THE UNITY OF CONSCIOUSNESS

BINDING, INTEGRATION, AND DISSOCIATION

Edited by

AXEL CLEEREMANS

(2003)

OXFORD
UNIVERSITY PRESS

CHAPTER 3.1

LINKING LEARNING AND CONSCIOUSNESS: THE SELF-ORGANIZING CONSCIOUSNESS (SOC) MODEL

PIERRE PERRUCHET AND ANNIE VINTER

Introduction

There are two main ways of thinking about learning in contemporary cognitive sciences. In the historically earlier mode of thought, people acquire knowledge about situations through the application of specialized algorithms whose functioning follows certain logical or rational principles. In the second conception, people become sensitive to the regularities embedded in the material thanks to non-specific, and predominantly associative, mechanisms. These two conceptions are typically implemented in classic (formal, symbolic) models and in neural network modeling, respectively. Our starting point is that both conceptions of learning share a similar view with regard to the relation between learning and consciousness. Indeed, in the two conceptions, most if not all of the information manipulation and computation involved during a training episode takes place at an unconscious level. Learner's phenomenal experience has no relation with the hypothesized mechanisms of learning, whether based on symbolic or connectionist architectures.

Reciprocally, the literature on consciousness usually makes no reference to learning. Consciousness is usually construed as making accessible the end-product of the unconscious computations involved during learning, with conscious thought taking no active part in the genuine learning process. The thesis developed in this chapter is that linking the notions of learning and consciousness provides us with an approach that may bring about a deep-seated revision of our understanding of the two concepts.

In the first section, we illustrate the above claims about the current divorce between learning and consciousness through a specific example. Then, using the same example, we introduce our alternative conception, in which learning is a natural by-product of on-line phenomenal experience. Subsequent sections are devoted to generalizing our proposal. Our integrative conception is more parsimonious, we argue, because it avoids

the need for two independent machineries: one, fully unconscious, devoted to learning, and the other generating phenomenal experience and devoid of any function in the learning process. Of special interest are the implications of our view with regard to the unity of consciousness. To anticipate our conclusions, instead of being a to-be-explained phenomenon without evident utility, the unity of consciousness appears to be a fundamental concept for development and learning.

Two prevalent approaches to learning

The situation selected to illustrate our proposal concerns the discovery of words from a non-segmented speech flow. This ability is worthy of interest in itself, in so far as the formation of words can be construed as a necessary prerequisite for learning about other aspects of language, such as its syntactic structure. Moreover, this situation may be taken as representative of a very large class of non-linguistic problems, such as perceiving objects in the visual environment. In a formal way, this situation taps our ability to build internal representations that are isomorphic to the world structure, when this structure is not salient in the sensory input.

Language acquisition initially proceeds from auditory input, and linguistic utterances usually consist of sentences linking several words without clear physical boundaries. The question thus arises: how do infants become able to segment a continuous speech stream into words? Recent psycholinguistic research has identified a number of prosodic and phonological cues that may potentially help infants, but they provide only probabilistic information. The importance of prosodic and phonological cues in word discovery is further questioned by recent experimental studies showing that these cues are not necessary. For instance, Saffran *et al.* (1996) used an artificial language consisting of six trisyllabic words, such as *tutibu* and *dutaba*. The words were read by a speech synthesizer in random order in immediate succession, without pauses or any other prosodic cues. Thus the participants heard a continuous series of syllables without any word boundary cues. In the following phase, they were asked to perform a forced choice test in which they had to indicate which of two items sounded more like a word from the artificial language. One of the items was a word from the artificial language, whereas the other was a new combination of three syllables belonging to the language. Participants performed significantly better than would be expected by chance. This and other studies (e.g., Saffran *et al.* 1997) offer impressive support for the hypothesis that people are able to learn words forming a continuous speech stream without any prosodic or phonological cues for word boundaries.

The symbolic model of word segmentation developed by Brent and Cartwright (1996) offers a first way to account for this ability. The authors construe segmentation as an optimization problem. The principle of the method is akin to establishing a list of all the possible segmentations of a given utterance (although the authors used computational tools which 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 (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. This method has been applied with some success for parsing phonetic transcripts of child-directed speech into words.

The performance observed in the Saffran *et al.* experiments can also be accounted for by connectionist models. Most of the connectionist models which address the word segmentation issue rely on the simple recurrent network, or SRN, initially proposed by Elman (e.g., 1990; see also Cleeremans 1993). An SRN is a network which is 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 in the network at each time step, in such a way as to decrease the discrepancy between the two events. Elman (1990) presented such a network with a continuous stream of phonemes, one phoneme at a time, the task being to predict the next phoneme in the sequence. The accuracy of prediction was assessed through the root mean square error for predicting individual phonemes. After training, the error curve had a marked saw-tooth shape. As a rule, the beginning of any word coincided with the tip of the teeth. This means that after a word, the network was unable to predict the next phoneme. However, as the identity of more and more of the phonemes in a word was revealed, the accuracy of prediction increased up to the last phoneme of the word, and the error curve therefore fell progressively. The start of the next tooth indexed the beginning of the next word. Therefore, an SRN provides information useful for the parsing of a continuous speech flow into words, although words can only be revealed using a subsequent cluster analysis (for more recent models, see Aslin *et al.* 1996; Christiansen *et al.* 1998).

Prevalent approaches and consciousness

Let us consider now the conscious experience of the learner. Certainly this experience may differ from one subject to another. However, the following description presumably captures the gist of everyone's experience. During the initial exposure to a new language, the same phenomenon presumably occurs as with any unstructured sequence of items: the material is chunked into parts composed of a few perceptual primitives. These chunks form the discrete content of successive attentional focuses, and are typically short lived. As a confirmation of this assumption, if the speech flow is stopped at a given moment,

a subject may be able to recall the most recent set of two or three syllables, but no more. Then, with further practice, some chunks may become increasingly familiar, so that they are automatically noticed in the speech flow. Stopping the training session may now allow the recall of a few chunks. The final step is not very different, except that all the material is now perceived as a succession of familiar chunks that match the words of the language.

