provide an explanation for the fact that a sequence of sounds turns out to elicit the conscious apprehension of sentence meaning. Understanding is seemingly due to the action of some unrelated and more sophisticated unconscious processors. My claim is that the elementary properties of conscious experiences are sufficient to account for the understanding of such a sentence, provided the dynamical interplay of those properties with those of language throughout

P. Perruchet
Université de Bourgogne, LEAD/CNRS, Pole AAFE,
Esplanade Erasme, 21000 Dijon, France
E-mail: pierre.perruchet@u-bourgogne.fr

development is considered. My framework is borrowed from the dynamicist approach, the application of which to human behavior is far from new (e.g., Thelen & Smith, 1994). The present application shares the same general principles, except in one crucial aspect. Whereas one of the grounding postulates of the current dynamical theories is the rejection of the notion of representation, and the neglect of consciousness, I consider the notion of conscious representation as central. My claim is that the content of the momentary phenomenal experience is self-organizing. This means that the properties of any conscious experiences on the one hand, and those of language on the other hand, are sufficient to explain the mastery of language, in so far as we consider their interactions over time. (The implications of the self-organizing consciousness model are not limited to language. I have provided with Annie Vinter a full treatment of the issue in Perruchet & Vinter, 2002, 2003.)

I shall discuss in the final section the legitimacy of including conscious representations within a dynamical view. Before that, it must be shown how such an explanation can work. I shall start from the Chomskyan approach to language at the end of the 1950s. The distance between this approach and my final objective may seem insurmountable, insofar as Chomsky's theory is certainly the conceptual system the furthest away from the idea behind the dynamical systems' models. I propose to split this distance in two. In a first step, I will borrow from the extant literature the idea that most, if not all, of language production and comprehension, which Chomsky saw as the end-product of an innate rule-based system, can be viewed rather as a learned sensitivity to the statistical regularities embedded in language. The second step is at the core of the thesis. The claim is that the behavioral sensitivity to statistical regularities is not necessarily the product of statistical computations, as posited in the current statistical approaches. Instead, I shall argue that the same phenomenon can be accounted for as successfully, and perhaps even more successfully, when conscious experience is considered within a dynamical perspective. Each of the two steps will now be considered in turn.

Step 1: From rule-based to statistical approaches in the language domain

The emergence of a statistical approach to language in recent years has been documented extensively elsewhere (e.g., Redington & Chater, 1998; Seidenberg & MacDonald, 1999). However, a brief outline of this research is useful here, because the departure from rule-based conceptions instantiated in the statistical approaches prefigures in some sense the departure from statistical approaches that I shall suggest in the second part of this paper.

The innatist Chomskyan perspective

At the root of Chomsky's theory is his focus on a formal description of language syntax. Chomsky (e.g., 1957) described a hierarchy of grammars of increasing complexity. At the lowest level of complexity are the finite state grammars (FSG), which generate sequences by concatenating a set of elements (states) while following pre-specified transitional probabilities. However, according to Chomsky, the use of human language requires mastery of the next level in the complexity hierarchy, termed the "phrase structure grammar" (PSG). In addition to the concatenation of items like an FSG, a PSG allows the embedding of strings within other strings, thus generating complex hierarchical structures. An instance of embedding in English is "the rat the cat ate stole the cheese," in which one relative clause ("the cat ate") is nested within the sentence ("the rat stole the cheese"). Phrase structure grammars, Chomsky asserts, cannot be learned. This is because the information available in the linguistic utterances is insufficient, a claim known as the "poverty of the stimulus" argument. Indeed, the learner would need to be informed about what is grammatically unacceptable to reject incorrect hypotheses. Now, Chomsky argues that negative evidence is lacking, or at least too limited to make this form of learning plausible. As a consequence, syntactical rules are construed as innate, at least in their very abstract forms, which are held to be universal.

These proposals are still advocated on occasion. For instance, Peña, Nespor, Bonatti, and Mehler (2002) presented what they construed as a demonstration that detection of nonadjacent dependencies, which is a prerequisite to dealing with a PSG grammar, needs rule-based language-specific mechanisms, and Fitch and Hauser (2004) claimed to have provided evidence that the mastery of a PSG distinguishes human from nonhuman primates. However, these conclusions were weakened by major theoretical and methodological drawbacks (see Perruchet, Tyler, Galland, & Peereman, 2004, for a reappraisal of Peña et al., and Perruchet & Rey, in press, for a reappraisal of Fitch & Hauser). By and large, the early claims of Chomsky are no longer the focal point of recent debates. They are undermined by at least two general classes of arguments. First, the initial statement that human languages need the mastery of PSG grammars has been challenged. Limited output from a PSG can always be approximated by an FSG, and recent work in linguistics tends to consider that much of language is a finite state (e.g., Sproat, 2002). Second, people's ability to master syntactically complex sentences may have been overemphasized in the Chomskyan tradition. For instance, even simple spontaneous oral productions are rarely error-free, and it is fairly difficult to capture the syntactic structure of a complex sentence whenever semantics cannot help. Let us add simply one more embedding to our instance of center-embedding above, and we get the oft-cited "the rat the cat the dog chased ate stole the cheese," which is unintelligible to most English speakers, as Miller and Chomsky (1963) noted themselves about a very similar example. The literature on self-embedding is entirely devoted to accounting for why self-embedded structures are *not* manageable whenever the depth of embeddings exceeds one or two, even when semantic biases are available (e.g., Gibson & Thomas, 1999).

This does not mean that the influence of Chomsky's views has gone away. However, most recent studies committed to this tradition argue in fact for a somewhat toned-down position. The main claim is still that language processing is the realm of rules, but the abstract principles of a putative Universal Grammar have been traded for the idiosyncratic rules of the grammar of specific languages.[1]

The best-known target of recent studies is past tense formation in English. The claim is that the formation of the past tense of regular verbs is proof of the application of a rule, which may be phrased as "append *-ed* to the base of regular verbs." An especially striking piece of evidence for this hypothesis is the U-shaped evolution of the past-tense formation of irregular verbs in childhood language. Children may begin to use the correct irregular form (e.g., went), and then erroneously apply a regular transformation (e.g., goed), before finally producing again the correct version. Although this phenomenon of over-generalization is in fact far less frequent than parents' informal observations would lead us to believe, it is argued to be the hallmark of rule guidance. Indeed, it is proof that the *-ed* suffix does not concern a specific token, but a variable, namely the syntactic class of regular verbs (regularization is conceived as the erroneous inclusion of some irregular verbs within this variable).

The rise of a distributional approach

The most influential challenge to this traditional view is undoubtedly the emergence of connectionist modeling. In response to the claim that mental rules are required to account for linguistic performance, connectionist researchers mount an empirical argument: Models including no pre-wired rules, and composed only of a network of associations, the weights of which gradually vary as a function of the input, can learn some fundamental aspects of language. Thus, Rumelhart and McClelland (1986) showed that a connectionist model is able to learn the past tense of both regular and irregular verbs from exposure to a phonological representation of those verbs, and also displays the transitory appearance of over-generalization.