Needless to say, as illustrated in Fig. 3.1.1, there is no possible way to link the above models and the conscious experience of the learner throughout the learning phase. The operations involved in the Brent and Cartwright (1996) model, such as the computation of all the possible segmentations of an utterance in order to choose the one responding to prespecified criteria, have no subjective counterpart. Likewise, the SRN model's contents do not form part of a learner's experience. Relying on the distribution of errors to draw a parallel with the subjective experience of the learner throughout training leads to a dead-end. For instance, errors are almost equally distributed at the outset, thus preventing any inference about subjective units, whereas the beginning learner perceives chunks that, although randomly determined, have the format of discrete representations. Even the final state, namely the representation of the input as a set of words, is not directly provided by the network: a significant amount of computational work is still needed to infer words from the graded distribution of errors after learning is completed.

These remarks can hardly be thought of as criticisms of these models, given that they were not devised to account for conscious experience. However, they highlight a striking paradox: the models of word segmentation considered up to now rely on mental states and representations we have no evidence of, while offering no explanation for the

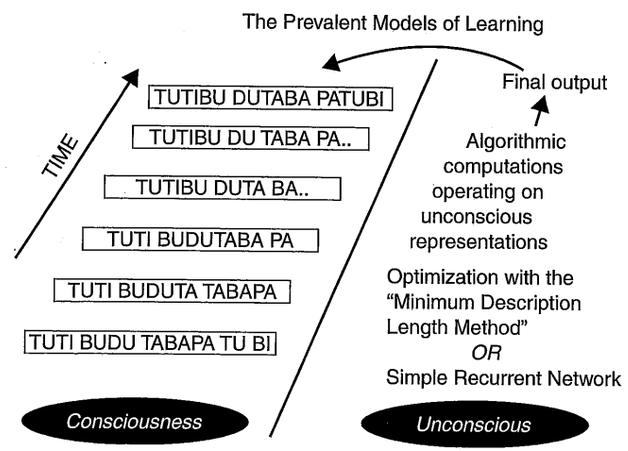

Figure 3.1.1 In the prevalent models of learning, whether symbolic or connectionist, learning proceeds from unconscious algorithms operating on representations that are unrelated to the learner's subjective experience. Consciousness provides optional access to the result of these computations.

representations of which we have direct and immediate evidence through conscious experience. The end result is that, in these models, assumptions are made about operations involving unconscious representations while the representational content of phenomenal experience is left both unexplained and without object. This statement can be generalized to virtually all the models composing the cognitive literature, whatever their domain of relevance.

Another view

We now intend to show that seriously considering the subjective experience of the learner makes the unconscious algorithmic computations involved in symbolic or connectionist models of learning useless. Indeed, the content of subjective experience is endowed with a remarkable property, namely its self-organizing nature.

Let us return to the word segmentation experiment presented above. Imagine that the initial chunks of a few syllables composing the conscious experience of the learner, instead of being an epiphenomenon doomed to fulfill the subjective scene while the serious learning mechanisms are operating in the deep unconscious, are in fact the prefiguration of what will become, a few minutes later, the words of the language. Of course, these initial chunks do not match the words, except in a few fortunate cases. More probably, they are part words, or straddle over word boundaries. How can these randomly drawn, structurally irrelevant chunks, become the relevant words?

A response is provided by the application of the three following principles, which have been confirmed by a huge amount of experimental studies. First, the components of any conscious percept tend to be retrieved as a whole, and perceived again as a whole when these components, or a part of them, are displayed again in the nominal stimulus. This means that if "tuti" was the object of an attentional focus for a given subject, "tuti" tends to form a new internal representation and, as such, may be retrieved and perceived more fluently if "tuti" is displayed again, completed if "tu" alone is displayed, and so on. Second, the phenomenon of unitization described above is short lived. If there is no subsequent opportunity to perceive the components of a given percept, the properties described above tend to disappear, as a consequence of both natural decay and interference from the processing of similar material. However, unitization is strengthened by repetition. The third principle is that unitization allows the new representation to become a component within a further conscious percept, and hence, due to the recurrence of the first principle, allows the formation of another, more complex representation. This means that if "tuti" has become a new perceptual unit due to its repeated perception, "tutibudu" may be a further conscious percept. If "tutibudu" reoccurs frequently enough in the corpus, it will evolve as a new stable representation.

In theory, these principles should be able to account for word extraction, due to a property inherent to any language. If the speech signal is segmented into small parts on a random basis, these parts have more chance of being repeated if they match a word,

or part of a word, than if they straddle word boundaries. In consequence, the primitives that emerge from natural selection due to the forgetting of infrequent percepts are more likely to match a word, or part of a word, than a between-word segment.

In order to provide evidence that these principles are sufficient to account for the transformation of the initial, irrelevant chunks, into the words of the artificial language, they have been implemented in a computer program, PARSER. The model has been described in Perruchet and Vinter (1998) and an on-line presentation is available at the URL (http://www.u-bourgogne.fr/LEAD/people/perruchet/SOC_fichiers/frame.htm). Suffice it to say that the model, except for unavoidable parameter specifications, includes no mechanisms other than those deriving directly from the above principles. Simulations have revealed that PARSER extracts the words of Saffran and collaborators' artificial language without any errors, well before exhausting the material presented to actual participants (Perruchet & Vinter 1998, 2002).

It is worth emphasizing that, in striking contrast with other models, the only representations included in PARSER closely match the conscious representations participants may have when performing the task. The early coding of the material as a set of short and disjunctive units, as well as the final coding of the input as a sequence of words, are assumed to match the phenomenal perceptual experience of the listeners. This correspondence also extends to the entire training phase, thus permitting our model to perform word segmentation while mimicking the on-line conscious processing of incoming information. Importantly, these conscious percepts enter into a causal chain, which takes subjects' representations from the initial chunks to the final words (Fig. 3.1.2).

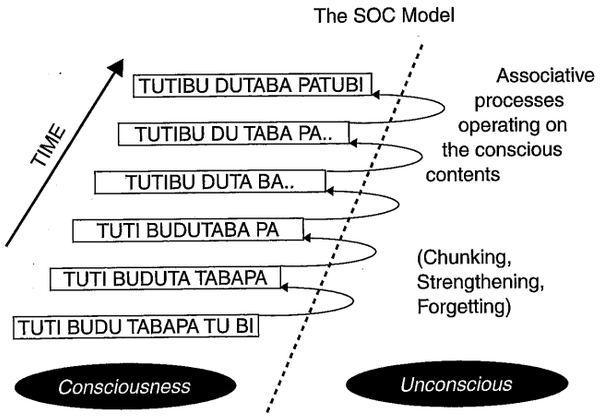

Figure 3.1.2 In the framework proposed here, unconscious associative processes operate on the representational contents of subjects' on-line conscious experiences. Although illustrated here for the word extraction issue, the conceptions represented in Figs 3.1.1 and 3.1.2 are quite general.

Concept of self-organizing consciousness

The above demonstration directly introduces the concept of self-organizing consciousness (SOC): the content of consciousness self-organizes, that is to say it becomes increasingly isomorphic to the world as a consequence of the interaction between its own properties and the properties of the world.

The SOC framework is primarily a theory of learning. This theory of learning shares many aspects with the conventional associative theory of learning and memory. The basic phenomenon that allows extraction of the structure of the material is that the unitization resulting from the simultaneous perception of individuated components is strengthened through repetition and vanishes through forgetting and interference. The only point of departure with regard to the conventional theory is our focus on *conscious* percept. Indeed, we assume that conscious representations are the very stuff on which associative processes operate. However, even if this principle is not commonly stated as such, it is in keeping with a widespread conception according to which associative learning and memory are nothing other than the by-products of *attentional* processing. For instance, many authors, using a different terminology, agree in claiming that associative learning is an automatic process that associates all the components that are present in the attentional focus at a given point (Wagner 1981; Frensch & Miner 1994; Logan & Etherton 1994; Stadler 1995; Jimenez & Mendez 1999). If one subscribes to the view that what constitutes the content of the phenomenal experience at a given moment is what is attended to at this moment, and vice versa (e.g. Miller 1962; Posner & Boies 1971; Mandler 1975; Cowan 1995), it then appears that our framework is akin to the contemporary conceptions of associative learning and memory. The novelty, in fact, is not in the principles involved, but in the demonstration of their power in extracting the world structure.

However, the SOC framework goes far beyond the field of learning as it is usually defined. More exactly, by giving to any conscious experience a role in the formation of subsequent conscious experiences, it opens up a new conception of the relations between the conscious and the unconscious.

In the mainstream of cognitive approaches, the function assigned to consciousness consists in making certain parts of cognitive functioning accessible. To quote a recent overview by Baars (1998), "Many proposals about brain organization and consciousness reflect a single underlying theme that can be labeled the 'theater metaphor'. In these views the overall function of consciousness is to provide very widespread access to unconscious brain regions." And elsewhere in the same paper: "A classical metaphor for consciousness has been a 'bright spot' cast by a spotlight on the stage of a dark theater . . . . Nearly all current hypotheses about consciousness and selective attention can be viewed as variants of this fundamental idea." Being conscious is seeing with the mind's eye. It is worth stressing that the very notion of access implies that consciousness is stripped of any major role. The states and operations that are accessed would still exist even if access were not possible. The unconscious mental life is fully autonomous.

This view, we suggest, implies a complete reversal of perspective. In keeping with a conception cogently advocated by Dulany (e.g. 1991, 1997), in the SOC framework, the conscious and the unconscious are so intimately related that no mental life is possible without both of them. However, they are quite different in nature. The first refers to the mental representations, and the second to the processes, predominantly associative, acting on these representations. The two are linked like the head and tail of a coin. To quote ourselves in an earlier paper: "Processes and mechanisms responsible for the elaboration of knowledge are intrinsically unconscious, and the resulting mental representations and knowledge are intrinsically conscious. No other components are needed" (Perruchet *et al.* 1997, p. 44; see also O'Brien & Opie 1999, for a link between the notions of representation and consciousness). In this theoretical context, the unity of consciousness is no longer an amazing phenomenon that calls for a separate explanation, as a subsidiary complement to the main tasks of the cognitive sciences. The property of the conscious contents to cohere the elements that are simultaneously processed is, when considered within a dynamic perspective of learning, the fundamental principle on which any explanation of mental life must be based.

Is the SOC framework advantageous with regard to the conventional approaches of cognition?

In various talks in which we have presented the SOC framework and its application to the word segmentation issue, an objection that has frequently been raised is: "In fact, your model accounts for the same data as the prevalent cognitive view. Now, your model needs consciousness, whereas consciousness fulfills an optional function in the prevalent view. For instance, PARSER relies on conscious thought to learn to segment a continuous speech flow, whereas the same performance can be obtained using a simple connectionist network. The conclusion follows that your model is of no interest, because it makes unnecessary and costful postulates."

We agree with the premises, and we acknowledge that, at first glance, the pessimistic final inference appears warranted. Indeed, introducing consciousness into psychological modeling seemingly increases complexity, because this concept involves a huge number of definitional and theoretical problems. Of primary concern is the fact that the notion of consciousness is faced with what is variously designated as the "brain/mind issue" or "the explanatory gap", which refers to how consciousness can emerge from physical matter. This kind of issue is indeed extremely complex. However, it must be realized that the prevalent view is not immune to the problems inherent to the brain/mind issue. Indeed, the main enigma is not the subjective feeling associated with a conscious content, but the formation of mental states or representations, whether conscious or unconscious, from biological events. Our ignorance about the transformation of neural patterns into mental events does not prevent cognitive psychologists from routinely introducing (unconscious) mental states and representations in their models. We see no reason to condemn this approach simply because consciousness is evoked.