How does a connectionist model proceed? Without entering into technical considerations, a connectionist model works by computing statistics. In a few cases, the end-result of training a network is exactly equivalent to the computations of standard statistical coefficients. For instance, one of the most widespread connectionist models in the research on language is the Simple Recurrent Network, or SRN (Elman, 1990). When an SRN is asymptotically trained with a sequence of events comprising only first-order dependency rules, the output activation of the event *e2* when *e1* is displayed as input is equal to the transitional probability P(e2/e1) (Perruchet & Peereman, 2004). When the sequence includes higher-order dependency rules, the results no longer match a standard statistical coefficient, and this is the most general outcome. However, from a formal standpoint, any network can be construed as a statistical device. Obviously, a network does not compute statistics as would a statistician. The pattern of its weights is shaped by the distributional properties of the input, and it thereby reveals those properties, in a similar way to which the depth of a nail reveals, under specified conditions, the number of hits that it has received (except that, needless to say, the involved operations are more sophisticated than a simple summation of individual events).

It remains to be understood why computing statistics amounts to following rules. Research on implicit learning may provide some insights. This field of enquiry is not specially oriented toward language, but is nevertheless relevant to the issue at hand. In studies on implicit learning, participants are first exposed to material structured by *arbitrary* rules, and then a test is designed to reveal participants' acquired sensitivity to the structure. It has been reported that participants can become sensitive to very complex arbitrary rules. A striking example was provided by Lewicki et al. (1988), in which a complex set of second-order dependency rules determined the order of events within subsequences that were hidden in a continuous sequence. Participants' reaction times proved to be sensitive to this structure. However, Perruchet, Gallego, and Savy (1990) showed that a by-product of these complex rules was a frequency bias in small units comprising two or three events, irrespective of the location of the hidden subsequences. Taking advantage of the control over the material allowed by its arbitrary nature, they were able to demonstrate that participants actually learned those elementary frequency biases, instead of the rules. Other illustrative examples are provided by other implicit learning situations (e.g., Wright & Burton, 1995).

In a similar way, it can be shown that a connectionist network is able to learn the correct past tense for regular and irregular verbs, because regular and irregular verbs differ according to the distribution of their phonological features. The U-shaped curve of acquisition described

[1] The shift from Universal to idiosyncratic grammar has obvious consequences with regard to an innatist position. A specific rule needs to be learned. However, there is no clear explanation of how a rule can be learned. For instance, when referring to a monograph of Marcus et al. on over-generalization, Bybee (1995, pp.449–450) noted that the authors "do not address the question of exactly how the rule is extracted and deposited in a separable module. Thus, despite the fact that the relevant section in their paper is named "How might a regular rule be learned?", the question of how the restructuring from lexical schema to symbolic rule takes place is actually not addressed!"

above stems from the way the verbs are presented to the network, which is intended to reflect the way children are exposed to the language. Irregular verbs are fewer in number, but they are also much more frequent than regular verbs. Therefore, irregular verbs are presumably the first to which children are exposed, and hence the first that they learn. The subsequent exposure to a large number of regular verbs would be responsible for the stage of over-generalization, before further training allows again the correct formation of irregulars. The initial study by Rumelhart and McClelland (1986) has been heavily criticized on a number of aspects, and hot debates have risen since then, with connectionist modelers addressing in succession the challenges raised by the advocates of a rule-based view (which, in this context, is known as the "dual-route" position, since the past tense for irregular verbs cannot be encompassed within rules and hence, requires other learning mechanisms).

Although the examination of this literature exceeds the scope of this paper, it is interesting for our concern to focus on a representative episode of the debate, because it illustrates an important characteristic of statistical approaches. As noted by Pinker and Prince (1988), an obvious limitation of the connectionist models' use of phonological information is their inability to account for the correct past tense inflection of homophone verb pairs that have different past tense forms (e.g., brake → braked, break → broke). However, this concerns the insufficiency of phonological cues, not a limit inherent to the connectionist approach. Ramscar (2002) has shown that semantic cues could complement phonological cues, and solve the homophone problem. In addition, Ramscar was able to demonstrate that actual participants are able to exploit semantic similarity to form the past tense of pseudo-verbs. The general point is the following. A statistical approach generally reaches an acceptable level of success through the exploitation of a variety of cues, which may be situated at levels of description that conventional linguistics considers as separate, such as syntax and semantics in the last example.

Nevertheless, it remains that a statistical description often provides only an approximation of rules, and this could be construed as detrimental for its explanatory power. Although the point remains a matter of debate, two remarks are worth making. First, the entire literature on implicit learning shows that the approximation provided by statistical mechanisms is good enough to have misled researchers for years. Indeed, the recurrent story in this area of study has been that the first publications exploring a new situation were put forth as evidence that participants had abstracted the rules underlying this situation. Subsequent reappraisals have not shown that performances were poorer than previously shown, but instead that the very same performances that were put forward as a proof of rule discovery were in fact due to other mechanisms. But there is a second remark, which goes beyond the first one. Claiming that statistical approaches provide only an approximation of rules is an argument against statistical approaches only in so far as it is taken for granted that human behavior follows rules perfectly. If the departures of statistical predictions from a genuine rule-based behavior are also present in actual human behavior, then the argument turns back against rule-based theories. Now, most empirical studies aimed at pitting rule-based and statistical approaches against one another have exploited these minor differences, and their results have provided fuel for statistical approaches. For instance, transfer to a new situation in which a rule applies is never as perfect as a rule-based theory would predict (e.g., Pacton, Perruchet, Fayol, & Cleeremans, 2001), because behavior is sensitive to the distributional properties of the original input. Even in conditions where a simple rule exists, and is taught at school, such as the "s" to mark regular plurals in English, participants do not seem to rely on it (Kemp & Bryant, 2003). To account for these observations, rule-based theories need to add to their (costly) postulate for rule guidance other (costly) postulates for mechanisms preventing the application of rules[2] (for an illustration about Marcus' account of past-tense formation, see Bybee, 1995, p. 448)

Thus, recent research demonstrates that a distributional approach to language is now gaining increasing plausibility. Besides the specific examples evoked above, studies have focused on cases of derivational morphology other than past tense production, the formation of syntactic categories (e.g., Redington, Chater, & Fintch, 1998), and the development of word meaning (e.g., McDonald & Ramscar, 2001), among other domains. I do *not* contend that current connectionist models account for all aspects of language, even though, for instance, connectionist models of recursivity, which Chomsky held as the hallmark of human syntactic competences, have been proposed (Christiansen & Chater, 1999). I limit my claim to the uncontroversial observation that, over the last two decades, a rule-based view of language has gradually waned, while connectionist models are increasingly able to account for various aspects of language behavior. Of course, this does not mean that the human mind is unable to manage rules. For instance, it is undisputable that linguists or grammarians are able to infer the rules of the language, as any scientist is able to infer the rules of his/her own domain of inquiry. It is also likely that we use rules on some occasions to consciously control our behavior. The point is that the day-to-day adaptation to the linguistic environment seemingly does not rely on rules.

[2] This is reminiscent of the traditional Chomskyan distinction between (inferred) competence and (observed) performance, and its associated requirement for researchers to account for the extraneous limitations that prevent unlimited competence from fully translating into performance.

Step 2: From statistical computation to the concept of self-organizing consciousness

The previous section suggests that researchers have become increasingly aware that the sensitivity of humans to the rules of their language does not automatically imply that rules exist somewhere in the mind. The exploitation of the distributional properties of the language can also account for linguistic performance. The present section illustrates that this reasoning can be applied recursively. Sensitivity to the distributional properties of the language does not automatically imply that statistics have been computed. I intend to show how introducing consciousness in an otherwise standard dynamical perspective accounts for this phenomenon as well.

It is quite obvious that the shift from a Chomskyan perspective to a connectionist framework represents a considerable move toward a dynamical approach. A key point is that most connectionist models do not attempt to account for full-blown adult competence, a project that entails an obligatory recourse to innate abilities. Instead, they show how adult adaptive behavior can be achieved through the progressive transformation of a network under the pressure of environmental regularities. On these bases, connectionist models are sometimes considered as instances of dynamic approaches (e.g., Smolensky, 1988).

Whether or not dynamicism has to be thought of as a new paradigm with regard to connectionism, my argument is the following. The core of a dynamical approach is that adapted behavior emerges as the end-result of the continuous interplay between the properties of the organism and those of the external environment. The current distributional approaches undoubtedly exploit the properties of the world, but I contend that they are often defective with regard to the properties, and especially to the processing constraints, of the learners.

About constraints

Admittedly, neural networks implement constraints. The most fundamental are the absence of explicit rule abstraction, reasoning, and other high-order cognitive abilities. In fact, the interest of the approach lies in their very absence. But how are the other constraints and resources theoretically motivated? For most connectionist modelers, the properties that an artificial network must implement are biological in nature, being dictated by the organization of the human brain. In truth, extant connectionist models only loosely respect the organization of the neural system. Certainly, the performance of a network depends on a set of structural (e.g., number of hidden layers, number of nodes, learning algorithms) and parametric (e.g., learning rate, momentum) properties, but it is nearly impossible to match those properties to those of human brains. For instance, it is known that limiting the number of nodes in hidden layers has the (often desirable) effect of promoting generalization, but this constraint is entirely artificial and ad hoc. The number of nodes is certainly not determined by empirical evidence that human learners would be unable to engage a larger pool of neurons in some putative layer in order to perform a given task.

However, my main concern lies not in the haziness with which biological constraints are implemented, but instead with the choice of these kinds of constraints in the first place. I think that the project of implementing biological constraints in a model devised to account for mental activities is questionable in itself. What I mean here can be made clearer by taking a broader perspective. Searle wrote:

> If we think of the making-up of the world, then of course everything in the world is made of particles, and particles are among our paradigms of the physical. If we are going to call anything that is made up of physical particles physical, then, trivially, everything in the world is physical. But to say that is not to deny that the world contains points scored in football games, interest rates, governments, and pains. All of those have their own way of existing—athletic, economic, political, mental, etc. (Searle, 1992)

Even though "everything in the world is made of particles," no science addresses the challenge of understanding a football match from the activity of billions of physical particles. A point in a match of football is determined by specific causal events, which have to be distinguished from the causal relationships relevant in the physics of particles. Likewise, understanding a sentence such as "look, a kangaroo" obviously involves neural activities, such the activation of the cochlear nuclei, the thalamus and the auditory cortex. However, understanding is a mental activity, and as such it is endowed with specific organizational principles that have to be distinguished from those that apply to the neural level. My contention is that neuronal constraints may be no more relevant to describing mental events than principles governing particles' movement are relevant to describing a football match. If we agree with the existence of different levels of description and explanation, the constraints that have to be implemented for a psychological model aimed at accounting for language mastery must be set at the mental level. But what are the relevant constraints inherent to mental life?

Introducing consciousness

If the notion of self-organization has emerged primarily for social insects, it is probably because their processing constraints are unquestionable. If it were possible that ants had sophisticated conceptual and linguistic abilities, there is no doubt that the selection of the shortest path between the anthill and the food source would have been attributed to those abilities, and that the dynamic explanation would not have been discovered yet. The information processing framework that emerged at the end of the 1950s has failed to provide any organizational constraints for the processes regulating mental life in humans. As I have argued with Annie Vinter (Perruchet & Vinter, 2002), this failure is due to the (often) tacit acceptance of a sophisticated cognitive unconscious. Indeed, the main characteristic of the cognitive unconscious is its freedom from the well-known limitations of the conscious/attentional system, such as ephemeral duration, limited capacity, seriality, and relative slowness of processing.

Perruchet and Vinter (2002) explored the explanatory power of the most constrained framework that can be envisioned for mental life. The position taken by Perruchet and Vinter and in the present paper is quite simply that mental life is coextensive with consciousness. That is to say, the existence of any unconscious mental activities is denied. All representations are conscious, and the only operations we perform on those representations are those that compose our momentary phenomenal experience (the view endorsed here is borrowed from Dulany, e.g., 1991, 2002, who coined it as a mentalistic metatheory. However, as Shanon (2002) pointed out, a similar view has been advocated by famous precursors, such as William James and James Gibson).

I am fully aware that the question of consciousness is usually framed in a diametrically opposed way, with the idea that accounting for some behavior without involving conscious thought is a pledge for simplicity and parsimony. This conception follows from the initial postulate of the information processing framework according to which cognitive activities can be either conscious or unconscious. Within this framework, it appears that the conscious mode is endowed with "something more" than its unconscious counterpart, which is usually called the subjective experience. Because this aspect looks somewhat unmanageable within a scientific approach, explaining behavior without introducing consciousness is generally construed as conceptual progress. But this logic is deeply flawed. Decoupling mental activities from their conscious expression provides no advantage. If conscious feeling were an invented construct, such as the phlogiston of eighteenth century physics, eliminating it in favor of a more objective concept would indeed be a step forward. But this is not the case: Consciousness is a fact that is observable by all. Thus, setting aside subjective experience does not eliminate it from the field of to-be-explained phenomena. Furthermore, decoupling mental activities from their conscious expression has a major negative consequence. Conscious and unconscious modes of thought do not only differ with regard to whether they fill the subjective scene or not. They also differ with regard to whether or not they are endowed with processing constraints. Therefore, the neglect of consciousness in the mainstream cognitive tradition, which was intended to improve parsimony, led to the opposite consequence, namely the acceptance of an unconscious device free from any processing limitations. The position advocated here is that the limitations of conscious thought are computationally relevant[3] and must be integrated into any model intended to be psychologically plausible. I intend to show that introducing those constraints within a dynamical approach appears to be surprisingly productive.