The implications of introducing consciousness into psychological modeling, we suggest, must be assessed at the computational level. It is often claimed that consciousness is computationally irrelevant, because qualifying a representation involved in a computational model as conscious or unconscious has no effect on the way this model works. As a case in point, qualifying the chunks formed in PARSER as conscious or unconscious is consequentless. However, the computational irrelevance of consciousness can no longer be maintained if, instead of considering piecemeal aspects of the models, we consider their overall conditions of functioning. It appears then that introducing consciousness entails the consideration of striking functional constraints, such as limited capacity, seriality, relative slowness of processing, and quick memory decay. The chunks formed in PARSER certainly do not need to be conscious: the model would work just as well if they were ascribed an unconscious status. However, they can be qualified as conscious, because nothing in the model violates the constraints of conscious thought. At any given moment, the system deals with a quite limited number of primitives, which are processed serially. No rapid computation is involved. And the perceptual chunks are subject to quick memory decay, which mimics the well-documented decay of the traces described in the literature on memory (note that 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 the very property that offers the system 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, PARSER would fail to extract the relevant units from the multiple candidate units processed by the system).

It is usual to frame the issue of consciousness in terms of *necessity*. The underlying idea is that it is more economical to ascribe, a priori, an unconscious status to any representations or mental operations, and to introduce consciousness only if this addition seems unavoidable on the basis of empirical evidence. For instance, the entire literature concerning the "conditioning without consciousness" issue, in vogue in the 1970s, raised the question: "is consciousness necessary for discovering the interstimulus relationships?" The literature on implicit learning raised similar questions two decades later.

In contrast, our claim is that questions about consciousness must be framed in terms of *sufficiency*, rather than necessity. The relevant question is not: "is consciousness necessary for a given adaptive response?" In fact, addressing this question appears experimentally intractable, because a positive response would require one to demonstrate that unconscious representations and computations do not exist, and demonstrating that any construct does not exist is beyond the reach of any empirical investigation. The question that needs to be addressed is: "is it sufficient to rely on the transient and labile representations that form one's momentary phenomenal experiences to account for a given adaptive response?"

About generalization

Up to now, we have shown the self-sufficiency of conscious representation only through PARSER, a computational model devised to find the words forming a simplistic micro-language composed of six trisyllabic words. In the remainder of this chapter, we will deal with the generalization issue in two ways, one following a bottom-up and the other a top-down approach. In the first approach, we will start from PARSER, in order to show how the general principles underlying the model are able to account for the formation of complex representations. In the second, we will start from a sample of complex adaptive phenomena, and we will show how these phenomena are in fact reducible to the formation of representations isomorphic to the world, whereas they have been primarily thought of as indicative of unconscious computations. Our hope is that the two approaches meet each other and thus trace the outline of a conceptual bridge between PARSER and the most complex aspects of human adaptation (a more detailed demonstration may be found in Perruchet and Vinter (in press)).

Generalization: the bottom-up approach

PARSER demonstrates limited learning abilities. Among other characteristics; (i) it forms units from immediately adjacent primitives, although there is evidence that representations may embed remote elements, irrespective of whether those elements are distant across time or space; (ii) it appears mostly sensitive to the frequency of co-occurrences, although there is well-documented evidence that people and animals are sensitive to co-variations or contingencies; and (iii) it learns only literal, surface aspects of the material, although there is evidence that some transfer of learning across surface features is possible. Our claim is that all these limits are linked to PARSER itself, and not to the SOC framework of which PARSER represents a specific implementation devised to address a specific problem.

In the SOC framework, any conscious representation is the end-product of associative processes acting on prior conscious representations. In PARSER, the content of the former conscious representations is almost uniquely determined by the nature of the material. The unique property of attention that is called on is the fact that only a few primitives can enter into a given processing unit. This is due to the properties of the material—a monotonous sequence of nonsense syllables—in which no other structuring cues are available. However, in most real settings, the current conscious content is not determined by its size only, because attention is selectively captured by certain features of the material. We intend to show now that the other properties of attention also serve the self-organizing process.

First, it is fairly obvious that the current conscious content may include both aspects perceived in the current external environment and the memorized or imagined representations that are either automatically triggered by the external stimuli or intentionally evoked. These components of the attentional focus, if they are jointly processed with

Table 3.1.1 In this table, a is the number of co-occurrences of X and Y, while the contingency between X and Y is given by the ratio $(a/(a+b) - c/(c+d))$

	X	$\overline{\text{X}}$
Y	a	b
$\overline{\text{Y}}$	c	d

sufficient frequency, can generate new stable units. This ability does not violate the contiguity principle that underpins the literature on associative learning. The point is that there must be a contiguity relation, not between the to-be-associated components in the external data, but between their representations in mind. This reasoning makes it possible to account for the fact that we are able to form representations of the world including primitives that are not adjacent in the sensory input.

The second point concerns the fact that a huge amount of literature has demonstrated the sensitivity of organisms to contingency rather than to pure co-occurrences. The difference is illustrated in Table 3.1.1: the number of co-occurrences is defined by a, which indicates the number of times that X and Y are jointly processed, whereas the contingency is normatively defined as the difference between the probability of X when Y is present and the probability of X when Y is absent $(a/(a+b) - c/(c+d))$. At first glance, the units found by PARSER are defined by the co-occurrences of their components. However, this limitation is more apparent than real, mainly due to the functional properties of attention. A well-documented finding is that attention is captured by novelty. This property accounts for the detection of contingency. Indeed, assuming a fixed number of XY co-occurrences, the amount of attention devoted to these co-occurrences depends on the number of occurrences of isolated X and Y events (b and c, respectively). When b and c decrease (and hence when the level of contingency increases), the amount of attention devoted to X and Y (and hence to the co-occurrences XY) increases, thereby making the formation of an internal unit XY more probable.