Let us now return to our initial problem. In the introduction of this paper, I proposed to account for the understanding of an oral sentence such as "Look, a kangaroo" within a dynamic perspective. This proposal can now be reformulated as that of accounting for the transition between the initial conscious state of a newborn hearing the sequence of sounds, and the final adult's conscious state, where the speech flow is perceived as a sequence of words with an associated meaning. In keeping with a dynamical approach, I intend to show that this transition implies no external rules or principles: It depends only on the interplay between the intrinsic properties of conscious states and the intrinsic properties of the language and more generally, of the linguistic and extra-linguistic environment. Of course, there is no reason to conceive of this transition as one-stepped. In the next section, I focus on a particular stage of the whole process. The starting point is the perception of the speech flow as a sequence of syllables (I assume here that listeners have solved the problem of mapping raw speech to syllables), and the endpoint is the perception of the speech flow as a sequence of words. How the principles at play in this stage of learning can generalize to other stages will be examined later. I also focus on a particular class of processes. A huge number of recent studies have shown that learning to segment natural language involves the exploitation of a large diversity of informational cues, such as the presence of single-word utterances (Brent & Siskind, 2001), and

[3] Consider the following story for a short illustration of the ambiguities inherent in the notion of "computational irrelevance." A cognitive scientist builds a psychological model that implies at some location the computation of $2 \times 2 = 4$. The scientist notes (rightly) that nothing has changed in his program according to whether or not this computation is thought to be accompanied by a conscious experience in the actual subjects he intends to simulate. From this observation, he coins the notion of "computational irrelevance of consciousness." In some later location of his program, he requires that his model now computes the cubic power of 6371, and he is first baffled by the fact that it is highly implausible that actual subjects are able to perform this computation mentally. Fortunately, he realizes that this limitation holds only if it is assumed that this computation is performed consciously. Because the issue of consciousness has been declared computationally irrelevant, this slight glitch is automatically removed, and even turns out to be an advantage: His program succeeds at simulating behavior without ever involving the need for consciousness! This story allows a contrast of two ways of conceiving constraints and parsimony: The conventional one, in which actual subjects are assumed to be able to compute the cubic power of 6371, but which calls for no conscious counterpart for this computation, and the one to which I subscribe, in which actual subjects are assumed to be able to compute only what they do consciously, which is hardly more than 2×2.

a variety of prosodic cues (e.g., Jusczyk, Houston, & Newsome, 1999). Those processes could be integrated in the self-organizational framework, but, for the sake of simplicity, I consider here only the mechanisms at play in a situation deprived of those informational sources. This situation consists in exposing children or adults to an artificial language, composed of imagined words without semantic referent, concatenated without pauses or other prosodic cues.

The case of spoken word segmentation

When we are confronted with material consisting of a succession of elements, each of them matching some of our processing primitives, our phenomenal experience is that of processing this material into small and disjunctive parts comprising a small number of primitives. Chunking, we contend, is a ubiquitous phenomenon, due to the constraints and resources intrinsic to attentional processing and conscious thought. As a case in point, when listeners are faced with an unknown spoken language, with syllables consisting of their conscious primitives, they presumably encode the speech flow by picking up some chunks of a few syllables. Importantly, however, the listener's initial conscious experience consists of a succession of chunks that have only a weak probability of matching the words of the language (Plunkett, 1990, has provided empirical evidence that before solving the segmentation problem, children indeed tend to use lexical units that are either parts of words or sequences of words).

Our proposal is that the final phenomenal experience of perceiving the correct sequence of words emerges through the progressive transformation of the primitives guiding the initial perception of the language, and that this transformation is due to the self-organizing property of the content of phenomenal experience. The basic principle is fairly simple. The primitives forming a chunk, that is those that are perceived within one attentional focus as a consequence of their experienced temporal proximity, tend to pool together and form a new primitive for the system. As a consequence, they can enter as a unitary component into a new chunk in a further processing step. This explains why the phenomenal experience changes with practice. But why do the initial primitives evolve into a small number of words instead of innumerable irrelevant processing units?

This is because the future of the chunk that forms a conscious episode depends on ubiquitous laws of associative learning and memory. If the same experience does not recur within some temporal lag, the possibility that a chunk will act as a processing primitive rapidly vanishes, as a consequence of both natural decay and interference with the processing of similar material. The chunks evolve into primitives only if they are repeated. Thus, some primitives emerge through a natural selection process, because forgetting and interference lead the human processing system to select the repeated parts from all of those generated by the initial, presumably mostly irrelevant, chunking of the material. The importance of this phenomenon becomes clear when viewed in relation to a property inherent to any language. If speech is initially segmented randomly into small parts, then these parts have greater chance of being repeated if they match a word, or a part of a word, than if they straddle word boundaries (for instance, in the sentence "look, a kangaroo," the string "kan/ga" has certainly more chance of occurring later in the speech stream than "a/kan," because "a/kan" is no longer present in expressions such as "...the kangaroo...," "...a little kangaroo...," and so on). Consequently, the primitives that emerge from natural selection due to forgetting and from interference are more likely to match a word, or a part of a word, than a between-word segment.

This account has been implemented in a computer program, Parser (Perruchet & Vinter, 1998; an on-line presentation of the model is available on the internet: http://www.u-bourgogne.fr/LEAD/people/perruchet/SOC.html), and applied to the artificial languages used by Saffran, Newport, and Aslin (1996). Simulations revealed that Parser extracted the words of the languages without any errors well before exhausting the material presented to participants in the Saffran et al. (1996) experiments.[4]

To summarize, I suggest that the discovery of words in the speech signal results from the interaction between one property of language—essentially that the probability of repeatedly selecting the same group of syllables by chance is higher if these syllables form intra-word rather than inter-words components—and the properties of the processing systems—essentially that repeated perceptual chunks evolve into processing primitives, which in turn determine the way further material is perceived.

It is worth noting that Parser exploits the very same distributional properties of the language as do the other models of word segmentation. Let us illustrate the point using the model of word segmentation developed by Brent and Cartwright (1996), in which segmentation is construed as an optimization problem. The principle of the method is akin to establishing a list of all possible segmentations of a given corpus (although the authors used computational tools that prevented the program from proceeding in this way). The choice between possible segmentations is then made in order to fulfill a number of criteria. These criteria are threefold

[4] Although they do not directly address the word segmentation issue, Boucher and Dienes (2003) also explore the possibility that the sensitivity to statistical regularities is not the result of statistical computations on individual elements, but rather the by-product of local representations of chunks of individual elements. In their view, to borrow the title of their paper, there are "two ways of learning associations," one in which chunking is an emergent property of statistical analyses, and the other in which chunking is a primitive process, the result of which amounts to simulating statistical computations. However, they do not equate their chunks with the focal content of phenomenal consciousness, as I do.

(according to the somewhat simplified presentation by Brent, 1996): Minimize the number of novel words, minimize the sum of the lengths of the novel words, and maximize the product of the relative frequencies of all the words. The process of optimization is performed thanks to a statistical inference method, called the "minimum representation (or description) length" method. Needless to say, nothing in Brent's model matches the conscious experience of the learner of a new language. The operations involved in the Brent and Cartwright model, such as the computation of all the possible segmentations of an utterance in order to choose the one responding to pre-specified criteria, far exceed the level of complexity that can be achieved by a conscious operator, whether complexity is assessed in terms of computational sophistication or memory capacity.

Parser is different from the Brent and Cartwright (1996) model for the following reasons. Instead of positing an arbitrary criterion of optimization, Parser finds those criteria in the properties of conscious thought. Brent's model selects the partition that generates the minimum number of different words. Parser exploits a logical corollary. For a given corpus, if a mode of segmentation minimizes the number of different words, it also maximizes the number of repetitions of each word. These two properties are essentially the same, except that detecting the minimum number of different words requires rather complex operations, while selecting the more frequent words is the natural by-product of memory decay (Note here the analogy with our initial example. Selecting the shortest path seems of intractable complexity for the abilities of ants, until we realize that the shortest path is traversed more often in a given unit of time, hence increasing pheromone concentration). Another difference involves the mode of selection. Brent's method amounts to listing all of the possible partitioning of the corpus. Parser relies on the variety generated by successive random drawing to reach the same result. Randomly drawing provisional perceptual units certainly does not allow for an exhaustive examination of the possible partitions, but provides a good approximation of this objective (again the analogy with the discovery of the shortest path in ants is striking, since this path is selected after random explorations). Finally, the length of the words is not an arbitrary decision, but a result of the interplay between the properties of the words and the limited capacity of the attentional focus.

It is worthwhile noting that the constraints inherent to conscious thought cannot be conceived of as prejudicial to model efficiency. Parser works well, not *despite* these constraints, but *thanks* to them. For instance, the fact that attention is limited to the simultaneous perception of a few primitives—a property of the conscious/attentional system usually thought of as a serious handicap—is that very property that offers Parser a set of candidate units. If humans perceived a complex scene as a single unit, Parser's principles would not work. Likewise, forgetting is essential to the functioning of the model because, if it did not forget, then Parser would fail to extract the relevant units from the multiple candidate units processed by the system. Memory breakdown, considered in conjunction with the preventing effect of repetitions, is adaptive, because it turns out that, in any language, a given segment has more chance of being repeated if it matches a word than if it straddles word boundaries. In this context, forgetting allows the selective disappearance of structurally irrelevant units.

Beyond frequency

The most obvious objection to the above proposal may be that the model applies (at best) to word segmentation, and that this stage of learning is only a part of the story. How the view applies to the other aspects of language processing will be discussed in the next section. However, a preliminary problem must be dealt with first. Parser is indisputably quite good at discovering the words in the artificial languages of Saffran and collaborators, but those languages are composed of six trisyllabic words, hence casting doubt on the possibility of extrapolating the segmentation algorithm to natural language. Notably, as presented above, Parser is seemingly responsive to a single statistical property of the input, namely raw frequency, because only frequent cooccurrences resist forgetting (as a case in point, Hunt & Aslin, 2001, wonder about the ability of Parser to discover words in a frequency-balanced task). Now, it is easy to illustrate that a selection exclusively based on frequency does not work well once we leave scaled-down languages. Indeed, in natural languages, many words are less frequent than between-word transitions composing certain expressions (e.g., the strings "kan/ga" or "ga/roo" may be less frequent than the string "air/con" present in the expression "air conditioned").

Before outlining how Parser might overcome this limitation, let us consider how connectionist models solve the word segmentation problem. In fact, connectionist models exploit a more subtle measure of statistical associations. Most of the connectionist models that address the word segmentation issue rely on the SRN, initially proposed by Elman (e.g., 1990; see also Cleeremans, 1993). SRNs are designed to learn to predict the next event of a sequence. To this end, at each time step, the activations of the hidden units are stored in a layer of context units, and these activations are fed back to the hidden units on the next time step (hence the term "recurrent"). In this way, at each step, the hidden layer processes both the current input and the results of the processing of the immediately preceding step, and so on recursively. With the exception of this feature, an SRN works as many networks do, using the back propagation of errors as a learning algorithm. The comparison between the predicted event and the next actual event of the sequence is used to adjust the weights of the network at each time step, in such a way as to decrease the discrepancy between the two events. Elman (1990) presented to such a network unbroken strings of phonemes one at a time, the task being to predict the next phoneme in the sequence. After training, the error curve had a strikingly marked saw-tooth shape. Usually, the beginning of a word coincided with the tip of the teeth. Therefore, an SRN appears able to parse a continuous speech flow into words (for more recent models, see Aslin, Woodward, LaMendola, & Bever, 1996; Christiansen, Allen, & Seidenberg, 1998).

An SRN exploits in fact the differences between the within-words and between-word transitional probabilities. On average, these probabilities are stronger between the intra-word components than between components spanning word boundaries, irrespective of the frequency of co-occurrences between those components (e.g., the probability of "ga" after "kan" is certainly stronger than the probability of "con" after "air," even if "kangaroo" has a lower token frequency than "air conditioned"). As a consequence, a network trained to predict the next syllable will achieve better predictions for a within-word transition than for a transition between the last syllable of a word and the first syllable of the next word, and this characteristic allows an inference of the location of the word boundaries within the speech flow.

Many authors (e.g., Peña et al., 2002; Saffran et al., 1996) consider only the transitional probability when they intend to quantify an association, for reasons that are somewhat unclear. Indeed, transitional probability, although more informative than co-occurrence frequency, provides only part of the relevant information about the tightness of an association. In order to obtain a more reliable measure, the transitional probability between two successive events, e1 and e2 (P(e2/ e1)), must be compared with the probability of e2 when not preceded by e1. Moreover, the strength of an association may also be related to the backward relationship between e2 and e1. Accordingly, the standard measure of correlation, Pearson r, measures the two-way dependency between e1 and e2. With dichotomous data, Pearson r is commonly called " r phi" (r_ϕ), and expressed as:

$$r_\varphi = \frac{ad - bc}{\sqrt{(a+b)*(c+d)*(a+c)*(b+d)}} \qquad (a)$$

where a stands for the number of e1–e2 co-occurrences, b for the number of occurrences of e1 followed by an event different from e2, c for the number of occurrences of e2 preceded by an event different from e1, and d for the number of events comprising neither e1 nor e2. There is at least some preliminary evidence that sensitivity to the structure of the language relies on a contingency measure, more than on co-occurrence frequency or even transitional probability (Perruchet & Peereman, 2004).

The relative complexity of this formula might suggest that Parser is unable to account for the human sensitivity to genuine contingency. However, this conclusion turns out to be incorrect (Perruchet and Peereman, 2004). The sensitivity of Parser to contingency relationships is not due to the addition of ad hoc mechanisms, but quite simply to the fact that Parser implements the ubiquitous properties of associative mechanisms. It has been known for years that forgetting may be due, in part, to the decay of traces over time, but also (and perhaps essentially) to interference. The sensitivity to interference makes Parser sensitive to contingency, for a simple reason. Let us consider an AB unit. Whenever A or B are presented in contexts other than the unit AB, they interfere with AB, resulting in a decrement of the weight of AB. Now, if AB is strongly contingent, this means that neither A nor B will be frequent in other contexts, and hence AB will receive no, or only a small amount of interference. Conversely, if AB is not (or negatively) contingent, this means that A and B are frequent events out of the AB unit, and therefore, interference will strongly reduce the weight of AB. Interestingly, modulating the parameter of interference makes Parser more sensitive to the contingency between events than to the relative frequency of co-occurrence (when the interference value is high) or more sensitive to the co-occurrence frequency than to the contingency (when the interference value is low).