Attention may also be captured by some abstract properties of the world. For instance, it is well documented that attention is captured by symmetry or repetition structure. This means that the repetition of two events, whatever they are, is automatically detected. This property is very important, because it provides a clue to explaining the formation of representations that are abstracted away from their sensory content, the third aspect on which PARSER seemed severely limited. Such an explanation has been shown to account for most of the findings demonstrating transfer across sensory features in learning (Perruchet & Vinter, in press).

To illustrate how these principles work, let us consider an example inspired by a question raised by Karmiloff-Smith (1992, p. 40) concerning word/object mapping. When an adult points a cat and says "look, a cat", how can the child pair the word "cat"

with the whole animal, rather than, say, with the cat's whiskers, the color of the cat's fur, or the background context? The SOC framework addresses this question. What is likely to become associated is what captures the infant's attention. As noted above, one of these properties 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. The same reasoning may hold with movement. It has been established that infants' attention is attracted by a moving display (e.g., Haith 1978). If several elementary features move concurrently, they have a high probability of being attentionally processed as a whole by infants. Returning to Karmiloff-Smith's question and considering the auditory input first, "cat" is presumably newer than "look", because "look" has been associated with many contexts before. As a consequence, it is highly probable that "cat", rather than "look", enters into the momentary attentional focus. On the other hand, it is also 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.

All these examples provide illustrations of the power of the concept of self-organizing consciousness. In each case, the general and specific properties of the attentional system appear to be tuned in such a way that the resulting conscious content becomes increasingly isomorphic to the world structure.

Generalization: the top-down approach

In the prior section, we suggested how the principles underpinning PARSER could account for the formation of conscious representations far more complex than the trisyllabic words of the Saffran *et al.* experiments. The value of building complex representations for adaptive behaviors that are commonly referred to as the end-products of sophisticated computations remains to be seen. We intend to show that complex and integrative conscious representations deprive unconscious computation of any object. Here, our thesis relies heavily on the idea that neural systems "trade representation against computation", to borrow the expression used by Clark and Thornton (1997). PARSER provides a first insight into the meaning of this claim. As shown by the comparison of Figs 3.1.1 and 3.1.2, the change in the conscious percept due to simple associative processes may replace, at a functional level, the operations hypothesized in conventional models, whether they are based on a formal or a connectionist architecture.

Although often indirect, supporting evidence for a representation/computation tradeoff can be found in various areas of psychology. Examples include the lexicalist approaches to syntactic processing (e.g., Bates & Goodman 1999), the notion of mental models in problem solving (e.g., Johnson-Laird 1983), and the memory-based theory of automatism (Logan 1988). These examples will be briefly outlined in turn. Other examples that space limitations prevent us from presenting here are the instance-based model of categorization (e.g., Brooks 1978) and the so-called episodic (e.g., Neal & Hesketh 1997) or fragmentary (e.g., Perruchet 1994) accounts of implicit learning.

Although they evolved in at least partial independence, all these avenues of research share the same general distrust with regard to the notions of abstract computation and rule-based processing, and stress the adaptive advantage of building complex representations.

Syntactic processing

Of special relevance for our position are the so-called lexicalist approaches to syntactic processing. As claimed by Bates and Goodman (1999, p. 37), "A general trend has characterized recent proposals in otherwise very diverse theoretical frameworks within linguistics: More and more of the explanatory work that was previously handled by the grammar has been moved into the lexicon." Converging lines of evidence have evolved in other contexts. For instance, careful scrutiny of the linguistic productions of young children shows that these productions are organized around particular words and phrases, instead of operating with abstract linguistic categories and schemas. This finding of the item-based learning and use of language appears fairly general (for a review, see Tomasello 2000). Of course, the lexicalist, "item-based", or "memory-based" (McKoon & Ratcliff 1998) approaches to grammar have not gone unchallenged. Some authors go on to argue that there is a modular dissociation between syntax and lexicon (e.g., Grodzinsky 2000). We are not familiar enough with the domain to offer new arguments in either direction. Our intention was simply to point out that distinguished figures in the psycholinguistic literature have been prepared to reject the idea that the mastery of at least some aspects of syntactic regularities necessarily implies syntactical rules. Such a view confers a high degree of plausibility on one of the main propositions of this chapter, namely that it may be possible to explain the apparent use of abstract rules in terms of the formation of complex representations.

Problem solving

In many cases, the solution to a problem springs to mind without the phenomenal experience of engaging in logical–analytic operations. Conclusions simply rise to consciousness, without being the outcome of a worked-out inference. In keeping with the dominant zeitgeist, this phenomenon is commonly attributed to the action of an unconscious and sophisticated processor. The underlying idea is that the solution to a problem may be worked out in the absence of conscious awareness of the operations required by this problem. Our suggestion is that intuition and insight, and all the cases in which logic-like operations are apparently performed by the mind in the absence of conscious thought, can be encompassed within the notion of self-organizing consciousness. We have seen above how the notion of self-organizing consciousness allows us to account for the formation of internal representations which are increasingly congruent with the world structure. If we expand the scope of these representations to the various dimensions involved in a given problem, it becomes conceivable that a representation contains, in some sense, both the data and the solution of the problem. The solution pops up in

the mind, because it is a part of the model of the world that people have built through automatic associative processes.

Let us take a simple example, one relating to the notion of transitivity. In the linear ordering tasks, two premises are presented, the formal expression of them being: "A is longer than B and B is longer than C". Participants have to judge whether an expression such as: "A is longer than C", is correct. It can be assumed that people solve this task because they have some formal notion about the transitivity of the expression "longer than", and that they apply the transitivity rule to the problem at hand. However, it is far simpler to assume that people have built an integrative representation of the premises in the form of a linear array, and then read the response to the question directly on this representation. There is now a consensus about the idea that people proceed in this way (Evans *et al.* 1993). This illustrates how a representation which is isomorphic to the world structure makes rule knowledge unnecessary.