To summarize, it would be misguided to consider that because Parser implements simplistic principles, it is underpowered whenever the information contained in the input becomes more sophisticated than raw frequency. On the contrary, Parser exploits an even more sophisticated measure of association than an SRN, which is limited to the computation of transitional probabilities. This outcome confirms that Parser is not limited to dealing with the Saffran et al. (1996) artificial language for which it was initially designed, and extends further Parser's ability to learn from various situations, as revealed in recent studies (Peereman, Dubois-Dunilac, Perruchet, & Content, 2004; Perruchet et al., 2004; Perruchet, Vinter, Pacteau, & Gallego, 2002).

Broadening the scope of the self-organizing consciousness model: Word-object mapping

Even if we consider that the prior sections demonstrate convincingly how the words composing a sentence such as "Look, a kangaroo" can be individuated by self-organizational processes, it is quite obvious that the understanding of language cannot be reduced to the segmentation into word units. Parser, as an instantiation of the self-organizing consciousness (SOC) model, is by construction unable to go further. However, the question is: Are the principles inherent to the SOC model able to account for more than word segmentation?

For instance, understanding language supposes that a word is mapped to its referent. Assuming that "kangaroo" has been isolated as a relevant unit, how can the word be mapped onto the animal that it stands for?[5] It may be argued that the relation between a word and its referent is a form of association, and hence, that a model able to extract contingency is well suited to achieving such a mapping. However, a problem arises if we intend to generalize the way Parser extracts contingency. In speech, the number of possible units is limited by the sequential nature of the auditory signal. For instance, a three-syllable message can be composed of three one-syllable words, two words consisting of one and two syllables, or one three-syllable word. This results in only four possibilities, and hence a model in which the correct choice is selected by elimination of the others is likely to succeed. However, the number of candidate units is much greater with a multidimensional presentation than with purely linguistic material, due to combinatorial explosion. In real life, infants may capture within a single attentional focus a virtually unlimited number of different chunks, each composed of unrelated components of the environment, such as a sound frequency together with the orientation of a segment of the visual field. Under these conditions, the formation of relevant units through a selective process would appear to be an intractable problem. To illustrate, let us consider the following question (paraphrasing the question raised by Karmiloff-Smith, 1992, p. 40). When an adult points toward a kangaroo and says "look, a kangaroo," how can the child pair the word "kangaroo," rather than, say, "look", with the whole animal, rather than with the kangaroo's ears, the color of the kangaroo's fur, or the background context?

This problem, again, finds a solution in the properties of consciousness, and more precisely in the assumption that units are formed by the concurrent attentional processing of a small number of primitives. The point is that infants' attention is not dispersed randomly all around, but is instead captured by an array of stimuli sharing specific properties. One of these properties, for instance, is novelty (e.g., Kagan, 1971). If, at a given moment, several primitives are new for the infants, it is highly probable that these primitives are processed conjointly in the attentional focus, hence forming a new unit. Now, if several primitives are new for a subject, there is also a good chance that they will be the components of one and the same meaningful unit, such as an actual object. The same line of reasoning may be followed with movement. It has been established that infants' attention is attracted by a moving display (e.g., Vinter, 1986). If several elementary features move concurrently, they have a high probability of being both attended to by infants, and of belonging to same real object (of course, many objects do not move; however, it is imaginable that the perceived movement generated by eye displacement in a 3D visual field makes it possible to generalize this phenomenon to motionless objects). Returning to the question above, what is likely to become associated is what captures the infant's attention, that is, essentially, what is new and/or moving. Thus, it is highly probable that the infant's attention is focused on the animal, which moves as a whole, rather than on one of its parts, or on the other elements of the context, which are presumably both more familiar and motionless. Considering now the auditory input, "kangaroo" is presumably newer than "look," because "look" has been associated with many contexts before. As a consequence, it is highly probable that "kangaroo," rather than "look," enters into the momentary attentional focus and become associated with the animal.

Of course, the process of mapping as described above may sometimes fail. The infant may be quite familiar with kangaroos, and surprised by the gray color of the fur of this specific kangaroo. We predict that, in this case, the infant would mismap the word "kangaroo" to the color gray. It is worth noting first that, in real world settings, this situation may be infrequent, because adults would tend to spell out what is presumably the most novel for infants, and more generally, what they infer to be their present object of attention. Furthermore, errors of mapping do in fact occur during language development. What is needed is not a theory predicting a perfect mapping from the outset, but a theory able to predict observed performance. Our model of learning is precisely adapted to extracting signals from noise. In general, the correct mapping will be the final outcome, because the infants will hear "kangaroo" for animals that are not gray, and will hear "gray" for objects or animals that are not kangaroos.

To summarize, our model of learning, initially applied to the word extraction issue, provides a new potential account of infants' basic ability to map words and objects. The apparent problem posed by the unmanageable number of potential units that can be initially perceived again finds a solution when the constraints and resources of conscious thought are considered. Attention is naturally captured by a tightly defined set of events that, as illustrated above, are precisely relevant to the problem at hand.

Beyond the word units

Even if an investigation of the problem of isolating words in speech and finding their meanings is on the way to a solution, we have still not accounted for the direct perception of the meaning of a sentence such as "Look, a kangaroo." What is required is some syntactic abilities. In fact, the statistical approaches evoked in the

[5] This presentation is obviously an over-simplification. I do not intend to mean that learning a language amounts to successively isolating the words from the speech flow, to map the word to its meaning, and so on. Those processes are considered here as independent steps for the sake of clarity, but it would be in keeping with a dynamic approach to consider them as closely interconnected processes.

preceding section have opened the way. If a large part of syntactic knowledge can be described as a set of associations and captured by a neural net, then any model able to extract contingencies must be able to reach the same outcome, at least in principle. However, the same problem again arises: How are the correct associations discovered? The difficulty no longer lies in capturing the relevant features from a multidimensional array, as above, given that language is sequentially organized. Instead, the problem now is to relate events that are not necessarily contiguous.

Let us consider the case of a sequence AXB in which A and B are associated irrespective of the length and nature of X. Such nonadjacent dependencies are found at different levels, from the subsyllabic level (e.g., the short versus long pronunciations of vowels according to the presence of a silent *e* ending, irrespective of the intermediary consonant, as in CAP—CAPE, CAR—CARE; Stanback, 1992) to morphosyntactic relationships (e.g., between auxiliaries and inflectional morphemes, as in "is writing," irrespective of the verb stem) and hierarchical structures (e.g., in center-embedded sentences, such as "the rat the cat ate stole the cheese"). Parser is a priori unable to capture the relation, because new units can only be formed between contiguous elements. However, the general principle that PARSER instantiates is that new units result from the processing of a few primitives within the same attentional focus. When people encounter sequential material, the most simple assumption is that each attentional focus embraces a small number of contiguous elements. In artificial, meaningless languages, there is no obvious reason to expect a different type of chunking. However, there are clearly no functional or structural constraints here. Each of us commonly mixes present and past events in his/her current phenomenal experience. This would be in keeping with our general approach of assuming that a new unit may be composed of spatially or temporally remote events, provided that there was some reason for those events to become associated in phenomenal experience.