This claim is reminiscent of various proposals, from the notion of mental models advanced by Johnson-Laird (1983), to the representation/computation trade-off envisaged by Clark and Thornton (1997). Shastri and Ajjanagadde's (1993) simulation model of reasoning relies on the same general view. These authors show how a neural network may simulate reasoning through the formation of a model of the world. To borrow their terms: "The network encoding of the Long term Knowledge Base is best viewed as a vivid internal model of the agent's environment, where the interconnections between (internal) representations directly encode the dependencies between the associated (external) entities. When the nodes in this model are activated to reflect a given state of affairs in the environment, the model spontaneously simulates the behavior of the external world and in doing so makes predictions and draws inferences." In Shastri and Ajjanagadde's framework, the internal model of the world takes the form of a neural network, and the authors do not provide a detailed account of the question of learning. Moreover, they say nothing about the issue of consciousness. However, it is easy to see how the same view can be held about the conscious representations which are built thanks to their self-organizing properties: representations become able to provide a model of the world in which some structural relations that have not been provided as such can be directly "read", instead of being computed through analytical inference processes.

Incubation

A marginal concern in the literature on problem solving relates to the phenomenon of incubation. Everyone has had the experience of a solution to a problem suddenly occurring after we have given up our deliberative and unsuccessful search for it. The phenomenon may occur for relatively simple problems of daily life, as well as in more sophisticated situations. For example, Henri Poincaré provided a fine-grained description of this effect based on his own experience of the resolution of very complex mathematical problems. The phenomenon was termed incubation by Wallas (1926). According to Wallas, when the solution to a problem is not directly reached through

explicit, step-by-step reasoning, it may be useful to suspend the search for a solution, in order to allow "the free working of the unconscious or partially conscious processes of the mind". Such phenomena provide, at first glance, clear-cut evidence for the fact that after suspension of deliberative search, a sophisticated cognitive unconscious takes over and goes on searching in parallel to the overt activities.

However, as claimed by Mandler (1994) in an overview of the phenomenon, "there is no direct evidence that complex unconscious 'work' (new elaborations and creations of mental contents) contributes to the incubation effects" (Mandler 1994, p. 20). This is because incubation can be accounted for in much more simple terms. Instead of imagining that the filling task leaves the cognitive unconscious free to search for a solution, it may be assumed that the intervening task makes it possible to forget certain aspects which are irrelevant to the solution of the problem at hand. The forgetting of inappropriate elements should promote the emergence of a new perceptual structuring. Smith and Blankenship (1989, 1991) have provided experimental evidence for this hypothesis: when misleading information was given to subjects while they were trying to solve various problems, an incubation delay led both to an improvement in problem solving and reduced memorization of the misleading information, with a close relation between the two effects. Here again we find the idea developed in PARSER that forgetting is crucial for the formation of perceptual representations isomorphic to the structure of the material.

Thus it appears that the formation of conscious representations thanks to elementary mechanisms of associative learning is able to account for many cases where the discovery of a solution has been attributed to some unconscious analytical reasoning. This conclusion could be expanded to other forms of learning that space limitation prevents us from examining in detail. For instance, studies on concept learning have yielded similar findings. In a study involving complex, ill-defined concepts, Carlson and Dulany (1985) concluded that "hypotheses of unconscious learning are most strongly disconfirmed by evidence that the content of conscious awareness could, given reasonable process assumption, account for the learning observed" (Carlson & Dulany 1985, p. 45).

Decision making

Going a step further in our speculation, decision making might prove to be another area of application of our framework. Most often, when faced with a choice, we have an immediate preference for one alternative, and explicit thoughts, when they occur, are merely able to suggest a posteriori justifications. It might again seem that spontaneous decisions are the product of an unconscious analysis of all the factors relevant for this decision. Our model suggests a more parsimonious explanation, provided that we make some additional assumptions. Phenomenal experience does not only comprise the cold representations of the world: it is emotionally valenced, either positively or negatively. Our proposal is that decision could be directly based on this affective valence, and that the affective valence is itself the end result of associative processes such as those involved in PARSER. In other words, we suggest that a situation is directly perceived as

positively or negatively valenced, this feature being a consequence of the self-organizing property of consciousness. Indeed, there is no reason to think that emotive components escape from the associative processes that shape conscious experience. On the contrary, we have experimental evidence, through the studies on conditioning, and especially the recent studies on evaluative conditioning (e.g. De Houver *et al.* 1997), that the emotive components are responsive to the same mechanisms as those involved in PARSER. Thus the conscious representations that have developed under natural conditions are probably endowed with an affective color which results from associative processes and which may be directly responsible for the decision.

Automaticity

At first sight, the fact that cognitive operations and representations can progressively relax their initial link with conscious awareness as a consequence of extended training seems irreducible to the SOC framework. Consciousness appears to be an optional quality of cognitive activities, a proposal that is in contradiction with our framework. The point we wish to make here is that although our framework is indeed incompatible with the possibility of transferring operations from a conscious to an unconscious mode, the idea that automatization consists in such a transfer is only one of several theoretical accounts of the phenomenon. This account is instantiated by the LaBerge and Samuel (1974) theory, in which automatization is equated with the progressive withdrawal of attention from operations that are otherwise left qualitatively unchanged. In a similar vein, Shiffrin and Schneider (1977) argue for a transition from serial to parallel processing. The notion of knowledge compilation, introduced by Anderson (e.g., 1983) also relies on the same principles: programs running in interpreted and compiled modes use different codes, but they follow the same algorithms. These theories converge to strengthen the view that the cognitive unconscious can perform the very same processing as conscious thought but with even greater proficiency.

These interpretations of automatisms have been challenged by Logan and his co-workers (e.g., Logan 1988). For Logan, the withdrawal of attention that characterizes automatization is not a cause, but a consequence of a change in the nature of the operations performed by the learner. The change is described as a transition from performance based on a general algorithm to performance based on memory retrieval. Logan illustrates this idea in the field of arithmetic computation: initially, children perform, say, additions, with a general counting algorithm but, after practice, they retrieve sums directly from memory without counting. The point is that step-by-step counting operations do not transfer from a conscious to an unconscious mode of control: they are simply deleted, and replaced by another operation.