Available experimental evidence indeed suggests that learning nonadjacent dependencies is possible only when there is some reason to pay attention to the associated events. For instance, it appears that learning A–C relationships is only possible (or at least much easier) when the AXC units are displayed as perceptually separate sequences, instead of being displayed within a continuous speech stream (Gomez; 2002, Perruchet et al., 2004). Now, both everyday experience and experimental evidence (e.g., Cowan, 1991) suggest that the start and the end of a sequence captures more attention than the intermediary events. Thus, it is likely that when the auditory stream is perceived as a succession of short sequences, participants pay more attention to their first and last elements than to their middle one, and then encode those elements as well as the relevant positional information. This prompts the formation of AXC units, where A and C are specific elements and X stands for unnoticed events. This interpretation of nonadjacent dependencies learning, centered around the role of attentional mechanisms, echoes that of Gomez (2002), and finds support in her results. Gomez showed that, in a situation where the successive AXC units were perceptually distinct, the degree to which the A–C relationships were learned depended on the variability of the middle element (X). More precisely, participants were presented with 2, 4, 8, or 24 different X elements, and it was observed that learning increased markedly under conditions of greatest variability in both adults and infants. Gomez argued that the high variability of the intermediary element led participants to focus attention on the nonadjacent elements, because they appeared to be the more stable features in the situation (Creel, Newport, & Aslin, 2004, also report data that can be accounted for along the same theoretical lines).

In natural language, various factors may concur to draw attention to nonadjacent events. The salient locations of these events could be one of these factors, although it is certainly not the only one. For example, prosody and semantics may help. It is also easy to imagine several developmental sketches accounting for how two remote events can be joined in a unitary experience. For instance, a link between A and B may emerge in situations where both events are contiguous. The occurrence of A without its usual successor may then result in the retention of A in working memory until B occurs in order to complete the percept AB. At this moment, A and B would be held simultaneously in the attentional focus despite their objective separation, thus providing conditions favoring both the strengthening of their association and the understanding of the sentence. Ellefson (2002) provides preliminary evidence for this possibility with visual material generated by an artificial grammar. This is again consonant with the SOC framework, which relies on the assumption that perception is shaped by earlier representations.

To summarize, we started from a simple problem, namely segmenting a continuous flow into relevant units, the solution of which requires the detection of associations between contiguous elements. I showed how a model based on the notion of self-organizing consciousness, Parser, achieved the task, at least with an artificial language, through the forgetting-based selection of candidate units. Then I considered increasingly difficult aspects of language acquisition, examining successively the problems raised by the scaling-up issue, by the virtually unlimited number of candidate units to achieve word–object mapping, and by the need to detect nonadjacent dependencies for performing complex syntactic processing. In each case, the solution emerged from the dynamic interplay between the constraints and resources of the conscious/attentional system and the properties of the linguistic and/or extralinguistic environment.

Discussion

The present proposal is at the meeting point of a dynamicist perspective and a mentalistic metatheory. Is such a coordination viable, or, beyond the analogy that may be drawn between the self-organization of consciousness and the self-organization of, say, an insect society, is this use of dynamicist principles purely metaphorical?

There are indeed important differences between the present approach and classical studies committed to the dynamicist framework. One of these is the lack of mathematical formalism. The dynamicist project heavily relies on a branch of mathematics called *dynamic system theory*, in which the interplay between variables is described as a set of differential equations. When applied to a cognitive system, these equations can be solved to predict the behavior of the system, using standard techniques that have been successfully applied to complex mechanical and general mathematical systems. If the dynamicist framework is inextricably tied to this mathematical formalism, our approach lies obviously out of its scope. The intrinsic value of mathematical formalism where it is applicable is not at issue, but I simply believe, following Barton, that it is unwarranted to "use rigorous terminology from nonlinear dynamics to refer to psychological variables that are multidimensional and difficult to quantify" (Barton, 1994, p.12).

There is another apparent paradox in our project. It is one of the hallmarks of dynamic explanations to reject the notion of representation. "To be a truly dynamicist model, there should be *no representation*" (Eliasmith, 1996, p. 452). Now, the notion of conscious representation is central in the mentalistic framework. Actually, I believe the contradiction is only apparent. The notion that is rejected by dynamicists is that of representation as an inferred intermediary variable, aimed at explaining the transfer of information between the components of a cognitive system. There is no need for such a kind of intermediary variable, because there is a direct coupling of the variables through the equations describing the system. However, in a mentalistic view, representations are not inferred constructs. They are the to-be-explained phenomenon. They are facts that are observable by all, because the only representations whose existence is acknowledged are those that compose the momentary phenomenal experience. A mentalistic framework rejects the need for unconscious representations, that is, representations inferred to exist to fulfil a function in a processing model. This entails that, despite the apparent contradiction, the dynamicist and the mentalist frameworks are in fact deeply convergent. Both of them reject the need for inferring mental representations that are not directly observed.

This does not mean that all of the studies committed to dynamicism are consistent with my approach. For instance, the dynamical account of various motor adjustments to changes in the physical environment is quite compatible with the endorsement of a mentalistic metatheory, because this account does not introduce the concept of mental representation and computation within an activity deprived from any conscious components. However, attempts to account for high level activities, such as those outlined in the final chapters of Thelen and Smith's (1994) book, run directly against a mentalistic approach, because they deny the proper existence of a genuine mental level. To conclude, the inclusion of the self-organizing model within the dynamicist framework largely depends on the extension given to the latter. If we consider that a description of variables along differential equations is required, and/or that any mental activity is by principle prohibited, then our approach ought to be excluded. However, it is also possible to consider that our approach is a genuine dynamicist one, because it focuses on temporal aspects, it is centered around self-organizing principles guiding dynamic and time-locked interactions between the organism and its environment, and it obviates the need for inferring an intentional unconscious.

In the title of this paper, I announced a "reversal of perspective" with regard to the statistical approaches to language. Because statistical approaches share many aspects with dynamical models, which were a main source of inspiration for the self-organizing consciousness model, the claim for a reversal seems like an overstatement.

The exact locus of the reversal is the following. As a rule, statistical approaches give no place to consciousness. When the issue is dealt with in the connectionist literature, consciousness is construed as an emergent by-product of statistical computations (O'Brien & Opie, 1999). Thus the status of consciousness is quite similar in connectionist and symbolic approaches of cognition: Conscious representations are the end-product of unconscious computations, with only the nature of the computation changing between the two frameworks. The properties of conscious experiences are irrelevant in the learning process. The present proposal amounts to a complete reversal of perspective, because *instead of being an emergent by-product of statistical computations, the properties of conscious experiences make us sensitive to statistical regularities*. It would be an overstatement, however, to claim that this proposal has the status of a full-blown theory. I have presented here no more than a first, highly speculative sketch for a possible alternative view, with many over-simplifications, in the hope that readers may find it worth exploring further.

Acknowledgments This work was supported by grants from the Centre National de la Recherche Scientifique (CNRS, UMR 5022), and the University of Bourgogne. I am indebted to many colleagues for helpful discussions about these issues, and especially to Barbara Tillmann for her thoughtful comments on a first draft.

References

Aslin, R. N., Woodward, J. Z., LaMendola, N. P., & Bever, T. G. (1996). Models of word segmentation in fluent maternal speech to infants. In J. L. Morgan, & K. Demuth (Eds.), *Signal to syntax* (pp. 117–134). Mahvah, NJ: Erlbaum.

Barton, S. (1994). Chaos, self-organization, and psychology. *American Psychologist*, *49*, 5–14.

Boucher, L., & Dienes, Z. (2003). Two ways of learning associations. *Cognitive Science*, *27*, 807–842.

Brent, M. R. (1996). Advances in the computational study of language acquisition. *Cognition*, *61*, 1–38.

Brent, M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, *61*, 93–125.

Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, *81*, B33–B44.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, *10*, 425–455.

Chomsky, N. (1957). *Syntactic structures*. Mouton, The Hague.

Christiansen, M. H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science*, *23*, 157–205.

Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes*, *13*, 221–268.

Cleeremans, A. (1993). *Mechanisms of implicit learning: A connectionist model of sequence processing* (pp. 227). Cambridge, MA: MIT Press/Bradford.

Cowan, N. (1991). Recurrent speech patterns as cues to the segmentation of multisyllabic sequences. *Acta Psychologica*, *77*, 121–135.

Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: Statistical learning of nonadjacent dependencies in tone sequences.*Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 1119–1130.

Dulany, D. E. (1991). Conscious representation and thought systems. In R. S. Wyer, & T. K. Srull (Eds.), *Advances in social cognition, Vol. 4* (pp. 97–120). Hillsdale, NJ: Erlbaum

Dulany, D. E. (2002). Mentalistic metatheory and strategies. *Behavioral and Brain Sciences*, *25*, 337–338.

Eliasmith, C. (1996). The third contender: A critical examination of the dynamicist theory of cognition. *Journal of Philosophical Psychology*, *9*, 441–463.

Ellefson, M. R. (2002). *The difficulty of learning complex structure: A comparative study of knowledge acquisition*. Unpublished doctoral dissertation, Southern Illinois University, Carbondale.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179–211.

Fitch, W. T., & Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science*, *303*, 377–380.

Gibson, E., & Thomas, J. (1999). Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes*, *14*, 225–248.

Gomez, R. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*, 431–436.

Hunt, R. H., & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: Access to separable statistical cues by individual learners. *Journal of Experimental Psychology: General*, *130*, 658–680.

Jusczyk, P. W., Houston DM, Derek M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, *39*, 159–207.

Kagan, J. (1971). *Change and continuity in infancy*. New York: Wiley.

Karmiloff-Smith, A. (1992). Beyond modularity: A developmental perspective on cognitive science. Cambridge, MA: Bradford/ MIT Press.

Kemp, N., & Bryant, P. (2003) Do beez buzz? Rule-based and frequency-based knowledge in learning to spell plural -s. *Child Development*, *74*, 63–74.

Lewicki, P., Hill, T., & Bizot, E. (1988). Acquisition of procedural knowledge about a pattern of stimuli that cannot be articulated. *Cognitive Psychology*, *20*, 24–37.

McDonald, S., & Ramscar, M. (2001). *Testing the distributional hypothesis: The influence of context on judgements of semantic similarity*. Paper presented at the 23rd Annual Conference of the Cognitive Science Society.

Miller, G., & Chomsky, N. (1963). Finitary models of language users. In R. Luce, R. Bush, & E. Galenter (Eds.), *Handbook of mathematical psychology, Vol. 2* (pp. 419–493). New York: Wiley.

O'Brien, G., & Opie, J. (1999). A connectionist theory of phenomenal experience. *Behavioral and Brain Sciences*, *22*, 127–196.

Pacton, S., Perruchet, P., Fayol, M., & Cleeremans, A. (2001). Implicit learning out of the lab: The case of orthographic regularities. *Journal of Experimental Psychology: General*, *130*, 401–426.

Peereman, R., Dubois-Dunilac, N., Perruchet, P., & Content, A. (2004). Distributional properties of language and sub-syllabic processing units. In P. Bonin (Ed.), *Mental lexicon*. New York: NovaScience.

Peña, M., Bonatti, L .L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, *298*, 604–607.

Perruchet, P., & Peereman, R. (2004). The exploitation of distributional information in syllable processing. *Journal of Neurolinguistics*, *17*, 97–119.

Perruchet, P., & Rey, A. (in press). Does the mastery of center-embedded linguistic structures distinguish humans from non human primates. *Psychonomic Bulletin and Review*.

Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language*, *39*, 246–263.

Perruchet, P., & Vinter, A. (2002). The self-organizing consciousness. *Behavioral and Brain Sciences*, *25*, 297–388.

Perruchet, P., Vinter, A. (2003). Linking learning and consciousness: The self-organizing consciousness (SOC) model. In A. Cleeremans (Ed.), *The unity of consciousness: Binding, integration, and dissociation*. Oxford: Oxford University Press.

Perruchet, P., Gallego, J., & Savy, I. (1990). A critical reappraisal of the evidence for unconscious abstraction of deterministic rules in complex experimental situations. *Cognitive Psychology*, *22*, 493–516

Perruchet, P., Vinter, A., Pacteau, C., & Gallego, J. (2002). The formation of structurally relevant units in artificial grammar learning. *Quarterly Journal of Experimental Psychology*, *55A*, 485–503.

Perruchet, P., Tyler, M. D., Galland, N., & Peereman, R. (2004). Learning nonadjacent dependencies: No need for algebraic-like computations. *Journal of Experimental Psychology: General*, *133*, 573–583.

Pinker, S., & Prince, A.(1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, *28*, 73–193.

Plunkett, K. (1990). The segmentation problem in early language acquisition. *Center for Research in Language Newsletter*, *5*, 1–17.

Ramscar, M. (2002). The role of meaning in inflection: Why the past tense does not require a rule. *Cognitive Psychology*, *45*, 45–94.

Redington, M., Chater, N. (1998). Connectionist and statistical approaches to language acquisition: A distributional perspective. *Language and Cognitive Processes*, *13*, 129–191.

Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science*, *22*, 425–469.

Rumelhart, D., & McClelland, J. (1986). On learning the past tenses of English verbs: Implicit rules or parallel distributed processing? In J. McClelland, D. Rumelhart, & the PDP

Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge, MA.: MIT Press.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*, 606–621.

Searle, J. R. (1992). *The rediscovery of the mind*. Cambridge, MA: MIT Press.

Seidenberg, M. S., & MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science*, *23*, 569–588.

Shanon, B. (2002). Remember the old masters! *Behavioral and Brain Sciences*, *25*, 353–354.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences*, *11*, 1–23.

Sproat, R. (2002). *The linguistic significance of finite state techniques*. Boston, MA: American Association for the Advancement of Science.

Stanback, M. L. (1992). Syllable and rime patterns for teaching reading: Analysis of a frequency-based vocabulary of 17,602 words. *Annals of Dyslexia*, *42*, 196–221.

Thelen, E., & Smith, L. B. (1994). A dynamic systems approach to the development of cognition and action. Cambridge, MA: MIT Press.

Vinter, A. (1986). The role of movement in eliciting early imitations. *Child Development*, *57*, 66–71.

Wright, R. L., & Burton, A. M. (1995). Implicit learning of an invariant: Just say no. *Quarterly Journal Experimental Psychology*, *48A*, 783–796.