Logan's theory provides the elements that allow the SOC model to encompass the data relating to automaticity. In the SOC model, automaticity may be construed as the possibility for a subject forming a new conscious representation whose components were previously perceived as independent primitives. The phenomenon is in fact very similar to that envisioned for the learning of the words of a language. When people create a new unit such as *tutibu*, this unit is also composed of initially independent primitives such as *tuti* and *bu*. The difference lies in the fact that, for instance, the final unit *tutibu* is given in the data, and needs only to be captured through selection from other possible units. By contrast, the final unit "5 + 3 = 8" may need to be built through time-consuming operations on the part of the subjects. But this difference does not mean that the final outcome differs: after training, people evoke the conscious unit "5 + 3 = 8" in the very same way that they evoke the conscious unit "*tutibu*". As Logan contends, automatic behavior is nothing other than memory retrieval.

Automaticity and the absence of consciousness are frequently referred to as identical (e.g., Jacoby *et al.* 1993). Our conclusion is, ironically, the exact opposite. The phenomenal experience is, to a large extent, the product of an automatization process. Tzelgov recently gave one chapter of a book the title: "Automatic but conscious: that is how we act most of the time" (Tzelgov 1997, p. 217). We fully agree with this claim. Obviously, we are conscious of the *output* of the mechanisms involved, and not of the mechanisms themselves. But the automatisms have no specificity in this regard: this is the case for all biological processes. Automatic behaviors are unconscious in the same sense that, say, the explicit remembering of past may be said to be unconscious: in both cases, we have no access to the mechanisms generating the current phenomenal experience. What gives us the feeling that some sophisticated computation on symbolic representations occurs unconsciously in automatized performance is linked to the belief that performance after extensive practice involves the very same set of operations that was requested at the beginning of practice. Once this assumption is abandoned, automatized activities can be qualified in the same way as any other activities: they are the conscious outcomes of unconscious mechanisms.

Conclusion

In contrast to the prevalent view, according to which learning processes and learners' conscious experience are unrelated phenomena, we propose that learning is due to the self-organizing nature of conscious thought: conscious contents become increasingly isomorphic to the represented world, as a result of the interaction between the properties inherent to the conscious/attentional system and the properties of the world. This view provides a powerful model of learning, because, on the one hand, the concept of self-organizing consciousness can account for the formation of surprisingly complex representations and, on the other, the formation of these representations can account for many adaptive phenomena commonly thought of as involving sophisticated computations following unconscious algorithms.

Conversely, the concept of self-organizing consciousness premises a thorough-going reappraisal of our understanding of phenomenal consciousness. Of primary concern is the unitary feeling associated with any subjective experience. To date, this major aspect of phenomenal experience has not been ascribed a clear function. Our approach highlights a function of having unitary conscious contents that has so far been ignored. This

function emerges when one considers the phenomenon in a dynamic perspective, instead of considering only the instant expression of conscious experiences. The fact of binding together the elements composing the ongoing percept allows a progressive structuring of subsequent percepts and representations. In consequence, the formation of unitary conscious contents appears to be a fundamental process deserving consideration when looking for the basic mechanisms underpinning learning and child development.

Obviously, this theoretical sketch is still overly speculative, and needs a huge amount of further empirical work and theoretical refinement. We hope we have convinced readers that this task is worth pursuing, because the resulting conception should be much more parsimonious than the current views.

References

Anderson, J.R. (1983) *The Architecture of Cognition.* Cambridge, MA: Harvard University Press.

Aslin, R.N., Woodward, J.Z., LaMendola, N.P. & Bever, T.G. (1996) Models of word segmentation in fluent maternal speech to infants. In: Morgan, J.L. & Demuth, K. (eds) *Signal to Syntax.* Mahvah. NJ: Lawrence Erlbaum Associates, 117–34.

Baars, B.J. (1998) Metaphors of consciousness and attention in the brain. *Trends in Neurosciences* **21**, 51–89.

Bates, E. & Goodman, J.C. (1999) On the emergence of grammar from the lexicon. In: Macwhinney, B. (ed.) *The Emergence of Language.* Mahwah, NJ: Lawrence Erlbaum Associates, 29–80.

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.

Brooks, L.R. (1978) Nonanalytic concept formation and memory for instances. In: Rosch, E. & Lloyd, B.B. (eds) *Cognition and Categorization.* Hillsdale, NJ: Lawrence Erlbaum Associates, 169–215.

Carlson, R.A. & Dulany, D.D. (1985) Conscious attention and abstraction in concept learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **11**, 45–58.

Christiansen, M.H., Allen, J. & Seidenberg, M.S. (1998) Learning to segment speech using multiple cues: a connectionist model. *Language and Cognitive Processes* **13** 221–68.

Clark, A. & Thornton, C. (1997) Trading spaces: computation, representation and the limits of uninformed learning. *Behavioral and Brain Sciences* **20**, 57–90.

Cleeremans, A. (1993) *Mechanims of Implicit Learning: a Connectionist Model of Sequence Processing.* Cambridge, MA: MIT Press Bradford Books.

Cowan, N. (1995) *Attention and Memory: an Integrated Framework.* New York: Oxford University Press.

De Houwer, J., Hendrickx, H. & Baeyens, F. (1997) Evaluative learning with "subliminally" presented stimuli. *Consciousness and Cognition* **6**, 87–107.

Dulany, D.E. (1991) Conscious representation and thought systems. In: Wyer, R.S. & Srull, T.K. (eds) *Advances in Social Cognition,* Vol. 4. Hillsdale, NJ: Lawrence Erlbaum Associates, 97–120.

Dulany, D.E. (1997) Consciousness in the explicit (deliberative) and implicit (evocative). In: Cohen, J.D. & Schooler, J.W. (eds) *Scientific Approaches to the Study of Consciousness.* Mahwah, NJ: Lawrence Erlbaum Associates, 179–212.

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

Evans, J.St.B., Newstead, S.E. & Byrne, P. (1993) *Human Reasoning.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Frensch, P.A. & Miner, C.S. (1994) Effects of presentation rate and individual differences in short-term memory capacity on an indirect measure of serial learning. *Memory and Cognition* **22**, 95–110.

Grodzinsky, Y. (2000) The neurology of syntax: language use without Broca's area. *Behavioral and Brain Sciences* **23**, 1–21.

Haith, J. (1978) Visual competence in early infancy. In: Held, R., Leibowitz, H. & Teuber, H.L. (eds) *Handbook of Sensory Physiology Vol. 3: Perception.* Berlin: Springer-Verlag, 311–56.

Jacoby, L.L., Ste-Marie, D. & Toth, J.P. (1993) Redefining automaticity: unconscious influences, awareness and control. In: Baddeley, A.D. & Weiskrantz, L. (eds) *Attention, Selection, Awareness and Control. A Tribute to Donald Broadbent.* Oxford: Oxford University Press, 261–82.

Jiménez, L. & Mendez, C. (1999) Which attention is needed for implicit sequence learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition* **25**, 236–59.

Johnson-Laird, P.N. (1983) *Mental Models.* Cambridge, MA: Harvard University Press.

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.

LaBerge, D. & Samuels, S.J. (1974) Toward a theory of automatic information processing in reading. *Cognitive Psychology* **6**, 293–323.

Logan, G.D. (1988) Towards an instance theory of automatization. *Psychological Review* **76**, 165–78.

Logan, G.D. & Etherton, J.L. (1994) What is learned during automatization? The role of attention in constructing an instance. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **20**, 1022–50.

Mandler, G. (1975) Consciousness: respectable, useful, and probably necessary. In: Solso, R. (ed.) *Information Processing and Cognition: the Loyola Symposium.* Hillsdale, NJ: Lawrence Erlbaum Associates, 229–54.

Mandler, G. (1994) Hypermnesia, incubation and mind popping: on remembering without really trying. In: Umilta, C. & Moscovitch, M. (eds) *Attention and Performance XV: Conscious and Nonconscious Information Processing.* Cambridge, MA: MIT Press, 3–33.

McKoon, G. & **Ratcliff, R.** (1998) Memory-based language processing: psycholinguistic research in the 1960s. *Annual Review of Psychology* **49**, 25–32.

Miller, G.A. (1962) *Psychology: the Science of Mental Life.* New York, Harper & Row.

Neal, A. & **Hesketh, B.** (1997) Episodic knowledge and implicit learning. *Psychonomic Bulletin and Review* **4**, 24–37.

O'Brien, G. & **Opie, J.** (1999) A connectionist theory of phenomenal experience. *Behavioral and Brain Sciences* **22**, 127–48.

Perruchet, P. (1994) Learning from complex rule-governed environments: on the proper functions of nonconscious and conscious processes. In: Umilta, C. & Moscovitch, M. (eds) *Attention and Performance XV: Conscious and Nonconscious Information Processing.* Cambridge, MA: MIT Press, 811–35.

Perruchet, P. & **Vinter, A.** (1998) PARSER: a model for word segmentation. *Journal of Memory and Language* **39**, 246–63.

Perruchet, P. & **Vinter, A.** (2002) The self-organizing consciousness: a framework for implicit learning. In: French, R. & Cleeremans, A. (eds) *Implicit Learning.* Hove, UK: Psychology Press, 41–67.

Perruchet, P. & **Vinter, A.** (in press) The self-organizing consciousness. *Behavioral and Brain Sciences.*

Perruchet, P., Vinter, A. & **Gallego, J.** (1997) Implicit learning shapes new conscious percepts and representations. *Psychonomic Bulletin and Review* **4**, 43–8.

Posner, M.I. & **Boies, S.J.** (1971) Components of attention. *Psychological Review* **78**, 391–408.

Saffran, J.R., Newport, E.L. & **Aslin, R.N.** (1996) Word segmentation: the role of distributional cues. *Journal of Memory and Language* **35**, 606–21.

Saffran, J.R., Newport, E.L., Aslin, R.N., Tunick, R.A. & **Barrueco, S.** (1997) Incidental language learning. *Psychological Science* **8**, 101–5.

Shastri, L. & **Ajjanagadde, V.** (1993) From simple associations to systematic reasoning. *Behavioral and Brain Sciences* **16**, 417–94.

Shiffrin, R.M. & **Schneider, W.** (1977) Controlled and automatic human information processing: II. Perceptual learning, automatic attention and a general theory. *Psychological Review* **84**, 127–90.

Smith, S.M. & **Blankenship, S.E.** (1989) Incubation effects. *Bulletin of the Psychonomic Society* **27**, 311–14.

Smith, S.M. & **Blankenship, S.E.** (1991) Incubation and the persistence of fixation in problem solving. *American Journal of Psychology* **104**, 61–87.

Stadler, M.A. (1995) Role of attention in implicit learning. *Journal of Experimental Psychology: Learning Memory, and Cognition* **21**, 674–85.

Tomasello, M. (2000) The item-based nature of children's early syntactic development. *Trends in Cognitive Sciences* **4**, 156–64.

Tzelgov, J. (1997) Automatic but conscious: that is how we act most of the time. In: Wyer, R.S. (ed.) *The Automaticity of Everyday Life, Advances in Social Cognition,* Vol. X. Mahwah, NJ: Lawrence Erlbaum Associates, 217–30.

Wagner, A.R. (1981) SOP: a model of automatic memory processing in animal behavior. In: Spear, N.E. & Miller, R.R. (eds) *Information Processing in Animals: Memory Mechanisms.* Hillsdale, NJ: Lawrence Erlbaum Associates, 5–47.

Wallas, G. (1926) *The Art of Thought.* New York: Harcourt, Brace.